NATIONAL SCIENCE FOUNDATION

JOURNAL OF THE NATIONAL SCIENCE FOUNDATION OF SRI LANKA

Contact details
Editorial Office, National Science Foundation,
47/5, Maitland Place, Colombo 07, Sri Lanka.

E-mail : jnsf@nsf.gov.lk
Phone : +94-11- 2696771
JNSF online submission portal :
https://jnsfsl.sljol.info/about/submissions
JNSF home page :
http://www.nsf.gov.lk/ index.php/nsfscience magazine

Publication : Published quarterly (March, June, September and December) by the National Science Foundation of Sri Lanka.

Manuscripts: Research Articles, Research Communications, Reviews and Correspondences in all fields of Science and Technology can be submitted for consideration for publication. A guide to the preparation of manuscripts is provided in each issue. The guidelines may also be obtained by visiting the NSF website or JNSF online submission portal.

Disclaimer: No responsibility is assumed by the National Science Foundation of Sri Lanka for statements and opinions expressed by contributors to this Journal.

Publication : free A publication fee of US$ 250 will be levied for each manuscript in two stages except, when the corresponding author is affiliated with a Sri Lankan institution.

- A processing fee of US$ 20 will belevied for each manuscript at peer-review stage.
- Remaining US$ 230 will be charged for accepted articles at the time of publication.

Indexing : **The JNSF is indexed in Science Citation Index Expanded, Journal Citation Reports/Science Edition, BIOSIS Previews, Zoological Record, Biological Abstracts, Chemical Abstracts, Scopus, DOAJ, TEEAL, Ulrich's, AGRICOLA and EBSCOhost, CAB Abstracts, SafetyLit, Journal TOCs, EBSCO Applied Science & Technology Source Ultimate**

JOURNAL OF THE
NATIONAL SCIENCE FOUNDATION
OF SRI LANKA

Volume 54 Number 1 March 2026

CONTENTS

RESEARCH COMMUNICATIONS

Cover: Kinniya cattle: A locally bred draught cattle population in Sri Lanka
Top left: Kinniya cattle female (left), Kinniya cattle male (right),
Bottom left: Kinniya cattle herd,
Right: Uniform orientation in horn morphology (straight or curved)
See J.Natn.Sci.Foundation Sri Lanka 2026 54(1): 63 - 75

J.Natn.Sci.Foundation Sri Lanka 2026 54 (1): 1
DOI: http://dx.doi.org/10.4038/jnsfsr.v54i1.13031

EDITORIAL

The human gut microbiome: A meeting place for Ayurveda and modern medicine

Ayurveda and modern medicine represent two formulations of the reality behind the same phenomenon with respect to human existence; maintenance of health and the management and cure of disease. These two formulations (medical systems) are rooted in two fundamentally different knowledge systems. Modern medicine is rooted in the reductionist approach of modern science, based on the concepts of physics, chemistry and biology. Ayurveda is rooted in the holistic approach of the Indian Knowledge system, based on the concept of the *Panchamahabhuta*. The *Panchamahabuta* refers to the five 'elements' of space, air, fire, water and earth, which are considered the fundamental constituents of all manifested reality.

Ayurvedic treatments for various diseases are based on diet, medication and practices. Particular emphasis is given to diet, specifying foods that should or should not be consumed. Many therapeutic regimens commence with purgation treatments. It is apparent that modulation of the gut microbiome can be one of the outcomes of these treatment modalities. One of the most interesting areas of research in modern medicine currently is the gut microbiome-immune system-brain axis. The manifestations of a variety of diseases ranging from psychiatric illness, Alzheimer's disease, cardiac disease, cancer and skin diseases to autoimmune diseases such as multiple sclerosis are being investigated for their interconnection with the gut microbiota.

The acknowledged applicability and relevance of both Ayurveda and modern medicine in many areas relating to health and disease pose the intriguing question of how the basic concepts of each formulation relate to one another. While Ayurveda does not refer to the gut microbiome, it uses the concept of *Agni* (fire), the energy that powers all physical functions, particularly digestion. It governs not only digestion, absorption and metabolism, but also all physical functions and mental and emotional processes. Thus, interpretation of Ayurvedic clinical practice in terms of the gut microbiome-immune system-brain axis, where applicable, can potentially yield valuable insights to both systems of medicine.

Ajit Abeysekera

J.Natn.Sci.Foundation Sri Lanka 2025 54 (1): 3 - 16
DOI: http://dx.doi.org/10.4038/jnsfsr.v54i1.12387

RESEARCH ARTICLE

Mathematical Modelling

Revan indices for the hexadentate-3-hydroxypyridinones-terminated dendrimers

DC Gunawardhana[1*], GHJ Lanel[2] and MGN Lakshani[2]
[1] *Department of Mathematical Sciences, Faculty of Applied Sciences, South Eastern University of Sri Lanka, Sammanthurai, Sri Lanka.*
[2] *Department of Mathematics, Faculty of Applied Sciences, University of Sri Jayewardenepura, Nugegoda, Sri Lanka.*

Submitted: 07 October 2024; Revised: 25 November 2025; Accepted: 30 January 2026

Abstract: Dendrimers are highly ordered, unimolecular, monodisperse, micellar nanostructures, applied in various fields such as drug delivery, nanotechnology, and biomedical applications. Hexadentate 3-hydroxypyridinones-terminated dendrimers are used in iron binding and anti-microbial activities. In the molecular graph, atoms are the vertices, and the bonds are the edges. A graph G consists of a pair of nonempty sets of vertices $V(G)$ and edges $E(G)$ and $d_G(u)$ is the degree of a vertex $u \in V(G)$. In this study, multiple bonds were modeled as multi-edges, and all hydrogen atoms were included in the molecular graph, thereby reducing the assumptions of traditional computational methods. Here, different Revan indices, which are topological descriptors derived from graph theory, were computed for hexadentate-3-hydroxypyridinones-terminated dendrimers. The Revan degree $r_G(u)$ of a vertex u is defined as $r_G(u) = \Delta(G) + \delta(G) - d_G(u)$, where $\Delta(G)$ is the maximum degree and $\delta(G)$ is the minimum degree among the vertices in G. In this paper, the first, second, and third Revan indices, forgotten Revan index, hyper Revan index, atom-bond connectivity Revan index, product connectivity Revan index, sum connectivity Revan index, harmonic Revan index, and geometric arithmetic Revan index were computed for different generations of hexadentate 3-hydroxypyridinones-terminated dendrimers. The results demonstrate that the complex topology in dendrimers can be accurately represented with the Revan indices, and as a result, their structural properties can be better understood. This study underlines the importance of these topological indices in directing design and development toward advanced applications with dendrimer-based systems.

Keywords: Dendrimers, hexadentate 3-hydroxypyridinones-terminated dendrimers, Revan degree, Revan Index, topological index.

INTRODUCTION

Dendrimers are synthetic polymeric macromolecules of nanometer dimensions with well-defined, homogeneous, and monodisperse structure, and their structure consists of tree-like arms or branches. They provide controlled release from the inner core, and the active compounds are incorporated in the interior as well as begin attached on the surface. Their unique architecture is highly useful in drug delivery, nanotechnology, and biomedical operations (Abbasi, 2014). The idea of the dendrimers was introduced first in 1978, and they are still laboriously delved into worldwide (Vogtle et al., 1978). Hexadentate 3-hydroxypyridinones-terminated dendrimers were synthesized for iron binding and implicit antibacterial operations (Zhou et al., 2018).

Chemical graph theory is a branch of mathematics that helps to study molecules in chemistry. This simplifies molecular structures, enabling quantitative analysis, property prediction, drug discovery, and the development of algorithms for efficient chemical research. A molecular graph is a simple graph that does not contain loops or

* Corresponding author (rehan@seu.ac.lk; https://orcid.org/0009-0003-2081-1273)

multiple edges. A graph $G = (V, E)$ consists of a finite nonempty set of vertices $V = V(G)$ and a set of edges $E = E(G)$. Two vertices of G connected by an edge, are said to be adjacent. The degree $d_v(G)$ or d_v of a vertex v in $V(G)$ is the number of vertices of G, adjacent to the vertex v (Gutman, 2013).

A topological index is a numerical descriptor that is computed based on the molecular graph of a chemical compound and used to determine the chemical and physical properties of any chemical compound. These indices derive from the mathematical field of graph theory, which studies networks or graphs. Different aspects of the molecule structure are captured by each topological index (Babujee et al., 2013).

The Wiener index is the first distance-based topological index in chemical graph theory and has been studied the most. In 1947, Harold Wiener presented this index as the path number. It was defined as the total distance between each carbon atom in the molecules, expressed in terms of carbon-carbon bonds (Wiener, 1947).

The Zagreb indices are the first vertex-degree-based topological indices (Gutman & Trinajstić, 1972). The definitions of the first Zagreb index $M_1(G)$ and the second Zagreb index $M_2(G)$ are $M_1(G) = \sum_{uv \in E(G)} (d_u + d_v)$ and $M_2(G) = \sum_{uv \in E(G)} d_u d_v$

Another popular and studied degree-based topological index is the Randić index. This is the first genuine degree-based topological index (Randić, 1975) and is defined as

$$R(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{d_u d_v}}$$

In 2009, the sum connectivity index of a graph G (Zhou & Trinajstić, 2009) was introduced and defined as,

$$S(G) = \sum_{uv \in E(G)} \sqrt{\frac{1}{d_u + d_v}}$$

In 1993, the harmonic index of a graph G (Favaron et al., 1993) was introduced and defined as,

$$H(G) = \sum_{uv \in E(G)} \frac{2}{d_u + d_v}$$

In 2009, the geometric arithmetic index of a graph G (Vukičević & Furtula, 2009) was introduced and defined as,

$$GA(G) = \sum_{uv \in E(G)} \frac{2\sqrt{d_u d_v}}{d_u + d_v}$$

The atom-bond connectivity index of a graph G (Estrada et al., 1998) was introduced and defined as,

$$ABC(G) = \sum_{uv \in E(G)} \sqrt{\frac{d_u + d_v - 2}{d_u d_v}}$$

The forgotten topological index of a graph G (Furtula & Gutman, 2015) was defined as,

$$F(G) = \sum_{uv \in E(G)} (d_u{}^2 + d_v{}^2)$$

In this study, the first, second, and third Revan indices, along with the forgotten Revan index, hyper Revan index, atom-bond connectivity Revan index, product connectivity Revan index, sum connectivity Revan index, harmonic Revan index, and geometric arithmetic Revan index, are computed for the hexadentate 3-hydroxypyridinone-terminated dendrimers. Connecting the Revan vertices u and v is the edge uv, which is a Revan edge. For any vertex v in G, the Revan vertex degree $r_G(v)$ is defined as

$$r_G(v) = \Delta(G) + \delta(G) - d_G(v).$$

Here $\Delta(G)$ is the maximum degree among the vertices in G and $\delta(G)$ is the minimum degree among the vertices in G (Kulli, 2017a).

MATERIALS AND METHODS

The molecular graph describes the dendrimeric structure in which the atoms and bonds are treated as vertices and edges, respectively. In this study, all the multi-bonds were treated as multi-edges, and all hydrogen atoms were included in a molecular graph. This is an enhanced method for computing topological indices to reduce the assumptions of traditional computing techniques (Gunawardhana et al., 2024). In this paper, we computed different versions of Revan indices for four types of hexadentate 3-hydroxypyridinones-terminated dendrimers using the following formulas.

The molecular structures considered in this study were adapted from Zhou et al. (2018). The analyzed structures correspond to compound 6b from Scheme 1 (Figure 1), structure 7 from Scheme 2 (Figure 2), compound 10 from Scheme 3 (Figure 3), and compound 13 from Scheme 4 (Figure 4) of Zhou et al. Each structure was treated as an independent molecular graph for the computation of Revan indices. Figure 2 represents the benzyl-protected dendron intermediate (structure 7), which was also independently analyzed in the present graph-theoretical study.

In 2017, the first, second, and third Revan indices of a graph G (Kulli, 2017b) were introduced and defined as,

$$R_1(G) = \sum_{uv \in E(G)} r_G(u) + r_G(v) \quad ...(01)$$

$$R_2(G) = \sum_{uv \in E(G)} r_G(u)\, r_G(v) \quad ...(02)$$

$$R_3(G) = \sum_{uv \in E(G)} |r_G(u) - r_G(v)| \quad ...(03)$$

The F-Revan index (also called forgotten Revan index) of a graph G (Kulli, 2018b) was defined as,

$$FR(G) = \sum_{uv \in E(G)} r_G(u)^2 + r_G(v)^2 \quad ...(04)$$

The first hyper Revan index of a graph G (Kulli, 2018c) was defined as,

$$HR_1(G) = \sum_{uv \in E(G)} (r_G(u) + r_G(v))^2 \quad ...(05)$$

The atom bond connectivity Revan index of a graph G (Estrada, 1998) was defined as,

$$ABCR(G) = \sum_{uv \in E(G)} \sqrt{\frac{r_G(u) + r_G(v) - 2}{r_G(u) r_G(v)}} \quad ...(06)$$

The product connectivity Revan index (also called Randi`c Revan index) of a graph G (Kulli, 2017a) was defined as,

$$RR(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{r_G(u) r_G(v)}} \quad ...(07)$$

The sum connectivity Revan index of a graph G (Kulli, 2017c) was defined as,

$$SR(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{r_G(u) + r_G(v)}} \quad ...(08)$$

The harmonic Revan index of a graph G (Kulli, 2020) was defined as,

$$HMR(G) = \sum_{uv \in E(G)} \frac{2}{r_G(u) + r_G(v)} \quad ...(09)$$

The geometric-arithmetic Revan index of a graph G (Kulli, 2018a) was defined as,

$$GAR(G) = \sum_{uv \in E(G)} \frac{2\sqrt{r_G(u) r_G(v)}}{r_G(u) + r_G(v)} \quad ...(10)$$

The Revan indices were computed using respective formulas. Zhou et al. synthesized a range of novel hexadentate 3-hydroxypyridinone-terminated dendrimers which were shown as Figures in this paper (Zhou et al., 2018). As Figure 1 illustrates, first generation dendrimeric chelators that is consists of three hexadentate moieties. The structure of the second generation dendrimeric chelators is depicted in Figure 2. It was conjugated with tri-acid and di-acid, respectively, to form structures in Figures 3 and 4, which have nine and six hexadentate centres, respectively. In this study, the first, second, and third Revan indices, along with the forgotten Revan index, hyper Revan index, atom-bond connectivity Revan index, product connectivity Revan index, sum connectivity Revan index, harmonic Revan index, and geometric arithmetic Revan index are computed for the above-mentioned dendrimers. Thus far, these selected Revan indices have not been calculated for the above dendrimeric structures.

RESULTS AND DISCUSSION

Theorem 1.

Let G be the first generation dendrimeric chelators. Then the first Revan index $R_1(G)$, the second Revan index $R_2(G)$, the third Revan index $R_3(G)$, the forgotten Revan index $FR(G)$, the first Hyper Revan index $HR_1(G)$, the ABC Revan index $ABCR(G)$, the Randi`c Revan index $RR(G)$, the Sum connectivity Revan index $SR(G)$, the harmonic Revan index $HMR(G)$, and the GA Revan index $GAR(G)$ for G are

i. $R_1(G) = 1506$,

ii. $R_2(G) = 1266$,

iii. $R_3(G) = 618$,

iv. $FR(G) = 4068$,

v. $HR_1(G) = 6600$,

vi. $ABCR(G) = 9\sqrt{\frac{5}{3}} + 75\sqrt{\frac{1}{2}} + 63\sqrt{3} + 66\sqrt{\frac{2}{3}}$,

vii. $RR(G) = 171 + \frac{135}{2}\sqrt{\frac{1}{2}} + 75\sqrt{\frac{1}{3}}$,

viii. $SR(G) = 18\sqrt{\frac{1}{7}} + 15\sqrt{\frac{1}{6}} + 126\sqrt{\frac{1}{5}} + 108\sqrt{\frac{1}{2}} + 33 + 60\sqrt{\frac{1}{3}}$,

ix. $HMR(G) = \frac{15753}{70}$, and

x. $GAR(G) = \frac{1044}{5} + 50\sqrt{2} + \frac{303}{7}\sqrt{3}$.

Proof:

From the Figure 1, it is easy to see that the vertices of G have degree 1,2,3, or 4. Thus $\Delta(G) = 4$, $\delta(G) = 1$, and therefore $r_G(u) = 4 + 1 - d_G(u)$ is the Revan degree of any vertex u. According to the Revan vertex degree, there are six types of Revan edge sets in the structure of the first generation dendrimeric chelators G shown in Figure 1, as follows:

$E_1 = \{uv \in E(G) | r_G(u) = 4, r_G(v) = 3\}$, $|E_1| = 18$

$E_2 = \{uv \in E(G) | r_G(u) = 4, r_G(v) = 2\}$, $|E_2| = 15$

$E_3 = \{uv \in E(G) | r_G(u) = 4, r_G(v) = 1\}$, $|E_3| = 126$

$E_4 = \{uv \in E(G) | r_G(u) = r_G(v) = 1\}$, $|E_4| = 108$

$E_5 = \{uv \in E(G) | r_G(u) = 3, r_G(v) = 1\}$, $|E_5| = 66$

$E_6 = \{uv \in E(G) | r_G(u) = 1, r_G(v) = 2\}$, $|E_6| = 60$

By using the above Revan edge partitions,

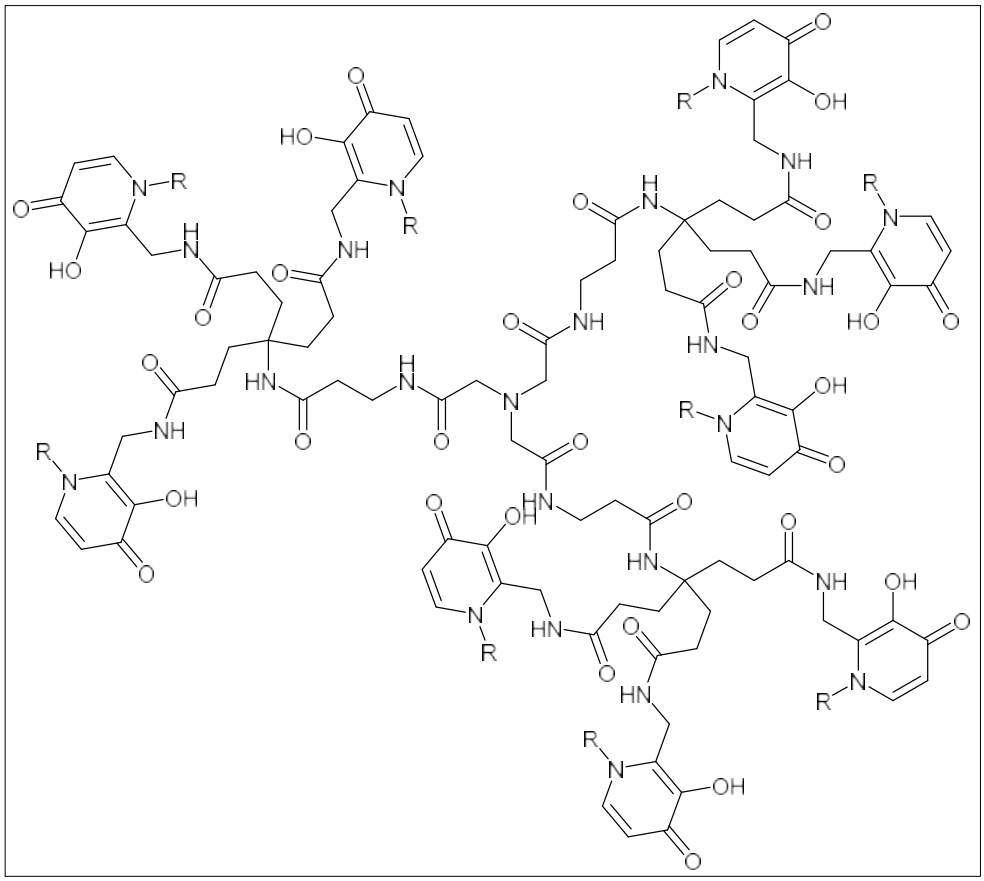


Figure 1: First-generation dendrimeric chelator corresponding to compound 6b in Scheme 1 of Zhou et al. (2018), containing three hexadentate chelating ligands and nine 3-hydroxypyridinone units (R = CH2CH2OH). Structure adapted from Zhou et al. (2018).

$R_1(G) = \sum_{uv\in E(G)} \quad r_G(u) + r_G(v)$

$= 18(4 + 3) + 15(4 + 2) + 126(4 + 1) +$

$108(1 + 1) + 66(3 + 1) + 60(1 + 2)$

$= 1506.$

$R_2(G) = \sum_{uv\in E(G)} \quad r_G(u)\, r_G(v)$

$= 18(4 \times 3) + 15(4 \times 2) + 126(4 \times 1) + 108(1 \times 1) +$

$66(3 \times 1) + 60(1 \times 2)$

$= 1266.$

$R_3(G) = \sum_{uv\in E(G)} \quad |r_G(u) - r_G(v)|$

$= 18|4 - 3| + 15|4 - 2| + 126|4 - 1| + 108|1 - 1| +$

$66|3 - 1| + 60|1 - 2|$

$= 618.$

$FR(G) = \sum_{uv\in E(G)} \quad r_G(u)^2 + r_G(v)^2$

$= 18(4^2 + 3^2) + 15(4^2 + 2^2) + 126(4^2 + 1^2) +$

$108(1^2 + 1^2) + 66(3^2 + 1^2) + 60(1^2 + 2^2)$

$= 4068.$

$HR_1(G) = \sum_{uv\in E(G)} \quad (r_G(u) + r_G(v))^2$

$= 18(4 + 3)^2 + 15(4 + 2)^2 + 126(4 + 1)^2 +$

$108(1 + 1)^2 + 66(3 + 1)^2 + 60(1 + 2)^2$

$= 6600.$

$ABCR(G) = \sum_{uv\in E(G)} \quad \sqrt{\frac{r_G(u)+r_G(v)-2}{r_G(u)r_G(v)}}$

$= 18\sqrt{\frac{4+3-2}{4\times3}} + 15\sqrt{\frac{4+2-2}{4\times2}} + 126\sqrt{\frac{4+1-2}{4\times1}} +$

$108\sqrt{\frac{1+1-2}{1\times1}} + 66\sqrt{\frac{3+1-2}{3\times1}} + 60\sqrt{\frac{1+2-2}{1\times2}}$

$= 9\sqrt{\frac{5}{3}} + 75\sqrt{\frac{1}{2}} + 63\sqrt{3} + 66\sqrt{\frac{2}{3}}.$

$RR(G) = \sum_{uv\in E(G)} \quad \frac{1}{\sqrt{r_G(u)r_G(v)}}$

$= 18\sqrt{\frac{1}{4\times3}} + 15\sqrt{\frac{1}{4\times2}} + 126\sqrt{\frac{1}{4\times1}} + 108\sqrt{\frac{1}{1\times1}} +$

$66\sqrt{\frac{1}{3\times1}} + 60\sqrt{\frac{1}{1\times2}}$

$= 171 + \frac{135}{2}\sqrt{\frac{1}{2}} + 75\sqrt{\frac{1}{3}}.$

$SR(G) = \sum_{uv\in E(G)} \quad \frac{1}{\sqrt{r_G(u)+r_G(v)}}$

$= 18\sqrt{\frac{1}{4+3}} + 15\sqrt{\frac{1}{4+2}} + 126\sqrt{\frac{1}{4+1}} + 108\sqrt{\frac{1}{1+1}}$

$+ 66\sqrt{\frac{1}{3+1}} + 60\sqrt{\frac{1}{1+2}}$

$= 18\sqrt{\frac{1}{7}} + 15\sqrt{\frac{1}{6}} + 126\sqrt{\frac{1}{5}} + 108\sqrt{\frac{1}{2}} + 33 + 60\sqrt{\frac{1}{3}}.$

$HMR(G) = \sum_{uv\in E(G)} \quad \frac{2}{r_G(u)+r_G(v)}$

$= 18\left(\frac{2}{4+3}\right) + 15\left(\frac{2}{4+2}\right) + 126\left(\frac{2}{4+1}\right) + 108\left(\frac{2}{1+1}\right)$

$+ 66\left(\frac{2}{3+1}\right) + 60\left(\frac{2}{1+2}\right)$

$= \frac{15753}{70}.$

$GAR(G) = \sum_{uv\in E(G)} \quad \frac{2\sqrt{r_G(u)r_G(v)}}{r_G(u)+r_G(v)}$

$= 18\left(\frac{2\sqrt{4\times3}}{4+3}\right) + 15\left(\frac{2\sqrt{4\times2}}{4+2}\right) + 126\left(\frac{2\sqrt{4\times1}}{4+1}\right)$

$+ 108\left(\frac{2\sqrt{1\times1}}{1+1}\right) + 66\left(\frac{2\sqrt{3\times1}}{3+1}\right) + 60\left(\frac{2\sqrt{1\times2}}{1+2}\right)$

$= \frac{1044}{5} + 50\sqrt{2} + \frac{303}{7}\sqrt{3}.$

Theorem 2.

Let G be the second generation dendrimeric chelators. Then the first Revan index $R_1(G)$, the second Revan index $R_2(G)$, the third Revan index $R_3(G)$, the forgotten Revan index $FR(G)$, the first Hyper Revan index $HR_1(G)$, the ABC Revan index $ABCR(G)$, the Randi`c Revan index $RR(G)$, the Sum connectivity Revan index $SR(G)$, the Harmonic Revan index $HMR(G)$, and the Geometric-Arithmatic Revan index $GAR(G)$ for G are

i. $R_1(G) = 2336,$

ii. $R_2(G) = 1749,$

iii. $R_3(G) = 1018$

iv. $FR(G) = 6260,$

v. $HR_1(G) = 9758$,

vi. $ABCR(G) = 39\sqrt{2} + 131\sqrt{3} + 68\sqrt{\frac{2}{3}}$,

vii. $RR(G) = 364 + 69\sqrt{\frac{1}{2}} + 68\sqrt{\frac{1}{3}}$

viii. $SR(G) = 34 + 233\sqrt{\frac{1}{2}} + 60\sqrt{\frac{1}{3}} + 262\sqrt{\frac{1}{5}} + 18\sqrt{\frac{1}{6}}$

ix. $HMR(G) = \frac{2089}{5}$, and

x. $GAR(G) = \frac{2213}{5} + 52\sqrt{2} + 34\sqrt{3}$.

Proof:

From the Figure 2, it is easy to see that the vertices of G have degree 1,2,3, or 4. Thus $\Delta(G) = 4$, $\delta(G) = 1$, and therefore $r_G(u) = 4 + 1 - d_G(u)$ is the Revan degree of any vertex u. According to the Revan vertex degree, there are five types of Revan edge sets in the structure of the second generation dendrimeric chelators G shown in Figure 2, as follows:

$E_1 = \{uv \in E(G) | r_G(u) = 4, r_G(v) = 2\}, \quad |E_1| = 18$

$E_2 = \{uv \in E(G) | r_G(u) = 4, r_G(v) = 1\}, \quad |E_2| = 262$

$E_3 = \{uv \in E(G) | r_G(u) = r_G(v) = 1\}, \quad |E_3| = 233$

$E_4 = \{uv \in E(G) | r_G(u) = 3, r_G(v) = 1\}, \quad |E_4| = 68$

$E_5 = \{uv \in E(G) | r_G(u) = 1, r_G(v) = 2\}, \quad |E_5| = 60$

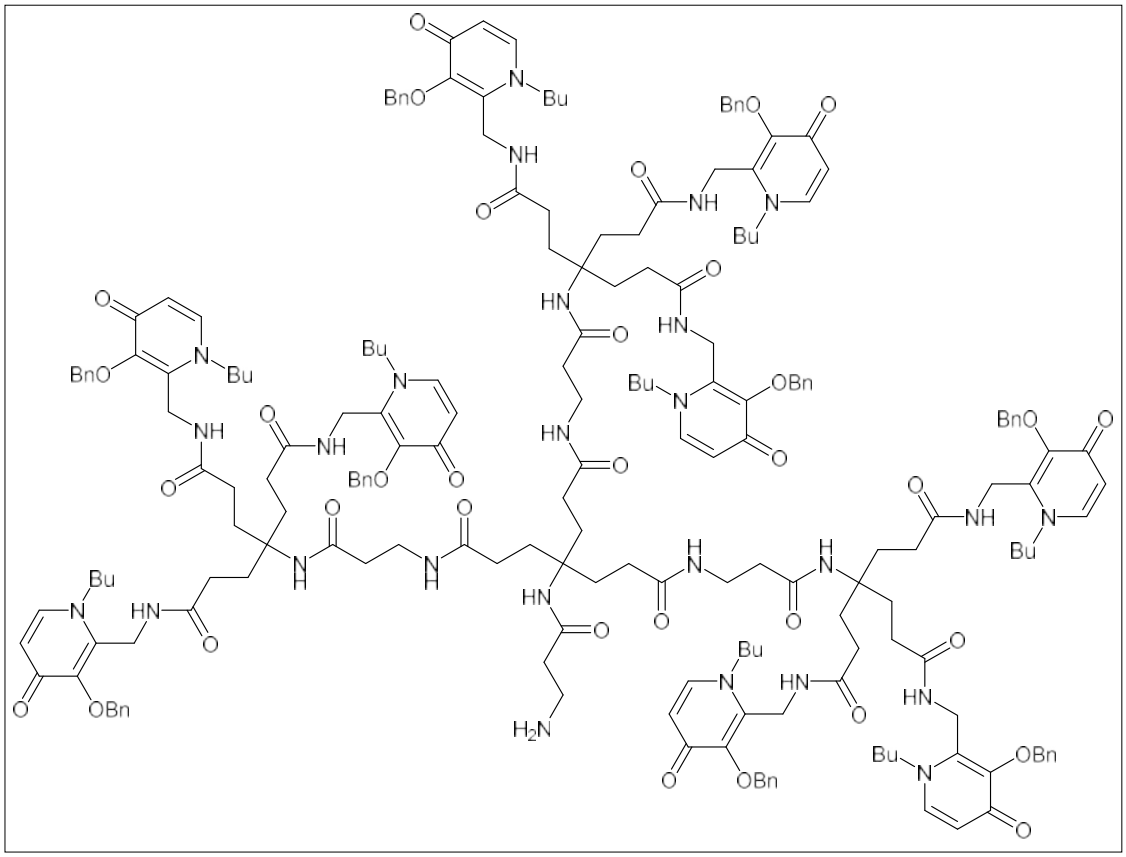


Figure 2: Benzyl-protected first-generation dendron intermediate corresponding to structure 7 in Scheme 2 of Zhou et al. (2018), containing three hexadentate chelating ligands and nine benzyl-protected 3-hydroxypyridinone units. Structure adapted from Zhou et al. (2018).

By using the above Revan edge partitions,

$$R_1(G) = \sum_{uv\in E(G)} \; r_G(u) + r_G(v)$$
$$= \; 18(4 \; + \; 2) \; + \; 262(4 \; + \; 1) \; + \; 233(1 \; + \; 1) \; + \; 68(3 \; + \; 1) \; + \; 60(1 \; + \; 2)$$
$$= 2336.$$

$$R_2(G) = \sum_{uv\in E(G)} \; r_G(u)\, r_G(v)$$
$$= \; 18(4 \times 2) \; + \; 262(4 \times 1) \; + \; 233(1 \times 1) \; + \; 68(3 \times 1) \; + \; 60(1 \times 2)$$
$$= \; 1749.$$

$$R_3(G) = \sum_{uv\in E(G)} \; |r_G(u) - \; r_G(v)|$$
$$= \; 18|4 - 2| + \; 262|4 - 1| + \; 233|1 - 1| + \; 68|3 - 1| + \; 60|1 - 2|$$
$$= \; 1018.$$

$$FR(G) = \sum_{uv\in E(G)} \; r_G(u)^2 + r_G(v)^2$$
$$= \; 18(4^2 \; + \; 2^2) \; + \; 262(4^2 \; + \; 1^2) \; + \; 233(1^2 \; + \; 1^2) \; + \; 68(3^2 \; + \; 1^2) \; + \; 60(1^2 \; + \; 2^2)$$
$$= \; 6260.$$

$$HR_1(G) = \sum_{uv\in E(G)} \; (r_G(u) + r_G(v))^2$$
$$= \; 18(4 \; + \; 2)^2 \; + \; 262(4 \; + \; 1)^2 + \; 233(1 \; + \; 1)^2 \; + \; 68(3 \; + \; 1)^2 \; + \; 60(1 + 2)^2$$
$$= \; 9758.$$

$$ABCR(G) = \sum_{uv\in E(G)} \; \sqrt{\frac{r_G(u)+r_G(v)-2}{r_G(u)r_G(v)}}$$
$$= 18\sqrt{\frac{4+2-2}{4\times 2}} + 262\sqrt{\frac{4+1-2}{4\times 1}} + 233\sqrt{\frac{1+1-2}{1\times 1}} + 68\sqrt{\frac{3+1-2}{3\times 1}} + 60\sqrt{\frac{1+2-2}{1\times 2}}$$
$$= 39\sqrt{2} + 131\sqrt{3} + 68\sqrt{\frac{2}{3}}.$$

$$RR(G) = \sum_{uv\in E(G)} \; \frac{1}{\sqrt{r_G(u)r_G(v)}}$$
$$= 18\sqrt{\frac{1}{4\times 2}} + 262\sqrt{\frac{1}{4\times 1}} + 233\sqrt{\frac{1}{1\times 1}} + 68\sqrt{\frac{1}{3\times 1}} + 60\sqrt{\frac{1}{1\times 2}}$$
$$= 364 + 69\sqrt{\frac{1}{2}} + 68\sqrt{\frac{1}{3}}.$$

$$HMR(G) = \sum_{uv \in E(G)} \frac{2}{r_G(u)+r_G(v)}$$
$$= 18\left(\frac{2}{4+2}\right) + 262\left(\frac{2}{4+1}\right) + 233\left(\frac{2}{1+1}\right) + 68\left(\frac{2}{3+1}\right) + 60\left(\frac{2}{1+2}\right)$$
$$= \frac{2089}{5}.$$

$$GAR(G) = \sum_{uv \in E(G)} \frac{2\sqrt{r_G(u)r_G(v)}}{r_G(u)+r_G(v)}$$
$$= 18\left(\frac{2\sqrt{4\times2}}{4+2}\right) + 262\left(\frac{2\sqrt{4\times1}}{4+1}\right) + 233\left(\frac{2\sqrt{1\times1}}{1+1}\right) + 68\left(\frac{2\sqrt{3\times1}}{3+1}\right) + 60\left(\frac{2\sqrt{1\times2}}{1+2}\right)$$
$$= \frac{2213}{5} + 52\sqrt{2} + 34\sqrt{3}.$$

Theorem 3.

Let G be the dendrimeric chelators contain six hexadentate centers. Then the first Revan index $R_1(G)$, the second Revan index $R_2(G)$, the third Revan index $R_3(G)$, the forgotten Revan index $FR(G)$, the first Hyper Revan index $HR_1(G)$, the ABC Revan index $ABCR(G)$, the Randi`c Revan index $RR(G)$, the Sum connectivity Revan index $SR(G)$, the Harmonic Revan index $HMR(G)$, and the Geometric-Arithmatic Revan index $GAR(G)$ for G are

i. $R_1(G) = 3876$,

ii. $R_2(G) = 3075$,

iii. $R_3(G) = 1698$,

iv. $FR(G) = 10632$,

v. $HR_1(G) = 16386$,

vi. $ABCR(G) = 81\sqrt{2} + 205\sqrt{3} + 126\sqrt{\frac{2}{3}} + 9\sqrt{\frac{5}{3}}$,

vii. $RR(G) = 506 + 72\sqrt{2} + 45\sqrt{3}$,

viii. $SR(G) = 63 + 301\sqrt{\frac{1}{2}} + 126\sqrt{\frac{1}{3}} + 410\sqrt{\frac{1}{5}} + 36\sqrt{\frac{1}{6}} + 18\sqrt{\frac{1}{7}}$,

ix. $HMR(G) = \frac{4404}{7}$, and

x. $GAR(G) = 629 + 108\sqrt{2} + \frac{513}{7}\sqrt{3}$.

Proof:

From the Figure 3, it is easy to see that the vertices of G have degree 1, 2, 3, or 4. Thus $\Delta(G) = 4$, $\delta(G) = 1$, and therefore $r_G(u) = 4 + 1 - d_G(u)$ is the Revan degree of any vertex *u*. According to the Revan vertex degree, there are six types of Revan edge sets in the structure of the dendrimeric chelators contain six hexadentate centres G shown in Figure 3, as follows:

$E_1 = \{uv \in E(G) | r_G(u) = 4, r_G(v) = 3\}$, $|E_1| = 18$

$E_2 = \{uv \in E(G) | r_G(u) = 4, r_G(v) = 2\}$, $|E_2| = 36$

$E_3 = \{uv \in E(G) | r_G(u) = 4, r_G(v) = 1\}$, $|E_3| = 410$

$E_4 = \{uv \in E(G) | r_G(u) = r_G(v) = 1\}$, $|E_4| = 301$

$E_5 = \{uv \in E(G) | r_G(u) = 3, r_G(v) = 1\}$, $|E_5| = 126$

$E_6 = \{uv \in E(G) | r_G(u) = 1, r_G(v) = 2\}$, $|E_6| = 126$

By using the above Revan edge partitions,

$$R_1(G) = \sum_{uv \in E(G)} r_G(u) + r_G(v)$$
$$= 18(4 + 3) + 36(4 + 2) + 410(4 + 1) + 301(1 + 1) + 126(3 + 1) + 126(1 + 2)$$
$$= 3876.$$

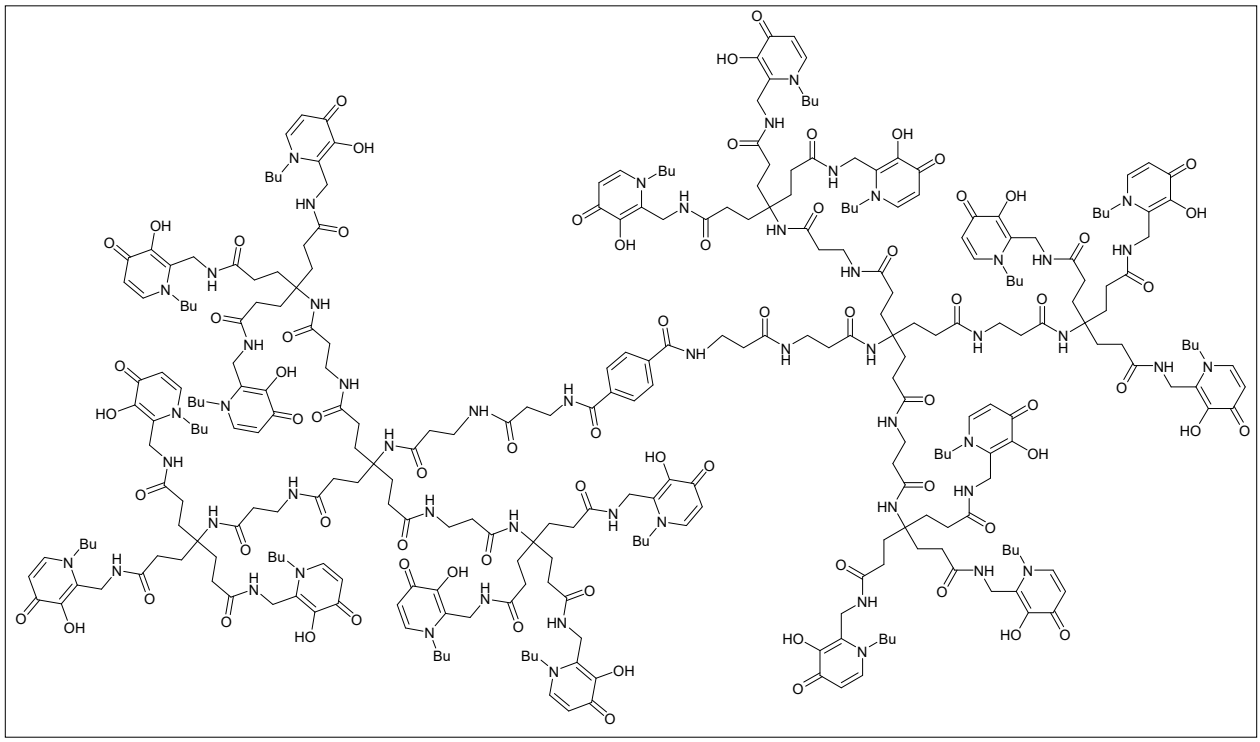


Figure 3: Second-generation dendrimeric chelator corresponding to compound 10 in Scheme 3 of Zhou et al. (2018), containing six hexadentate chelating ligands and eighteen 3-hydroxypyridinone units. Structure adapted from Zhou et al. (2018).

$$R_2(G) = \sum_{uv \in E(G)} \; r_G(u)\, r_G(v)$$

$$= 18(4 \times 3) + 36(4 \times 2) + 410(4 \times 1) + 301(1 \times 1) + 126(3 \times 1) + 126(1 \times 2)$$

$$= 3075.$$

$$R_3(G) = \sum_{uv \in E(G)} \; |r_G(u) - r_G(v)|$$

$$= 18|4 - 3| + 36|4 - 2| + 410|4 - 1| + 301|1 - 1| + 126|3 - 1| + 126|1 - 2|$$

$$= 1698.$$

$$FR(G) = \sum_{uv \in E(G)} \; r_G(u)^2 + r_G(v)^2$$

$$= 18(4^2 + 3^2) + 36(4^2 + 2^2) + 410(4^2 + 1^2) + 301(1^2 + 1^2) + 126(3^2 + 1^2) + 126(1^2 + 2^2)$$

$$= 10632.$$

$$HR_1(G) = \sum_{uv \in E(G)} \; (r_G(u) + r_G(v))^2$$

$$= 18(4 + 3)^2 + 36(4 + 2)^2 + 410(4 + 1)^2 + 301(1 + 1)^2 + 126(3 + 1)^2 + 126(1 + 2)^2$$

$$= 16386.$$

$$ABCR(G) = \sum_{uv\in E(G)} \sqrt{\frac{r_G(u)+r_G(v)-2}{r_G(u)r_G(v)}}$$

$$= 18\sqrt{\frac{4+3-2}{4\times3}} + 36\sqrt{\frac{4+2-2}{4\times2}} + 410\sqrt{\frac{4+1-2}{4\times1}} + 301\sqrt{\frac{1+1-2}{1\times1}} + 126\sqrt{\frac{3+1-2}{3\times1}} + 126\sqrt{\frac{1+2-2}{1\times2}}$$

$$= 81\sqrt{2} + 205\sqrt{3} + 126\sqrt{\frac{2}{3}} + 9\sqrt{\frac{5}{3}}.$$

$$RR(G) = \sum_{uv\in E(G)} \frac{1}{\sqrt{r_G(u)r_G(v)}}$$

$$= 18\sqrt{\frac{1}{4\times3}} + 36\sqrt{\frac{1}{4\times2}} + 410\sqrt{\frac{1}{4\times1}} + 301\sqrt{\frac{1}{1\times1}} + 126\sqrt{\frac{1}{3\times1}} + 126\sqrt{\frac{1}{1\times2}}$$

$$= 506 + 72\sqrt{2} + 45\sqrt{3}\,.$$

$$SR(G) = \sum_{uv\in E(G)} \frac{1}{\sqrt{r_G(u)+r_G(v)}}$$

$$= 18\sqrt{\frac{1}{4+3}} + 36\sqrt{\frac{1}{4+2}} + 410\sqrt{\frac{1}{4+1}} + 301\sqrt{\frac{1}{1+1}} + 126\sqrt{\frac{1}{3+1}} + 126\sqrt{\frac{1}{1+2}}$$

$$= 63 + 301\sqrt{\frac{1}{2}} + 126\sqrt{\frac{1}{3}} + 410\sqrt{\frac{1}{5}} + 36\sqrt{\frac{1}{6}} + 18\sqrt{\frac{1}{7}}.$$

$$HMR(G) = \sum_{uv\in E(G)} \frac{2}{r_G(u)+r_G(v)}$$

$$= 18\left(\frac{2}{4+3}\right) + 36\left(\frac{2}{4+2}\right) + 410\left(\frac{2}{4+1}\right) + 301\left(\frac{2}{1+1}\right) + 126\left(\frac{2}{3+1}\right) + 126\left(\frac{2}{1+2}\right)$$

$$= \frac{4404}{7}.$$

$$GAR(G) = \sum_{uv\in E(G)} \frac{2\sqrt{r_G(u)r_G(v)}}{r_G(u)+r_G(v)}$$

$$= 18\left(\frac{2\sqrt{4\times3}}{4+3}\right) + 36\left(\frac{2\sqrt{4\times2}}{4+2}\right) + 410\left(\frac{2\sqrt{4\times1}}{4+1}\right) + 301\left(\frac{2\sqrt{1\times1}}{1+1}\right) + 126\left(\frac{2\sqrt{3\times1}}{3+1}\right) + 126\left(\frac{2\sqrt{1\times2}}{1+2}\right)$$

$$= 629 + 108\sqrt{2} + \frac{513}{7}\sqrt{3}\,.$$

Theorem 4.

Let G be the dendrimeric chelators contain nine hexadentate centers. Then the first Revan index $R_1(G)$, the second Revan index $R_2(G)$, the third Revan index $R_3(G)$, the forgotten Revan index $FR(G)$, the first Hyper Revan index $HR_1(G)$, the ABC Revan index $ABCR(G)$, the Randi`c Revan index $RR(G)$, the Sum connectivity Revan index $SR(G)$, the Harmonic Revan index $HMR(G)$, and the Geometric-Arithmatic Revan index $GAR(G)$ for G are

i. $R_1(G) = 5790,$

ii. $R_2(G) = 4596,$

iii. $R_3(G) = 2538,$

iv. $FR(G) = 15888,$

v. $HR_1(G) = 25080,$

vi. $ABCR(G) = 243\sqrt{\frac{1}{2}} + 306\sqrt{3} + 189\sqrt{\frac{2}{3}} + \frac{27}{2}\sqrt{\frac{5}{3}}$,

vii. $RR(G) = 753 + 108\sqrt{2} + \frac{405}{2}\sqrt{\frac{1}{3}}$,

viii. $SR(G) = \frac{189}{2} + 447\sqrt{\frac{1}{2}} + 189\sqrt{\frac{1}{3}} + 612\sqrt{\frac{1}{5}}$

$+ 54\sqrt{\frac{1}{6}} + 27\sqrt{\frac{1}{7}}$,

ix. $HMR(G) = \frac{65661}{70}$, and

x. $GAR(G) = \frac{4683}{5} + 162\sqrt{2} + \frac{1539}{14}\sqrt{3}$.

Proof:

From the Figure 4, it is easy to see that the vertices of G have degree 1, 2, 3, or 4. Thus $\Delta(G) = 4$, $\delta(G) = 1$, and therefore $r_G(u) = 4 + 1 - d_G(u)$ is the Revan degree of any vertex u. According to the Revan vertex degree, there are six types of Revan edge sets in the structure of the dendrimeric chelators contain nine hexadentate centres G shown in Figure 4, as follows:

$E_1 = \{uv \in E(G) | r_G(u) = 4, r_G(v) = 3\}$, $|E_1| = 27$

$E_2 = \{uv \in E(G) | r_G(u) = 4, r_G(v) = 2\}$, $|E_2| = 54$

$E_3 = \{uv \in E(G) | r_G(u) = 4, r_G(v) = 1\}$, $|E_3| = 612$

$E_4 = \{uv \in E(G) | r_G(u) = r_G(v) = 1\}$, $|E_4| = 447$

$E_5 = \{uv \in E(G) | r_G(u) = 3, r_G(v) = 1\}$, $|E_5| = 189$

$E_6 = \{uv \in E(G) | r_G(u) = 1, r_G(v) = 2\}$, $|E_6| = 189$

By using the above Revan edge partitions,

$R_1(G) = \sum_{uv \in E(G)} r_G(u) + r_G(v)$

$= 27(4 + 3) + 54(4 + 2) + 612(4 + 1) +$

$447(1 + 1) + 189(3 + 1) + 189(1 + 2)$

$= 5790.$

$R_2(G) = \sum_{uv \in E(G)} r_G(u)\, r_G(v)$

$= 27(4 \times 3) + 54(4 \times 2) + 612(4 \times 1) +$

$447(1 \times 1) + 189(3 \times 1) + 189(1 \times 2)$

$= 4596.$

$R_3(G) = \sum_{uv \in E(G)} |r_G(u) - r_G(v)|$

$= 27|4 - 3| + 54|4 - 2| + 612|4 - 1| +$

$447|1 - 1| + 189|3 - 1| + 189|1 - 2|$

$= 2538.$

$FR(G) = \sum_{uv \in E(G)} r_G(u)^2 + r_G(v)^2$

$= 27(4^2 + 3^2) + 54(4^2 + 2^2) +$

$612(4^2 + 1^2) + 447(1^2 + 1^2) +$

$189(3^2 + 1^2) + 189(1^2 + 2^2)$

$= 15888.$

$HR_1(G) = \sum_{uv \in E(G)} (r_G(u) + r_G(v))^2$

$= 27 + 15(4 + 2)^2 + 54(4 + 1)^2 + 612(1 + 1)^2 + 447(3 + 1)^2 + 189(1 + 2)^2$

$= 25080.$

$ABCR(G) = \sum_{uv \in E(G)} \sqrt{\frac{r_G(u) + r_G(v) - 2}{r_G(u) r_G(v)}}$

$= 27\sqrt{\frac{4+3-2}{4\times3}} + 54\sqrt{\frac{4+2-2}{4\times2}} + 612\sqrt{\frac{4+1-2}{4\times1}} + 447\sqrt{\frac{1+1-2}{1\times1}} + 189\sqrt{\frac{3+1-2}{3\times1}} + 189\sqrt{\frac{1+2-2}{1\times2}}$

$= 243\sqrt{\frac{1}{2}} + 306\sqrt{3} + 189\sqrt{\frac{2}{3}} + \frac{27}{2}\sqrt{\frac{5}{3}}.$

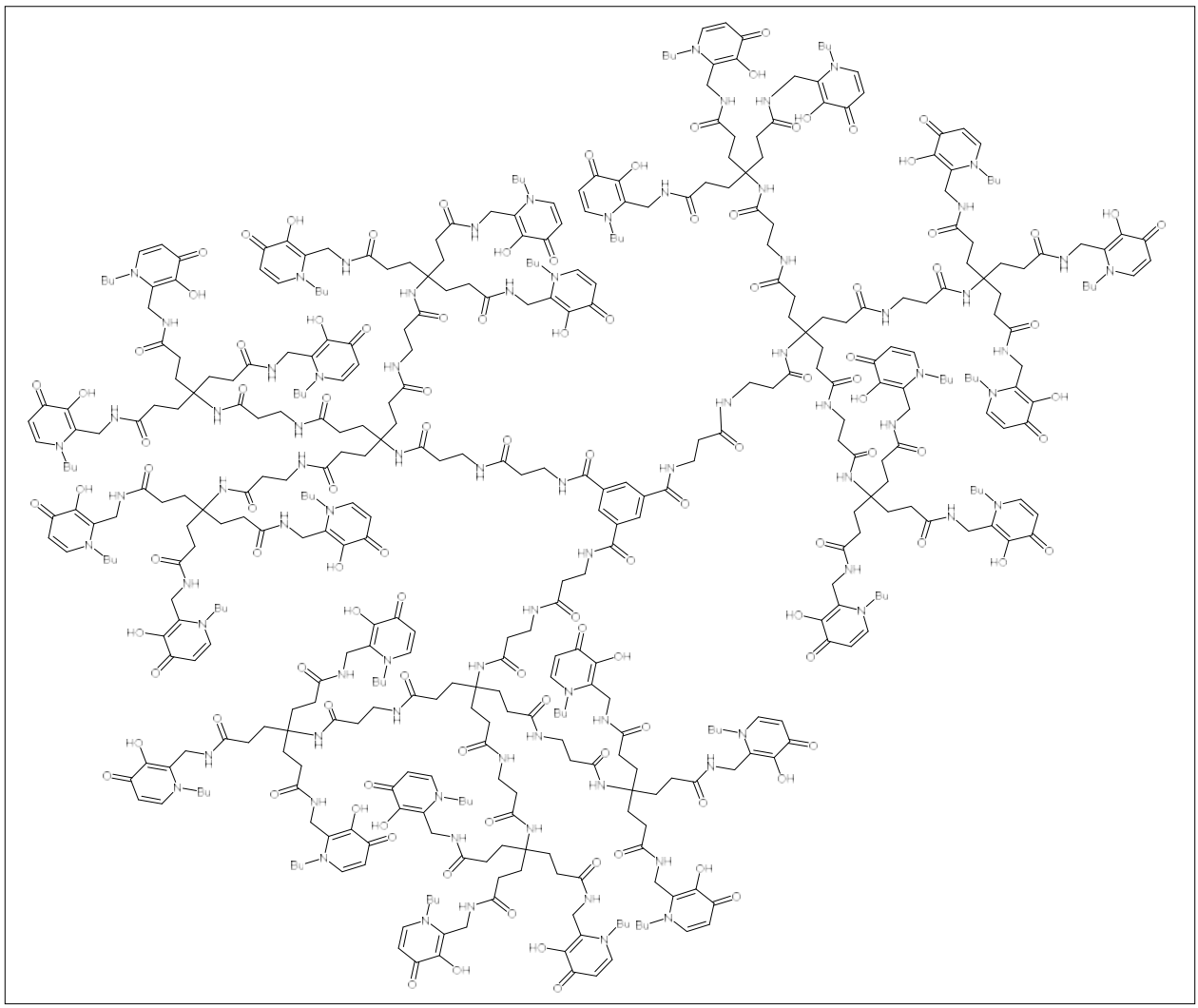


Figure 4: Second-generation dendrimeric chelator corresponding to compound 13 in Scheme 4 of Zhou et al. (2018), containing nine hexadentate chelating ligands and twenty-seven 3-hydroxypyridinone units. Structure adapted from Zhou et al. (2018).

$$RR(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{r_G(u) r_G(v)}}$$

$$= 27\sqrt{\frac{1}{4\times 3}} + 54\sqrt{\frac{1}{4\times 2}} + 612\sqrt{\frac{1}{4\times 1}} + 447\sqrt{\frac{1}{1\times 1}} + 189\sqrt{\frac{1}{3\times 1}} + 189\sqrt{\frac{1}{1\times 2}}$$

$$= 753 + 108\sqrt{2} + \frac{405}{2}\sqrt{\frac{1}{3}}.$$

$$SR(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{r_G(u)+r_G(v)}}$$

$$= 27\sqrt{\frac{1}{4+3}} + 54\sqrt{\frac{1}{4+2}} + 612\sqrt{\frac{1}{4+1}} + 447\sqrt{\frac{1}{1+1}} + 189\sqrt{\frac{1}{3+1}} + 189\sqrt{\frac{1}{1+2}}$$

$$= \frac{189}{2} + 447\sqrt{\frac{1}{2}} + 189\sqrt{\frac{1}{3}} + 612\sqrt{\frac{1}{5}} + 54\sqrt{\frac{1}{6}} + 27\sqrt{\frac{1}{7}}.$$

$$HMR(G) = \sum_{uv \in E(G)} \frac{2}{r_G(u)+r_G(v)}$$

$$= 27\left(\frac{2}{4+3}\right) + 54\left(\frac{2}{4+2}\right) + 612\left(\frac{2}{4+1}\right) + 447\left(\frac{2}{1+1}\right) + 189\left(\frac{2}{3+1}\right) + 189\left(\frac{2}{1+2}\right)$$

$$= \frac{65661}{70}.$$

$$GAR(G) = \sum_{uv \in E(G)} \frac{2\sqrt{r_G(u)r_G(v)}}{r_G(u)+r_G(v)}$$

$$= 27\left(\frac{2\sqrt{4\times3}}{4+3}\right) + 54\left(\frac{2\sqrt{4\times2}}{4+2}\right) + 612\left(\frac{2\sqrt{4\times1}}{4+1}\right) + 447\left(\frac{2\sqrt{1\times1}}{1+1}\right) + 189\left(\frac{2\sqrt{3\times1}}{3+1}\right) + 189\left(\frac{2\sqrt{1\times2}}{1+2}\right)$$

$$= \frac{4683}{5} + 162\sqrt{2} + \frac{1539}{14}\sqrt{3}.$$

CONCLUSION

In this work, four new hexadentate 3-hydroxypyridinone-terminated dendrimers were analyzed using different Revan indices. These indices, comprising the first, second, and third Revan indices, as well as the forgotten Revan index and others, offer important new perspectives on the topological properties of these dendrimeric structures. Using the Revan edge set partitions of the structures of dendrimers, these indices were calculated. The calculated indices reveal a strong correlation between the dendrimeric structures and its potential applications, thereby underscoring the utility of Revan indices in the study and development of advanced molecular systems. The results of this study, the computed values of Revan indices differ with various dendrimeric structures, and they assist in recognizing those structures' complexities. This study can be extended to study the physicochemical and biological properties of dendrimeric structures using the QSPR (quantitative structure-property relationship) and QSAR (quantitative structure- activity relationship) models as a future study.

REFERENCES

Abbasi, E., Aval, S., Akbarzadeh, A., Milani, M., Nasrabadi, H., Joo, S., Hanifehpour, Y., Nejati-Koshki, K., & Pashaei-Asl, R. (2014). Dendrimers: synthesis, applications, and properties. *Nanoscale Research Letters*, *9*(1), 247. https://doi.org/10.1186/1556-276x-9-247

Babujee, J.B., Ramakrishnan, S., & Senbagamalar, J. (2013). Topological Indices of Molecular Graphs Under Specific Chemical Reactions. *International Journal of Computing Algorithm*, *2*(1), 68–74. https://doi.org/10.20894/ijcoa.101.002.001.019

Buhleier, E., Vogtle, F., & Wehner, W. (1978). "Cascade"- and "Nonskid-Chain-like" Syntheses of Molecular Cavity Topologies. *Synthesis*, *1978*(02), 155–158. https://doi.org/10.1055/s-1978-24702

Zhou, T., Chen, K., Kong, L.-M., Liu, M.-S., Ma, Y.-M., Xie, Y.-Y., & Hider, R. C. (2018). Synthesis, iron binding and antimicrobial properties of hexadentate 3-hydroxypyridinones-terminated dendrimers. *Bioorganic & Medicinal Chemistry Letters*, 28(14), 2504–2512. https://doi.org/10.1016/j.bmcl.2018.05.058

Estrada, E., Torres, L., Rodriguez, L., & Gutman, I. (1998). An atom-bond connectivity index: modelling the enthalpy

of formation of alkanes. *Indian Journal of Chemistry-Section A*, 37(10), 849–855. https://nopr.niscpr.res.in/bitstream/123456789/40308/1/IJCA%2037A(10)%20849-855.pdf

Favaron, O., Mahéo, M., & Saclé, J. F. (1993). Some eigenvalue properties in graphs (conjectures of Graffiti — II). *Discrete Mathematics*, *111*(1-3), 197–220. https://doi.org/10.1016/0012-365x(93)90156-n

Furtula, B., & Gutman, I. (2015). A forgotten topological index. *Journal of Mathematical Chemistry*, *53*(4), 1184–1190. https://doi.org/10.1007/s10910-015-0480-z

Vukičević, D., & Furtula, B. (2009). Topological index based on the ratios of geometrical and arithmetical means of end-vertex degrees of edges. *Journal of Mathematical Chemistry*, 46(4), 1369–1376. https://doi.org/10.1007/s10910-009-9520-x

Randic, M. (1975). Characterization of molecular branching. *Journal of the American Chemical Society*, *97*(23), 6609–6615. https://doi.org/10.1021/ja00856a001

Gutman, I. (2013). Degree-Based Topological Indices. *Croatica Chemica Acta*, *86*(4), 351–361. https://doi.org/10.5562/cca2294

Gutman, I., & Trinajstić, N. (1972). Graph theory and molecular orbitals. Total φ-electron energy of alternant hydrocarbons. *Chemical Physics Letters*, *17*(4), 535–538. https://doi.org/10.1016/0009-2614(72)85099-1

Zhou, B., & Nenad Trinajstić. (2009). On a novel connectivity index. *Journal of Mathematical Chemistry*, *46*(4), 1252–1270. https://doi.org/10.1007/s10910-008-9515-z

Kulli, V. R. (2017a). On the Product Connectivity Revan Index of Certain Nanotubes. *Journal of Computer and Mathematical Sciences*, *8*(10), 562–567. https://doi.org/10.29055/jcms/694

Kulli, V. R. (2017b). Revan Indices of Oxide and Honeycomb Networks. *International Journal of Mathematics and Its Applications*, 5(4 - E), 663–667. https://ijmaa.in/index.php/ijmaa/article/view/1324

Kulli, V. R. (2017c). The Sum Connectivity Revan Index of Silicate and Hexagonal Networks. *Annals of Pure and Applied Mathematics*, 14(3), 401–406. https://doi.org/10.22457/apam.v14n3a6

Kulli, V.R. (2018a). Connectivity revan indices of chemical structures in drugs. *International Journal of Engineering Science Research Technology,* 7(5), 11-16. doi: 10.5281/zenodo.1241360

Kulli, V.R. (2018b). F-Revan index and F-Revan polynomial of some families of benzenoid systems. *Journal of Global Research in Mathematical Archives (JGRMA)*, *5*(11), 1–6.

Kulli V.R. (2018c). Hyper Revan indices and their polynomials of silicate networks. *International Journal of Current Research in Science and Technology,* 4(3), 17–21.

Kulli V.R. (2020). Revan indices of chloroquine, hydroxychloroquine, remdesivir: Research advances for the treatment of covid-19. *International Journal of Engineering Sciences and Research Technology* 9(5), 73–84. DOI: 10.5281/zenodo.3828871

Wiener, H. (1947). Structural determination of paraffin boiling points. *Journal of the American Chemical Society*, *69*(1), 17–20. https://doi.org/10.1021/ja01193a005

Gunawardhana, D. C., Lanel, G. H. J., Perera, K. K. K. R., & Gunaratna, A. G. M. J. (2024). On novel approach for computing distance-based indices of anti-tuberculosis drugs. *arXiv preprint* arXiv:2411.02416.

J.Natn.Sci.Foundation Sri Lanka 2026 54 (1): 17 - 30
DOI: http://dx.doi.org/10.4038/jnsfsr.v54i1.12522

RESEARCH ARTICLE

Statistics

Versatility in control theory and real-world applications

R Jabeen[1*], FM Alghamdi[2], HM Aljohani[3] and A Zaka[4]
[1] *Department of Statistics, COMSATS University Islamabad, Lahore Campus, Lahore, Pakistan.*
[2] *Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia.*
[3] *Department of Mathematics and Statistics, College of Science, Taif University, Taif 21944, Saudi Arabia.*
[4] *Department of Statistics, Government Graduate College of Science, Wahdat Road, Lahore, Pakistan.*

Submitted: 17 March 2025; Revised: 28 January 2026; Accepted: 18 February 2026

Abstract: The inspiration of this manuscript is reflected power function (RPF) and size biased distributions. We present a size-biased reflected power function (SBRPF) distribution using highly flexible reflected power function distribution. The resulting distribution is also flexible enough to fit all types of data either J shaped, reverse J shaped, positive skewed and negative skewed. We also drive various important properties of the suggested model. We show the comprehensive analysis of the proposed model detailing the asymptotic behavior of the function. We use diverse methods of estimation such as modified maximum likelihood method (MMLM) , maximum likelihood method (MLE) , percentile estimator (PE). The numerical analysis shows that the SBRF distribution remains consistent while mean square error (MSE) does not decrease as sample size increases. The real life data sets demonstrate that SBRPF distribution is a better choice to be compared with other models exist in the literature. We also reported the use of two control charts Exponential weighted moving averages (EWMA) and Extended Exponentially weighted moving averages (EEWMA) for the shape parameter of the proposed model. From both Simulation studies and real life application, we observe that EEWMA is a more effective control chart for detecting early change during a process for SBRPF distribution.

Keywords: Failure rate components, power function distribution, reflected distribution, size-biased distribution.

INTRODUCTION

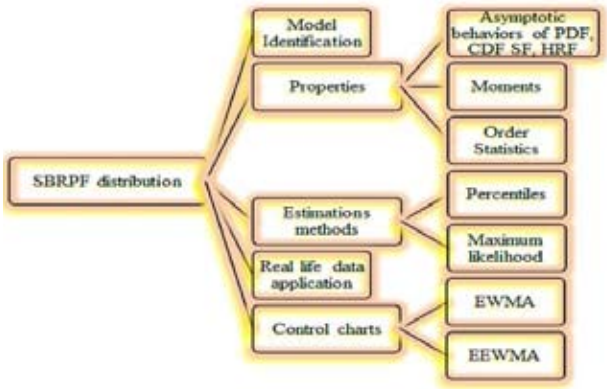
In engineering and related fields, researchers typically aim to assess the reliability of a module over a specified period under defined conditions. This need urges the new researchers to employ some new flexible distributions to model the lifetime data of different engineering processes. Probability distributions are used to draw meaningful inferences from real life problems and analyze the variations in the data. The probability distributions basically represent the variability in the random variables. Due to the continuous evolve in technology, the complexity of data increases and the existing distribution falls short of capturing all trends in data. Keeping in view these circumstances, efforts are made to produce new distributions, and some modifications are proposed to the existing distributions. These efforts include generalization, compounding and mixing the already existing distributions.

In this research, we incorporate reflected power and size-biased distributions to develop a unique distribution. Basically, reflected power function (RPF) distribution [1] is given as a modification of power function (PF) distribution. This model is famous for its flexibility in modeling lifetime data and good for fitting certain failure data sets. As claimed in [2], the PF distribution is interconnected to the Pareto distribution. The authors of [3] explore the PF distribution as a preferable choice over others for assessing the durability of semiconductor devices and products.

The one particular case for weighted distribution is named as the size-biased distribution. It was first

* Corresponding author (drriffatjabeen@cuilahore.edu.pk; https://orcid.org/0000-0003-1461-0611)

introduced by [4] in the context of bias modeling and later expanded into a unified theory by [5]. In practice, this type of distribution often arises when sample observations are recorded with unequal probabilities, such as in probability proportionate to size (PPS) sampling designs.



The size-biased distribution can be obtained using

$$f_s(x) = \frac{x^{\alpha} f(x)}{E[x^{\alpha}]},$$

where $f(x)$ is the baseline distribution, $w(x)$ is weight function. It is compulsory that the term $E[x^{\alpha}]$ exist.

A particularly interesting case occurs as the weight function becomes $w(x) =$ which is known as sized-biased distribution. For more details, please refer to [6]–[8].

The application of weighted distributions is discussed details by [9]. Subsequently, [10] analyzed HPS diameter increment data and applied weighted distributions. More recently, [11] utilized weight distributions to model a stochastic population with predictive effects, employing a stochastic differential equation to compute the gamma distribution as a weight for the stationary probability density function (PDF).

Taillie et al. [12] applied weighted distributions to sample fish stock populations. Also, Gove and Patil [13] used the same theory for the quadratic relationship between diameter and basal area. In both above literature, the said distributions served as straightforward model for the observed data.

The aim of this research is to apply size-biased distribution theory to real-world data. For this we propose the size-biased reflected power function (SBRPF) distribution as a modification to power function distribution (PFD) introduced in [14].

The SBRPF distribution is highly versatile due to its straightforward formulation, making it applicable in reliability engineering, biosciences, and forestry. The parameters of the proposed distribution will be estimated using MLE, MMLM, and MPE. Additionally, we will demonstrate the practical application of the SBRPF distribution.

The weighted distributions, introduced by [4] and further developed by [6], includes size-biased and area-biased distributions as special cases. These distributions naturally arise in practice when sample observations are recorded with unequal probabilities Additionally, the transmuted weighted exponential [15] and other weighted distributions are given by [16].

Control charts have become crucial tools for monitoring product quality across various sectors, particularly in agriculture and industry. It was [17] who introduced. The exponentially weighted moving average (EWMA). Then [18] provided modified EWMA (MEWMA) control chart for monitoring the scale parameter of Weibull distribution. The EWMA was modified by [19] and named as exponentially EWMA (EEWMA) control chart.

A new EWMA control chart, based on the log transformation of sample variance, was discussed in [20]. This new chart was compared with several conventional charts. Additionally, [21]-[26] explored new additional monitoring charts

Model identification

The power function (PF) distribution is defined as

$$f_{PF}(y) = \frac{\gamma y^{\gamma-1}}{\beta^{\gamma}}, \qquad 0 < y < \beta. \qquad ...(1)$$

where γ and β is shape and scale parameter.

The RPF distribution is defined as

$$f_{RPF}(x) = \frac{\gamma(\theta-x)^{\gamma-1}}{\beta^{\gamma}}, \quad \theta-\beta < x < \theta. \qquad ...(2)$$

where γ and β is shape and scale parameter. Also θ is reflecting parameter.

The PDF and cumulative distribution function (CDF) of the SBRPF distribution is given as

$$f_{SBRPF}(x) = \frac{x\gamma(\theta-x)^{\gamma-1}(\gamma+1)}{\beta^{\gamma}\{\theta(\gamma+1)-\beta\gamma\}}, \qquad \theta-\beta < x < \theta. \quad ...(3)$$

$$F_{SBRPF}(x) = \frac{(\theta-x)^{\gamma}(x\gamma+\theta)+\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}{\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}. \quad ...(4)$$

The survival function (SF), hazard rate function (HRF) and quantile function (QF) for SBRPF distribution are given by:

$$S_{SBRPF}(x) = 1 - \frac{(\theta-x)^{\gamma}(x\gamma+\theta)+\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}{\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}. \quad ...(5)$$

$$h_{SBRPF}(x) = \frac{x\gamma(\gamma+1)}{(\theta-x)(x\gamma+\theta)} \quad ...(6)$$

and

$$x_p = [\theta^2 - \beta\{\beta - 2\theta\}(p_i - 1)]^{1/2}. \quad ...(7)$$

Note that x_p can be used to generate the SBRPF random values.

Special cases

We reported two special cases of our proposed distribution which are given below

1. Multiplying SBRPF distribution by the quantity $\frac{\theta(\gamma+1)-\beta\gamma}{x(\gamma+1)}$, gives the density function of the RPF distribution.
2. Multiplying SBRPF distribution by $\frac{\theta(\gamma+1)-\beta\gamma}{x(\gamma+1)}$ and then by putting $x = \theta - y$, we obtain the PF distribution.

Two special cases are presented in Figure 1, which provides the loop of both cases.

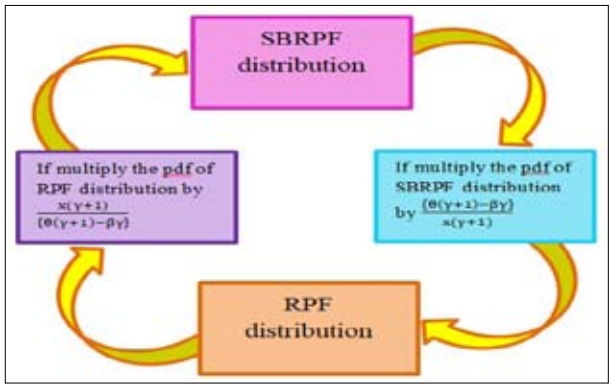


Figure 1: Probability distributions loop plot.

Figure 1 shows probability distributions loop plot which shows that we can get SBRPF from RPF distribution and similarly we get RPF distribution from SBRPF model as a special case by multiplying with $\frac{(\theta(\gamma+1)-\beta\gamma)}{x(\gamma+1)}$.

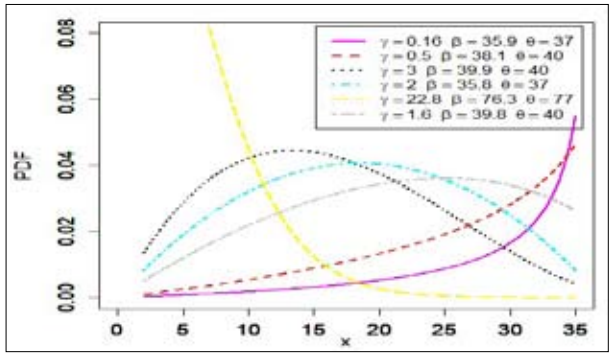


Figure 2: Probability density function of the SBRPF distribution.

Figures 2 and 3 present the plots of the PDF of the SBRPF distribution. The PDF plots demonstrate that this proposed distribution can effectively fit a wide variety of data types, including positively skewed, negatively skewed, J-shaped, reversed-J-shaped, and symmetric data.

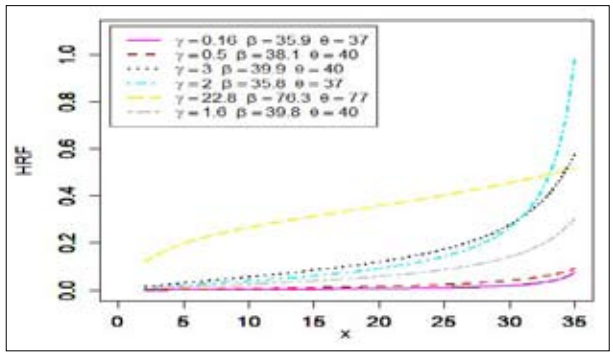


Figure 3: Plots for the HRF of the SBRPF distribution.

Figure 3 presents the plots of the HRF of the SBRPF distribution. The HRF plot demonstrates that this proposed distribution can effectively fit a wide variety of data types, including positively skewed, negatively skewed, J-shaped, reversed-J-shaped, and symmetric data.

Properties of SBRPF distribution

In this section various properties of the SBRPF distribution are discussed.

Limiting behaviors of SBRPF distribution

The limiting behaviors of the SBRPF distribution, for $x \to 0$ and $x \to \infty$, are given by

1. $\lim_{x\to 0} f_{SBRPF}(\mathrm{x}) = \lim_{x\to 0} \frac{\mathrm{x}\gamma(\theta-\mathrm{x})^{\gamma-1}(\gamma+1)}{\beta^{\gamma}\{\theta(\gamma+1)-\beta\gamma\}}.$

2. $\lim_{x\to \infty} f_{SBRPF}(\mathrm{x}) = \lim_{x\to \infty} \frac{\mathrm{x}\gamma(\theta-\mathrm{x})^{\gamma-1}(\gamma+1)}{\beta^{\gamma}\{\theta(\gamma+1)-\beta\gamma\}} = \infty.$

3. $\lim_{x\to 0} F_{SBRPF}(\mathrm{x}) = \lim_{x\to 0} \frac{(\theta-\mathrm{x})^{\gamma}(\mathrm{x}\gamma+\theta)+\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}{\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}} = 0,$ for $\gamma = 1, \beta = 0, \theta = 0.$

4. $\lim_{x\to \infty} F_{SBRPF}(\mathrm{x}) = \lim_{x\to \infty} \frac{(\theta-\mathrm{x})^{\gamma}(\mathrm{x}\gamma+\theta)+\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}{\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}} = 1,$ for $\gamma = 0, \beta = 1, \theta = 0.$

5. $\lim_{x\to 0} S_{SBRPF}(\mathrm{x}) = \lim_{x\to 0} 1 - \frac{(\theta-\mathrm{x})^{\gamma}(\mathrm{x}\gamma+\theta)+\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}{\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}} = 1,$ for $\gamma = 1, \beta = 0, \theta = 0.$

6. $\lim_{x\to \infty} S_{SBRPF}(\mathrm{x}) = \lim_{x\to \infty} 1 - \frac{(\theta-\mathrm{x})^{\gamma}(\mathrm{x}\gamma+\theta)+\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}{\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}} = 0,$ for $\gamma = 0, \beta = 1, \theta = 0.$

7. $\lim_{x\to 0} h_{SBRPF}(\mathrm{x}) = \lim_{x\to 0} \frac{\mathrm{x}\gamma(\gamma+1)}{(\theta-\mathrm{x})(\mathrm{x}\gamma+\theta)} = 0.$

8. $\lim_{x\to \infty} h_{SBRPF}(\mathrm{x}) = \lim_{x\to \infty} \frac{\mathrm{x}\gamma(\gamma+1)}{(\theta-\mathrm{x})(\mathrm{x}\gamma+\theta)} = 0,$ for $\gamma = 0, \theta = 0.$

Moments

The r^{th} moment about zero of SBRPF is below as

$$\mu_r = \frac{v(v+1)\theta^{r+v+1}}{\beta^{v}(\theta(v+1)-v\beta)} \sum_{j=0}^{\infty} \frac{(-1)^j(\Gamma v)\left(1-\left(\frac{\theta-\beta}{\theta}\right)^{(r+j+2)}\right)}{(\Gamma v-j)j!(r+j+2)} \quad \ldots(8)$$

The moment generating function defines the characteristics of a random variable.

$$M_o(t) = \frac{v(v+1)}{\beta^{v}\{\theta(v+1)-\beta v\}} \sum_{r=0}^{\infty} \frac{t^r}{r!} \theta^{v+r+1} \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma v}{(\Gamma v-j)j!} \left\{ \frac{1-\left(\frac{\theta-\beta}{\theta}\right)^{r+j+2}}{r+j+2} \right\} \quad \ldots(9)$$

Order statistics

The density function of j^{th}, largest and smallest order statistics of the SBRPF distribution is obtained as

$$g_{(j)}(x) = \frac{n!}{(j-1)!\,(n-j)!}\left[\frac{x\gamma(\theta-x)^{\gamma-1}(\gamma+1)}{\beta^{\gamma}\{\theta(\gamma+1)-\beta\gamma\}}\right]\left[\frac{(\theta-x)^{\gamma}(x\gamma+\theta)+\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}{\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}\right]^{j-1}$$
$$\left[1-\frac{(\theta-x)^{\gamma}(x\gamma+\theta)+\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}{\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}\right]^{n-j}. \quad ...(10)$$

$$g_{(n)}(x) = n\left[\frac{x\gamma(\theta-x)^{\gamma-1}(\gamma+1)}{\beta^{\gamma}\{\theta(\gamma+1)-\beta\gamma\}}\right]\left[\frac{(\theta-x)^{\gamma}(x\gamma+\theta)+\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}{\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}\right]^{n-1}. \quad ...(11)$$

$$g_{(1)}(x) = n\left[\frac{x\gamma(\theta-x)^{\gamma-1}(\gamma+1)}{\beta^{\gamma}\{\theta(\gamma+1)-\beta\gamma\}}\right]\left[1-\frac{(\theta-x)^{\gamma}(x\gamma+\theta)+\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}{\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}\right]^{n-1}. \quad ...(12)$$

Parameters estimation of the SBRPF

The parameters of SBRPF are estimated with two PE and MLE. The results are obtained using the R programming language. These estimation methods have also been employed in studies [30]-[34].

Percentile estimators

The PE was given by [35]-[36]. Let $p_i = i/n + 1$ be an unbiased estimator of $F_{SBRPF}(x_{(i)};\gamma,\beta,\theta) = \frac{(\theta-x)^{\gamma}(x\gamma+\theta)+\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}{\beta^{\gamma}\{\beta\gamma-\theta(\gamma+1)\}}$, the PE of the parameters from SBRPF obtained as after minimizing the following function

$$P(\gamma,\beta,\theta) = \sum_{i=1}^{n}\left[x_{(i)} - [\theta^2 - \beta\{\beta-2\theta\}(p-1)]^{1/2}\right]^2, \quad ...(13)$$

with respect to γ, β and θ.

Maximum likelihood estimators

SBRPF model's log-likelihood function reduces to

$$\ell(x) = \sum_{i=1}^{n}\log(x_i) + n\log(\gamma) + (\gamma-1)\sum_{i=1}^{n}(\theta-x_i)$$
$$+ n\log(\gamma+1) - \log(\beta)^{n\gamma} - n\log(\theta) - n\log(\gamma+1)$$
$$+ n(\gamma+1)\log(\beta) + n\log(\gamma). \quad ...(14)$$

We can get MLE of the parameters by maximizing the log-likelihood function.

Numerical analysis

We generated random numbers using SBRPF distribution. We use different combinations of parameters to numerically compare PE and MLE. For three sample sizes: n = 150, 250 and 400 results are generated. We repeated the procedure for N = 2000, during which the SBRPF parameters are estimated for each parameter combination and sample using both the PCE and MLE. The average values (Avg) of the estimates and mean square errors (MSE) are reported for all samples.

Method validation

From Table 1, it is clear that all estimation methods exhibit the consistency; specifically, for all combinations of parameters, the MSE decreases as we increase the n.

Real-life applications

The actual dataset consists of failure times for 20 components [37] is considered to compare the performance of proposed distributions. We compared the proposed model with the reflected power function (RPF) [1] and the size-biased Mukherjee-Islam (SBMI) distribution [38]. The Kolmogorov-Smirnov (KS) test and its p-value, as well as the Hannan-Quinn information

criterion (HQIC), Akaike information criterion (AIC), consistent AIC (CAIC), negative log-likelihood function (-log($\hat{\theta}$)) and Bayesian information criterion (BIC),were used to make the comparison.

Table 1: Average and MSE of the SBRPF parameters for various sample sizes.

Methods	n	Parameters			Avg of estimates			MSE		
		γ	β	θ	Avg$(\hat{\gamma})$	Avg$(\hat{\beta})$	Avg$(\hat{\theta})$	MSE$(\hat{\gamma})$	MSE$(\hat{\beta})$	MSE$(\hat{\theta})$
PCE	150	1	2.1	5	7.00012	2.09933	4.99971	36.00144	0.00301	0.00136
	250				1.00952	2.09973	5.00122	0.00009	0.00180	0.00088
	400				1.00021	2.10086	4.99962	0.00000	0.00112	0.00055
	150	1	1.9	4	5.43512	1.90188	4.00093	19.67029	0.00264	0.00113
	250				1.00323	1.89998	3.99930	0.00001	0.00148	0.00068
	400				1.00008	1.90164	4.00026	0.00000	0.00091	0.00042
	150	1	1.5	2	1.74984	1.50452	1.99892	0.56226	0.00249	0.00056
	250				1.00027	1.50495	1.99942	0.00000	0.00146	0.00034
	400				1.00001	1.50107	1.99941	0.00000	0.00082	0.00022
MLE	150	1	2.1	5	18.95848	3.72082	4.99939	322.50761	2.62707	0.00060
	250				2.46269	2.87735	4.99939	2.13946	0.60427	0.00016
	400				1.00376	2.10808	4.99930	0.00001	0.00007	0.00004
	150	1	1.9	4	15.08033	3.08278	3.99940	198.25580	1.39896	0.00037
	250				2.51481	2.69164	3.99934	2.29466	0.62670	0.00011
	400				1.00339	1.90719	3.99926	0.00001	0.00005	0.00003
	150	1	1.5	2	3.87385	1.93599	1.99934	8.25931	0.19009	0.00013
	250				1.00378	1.50711	1.99928	0.00001	0.00005	0.00003
	400				1.00116	1.50235	1.99935	0.00000	0.00001	0.00001

Table 2: Descriptive statistics summary of real data.

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
0.072	4.471	8.662	8.429	11.648	19.809

Table 2 presents a concise summary of the statistical data. It provides us with minimum and maximum values in the data. Also provide us mean, median and quartile values of the data.

Table 3: The MLEs, and goodness-of-fit measures for failure time of 20 components.

Model	MLEs	KS	p-value	AIC	BIC	CAIC	HQIC	$-\log(\hat{\theta})$
SBRPF	2.4764 19.7978 20.4586	0.14846	0.716	135.0122	137.9994	136.5122	135.5953	64.5061
RPF	3.4010 19.7904 19.9185	0.45654	0.0002	154.0047	156.9919	155.5047	154.5879	74.00236
SBMI	8.958771 19.9749	0.87201	0.0000	498.8489	500.8404	499.5548	499.2376	247.4244

Table 3 reports the MLE and all other measures for comparison among the distributions. The data in this table indicates that the SBRPF is superior for the studied data compared to the competing models.

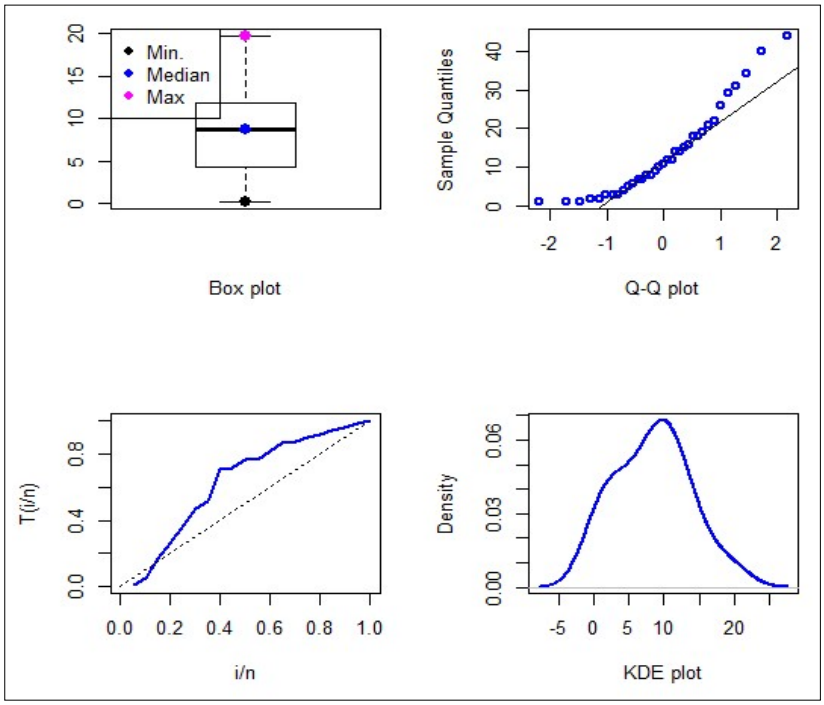


Figure 6: Plots of the real dataset.

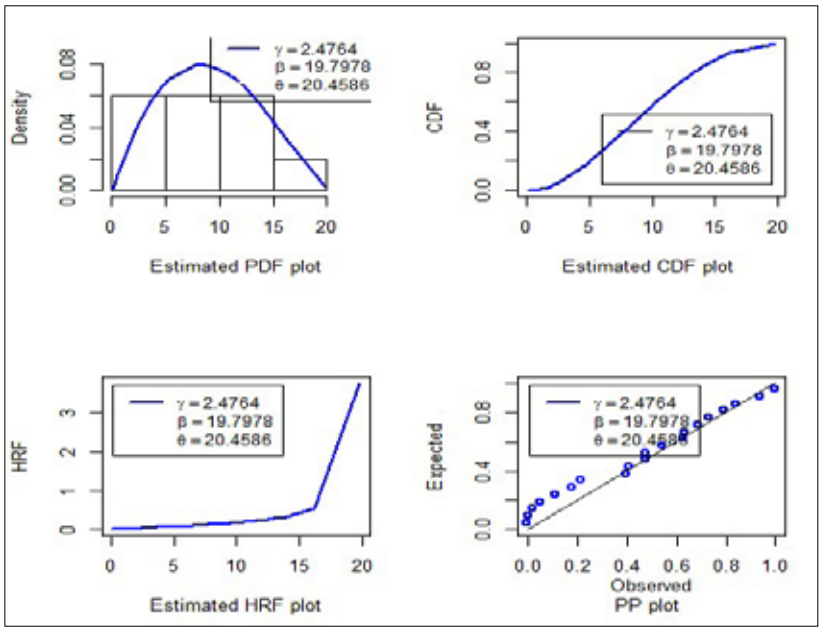


Figure 7: Some plots of the real dataset.

To assess the nature and behavior of the data, we employed several visualizations, including Box plot, Q-Q plots, total time in test (TTT) plot, and Kernal density estimation (KDE) plots for the real dataset, as shown in Figure 6. Using the same MLEs obtained for the goodness-of-fit measures, in Figure 7, some important plots are given. Additionally, the estimated density plot and HRF plots for the real dataset are illustrated in Figure 8. Figure 9 compares the AIC, BIC, CAIC, and HQIC values for the real dataset. These plots collectively demonstrate the advantage of the suggested distribution for the real-life dataset.

From Figures 6-9, it is clear that the SBRPF distribution is more appropriate to use for this data and demonstrate its flexibility to fit all types of data sets.

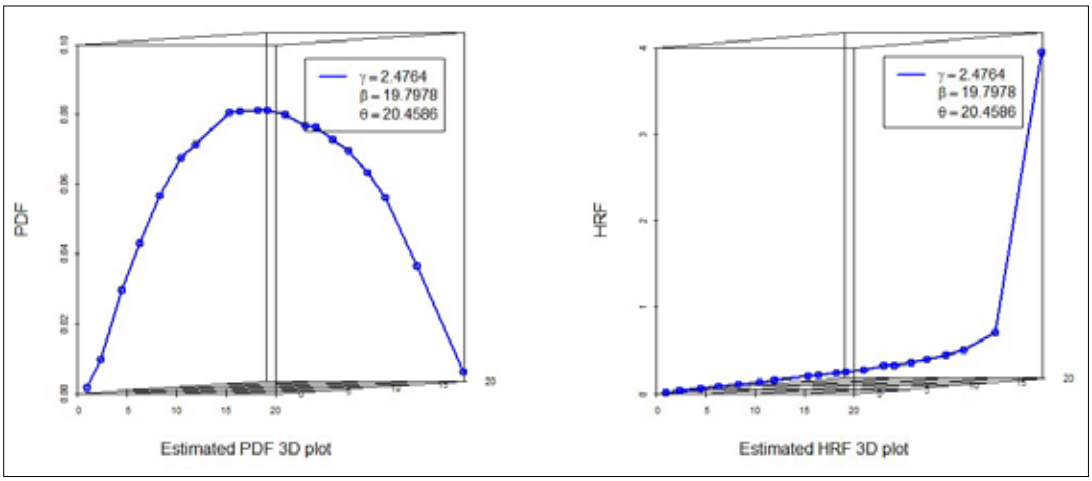


Figure 8: Estimated PDF (left panel) and estimated HRF (right panel).

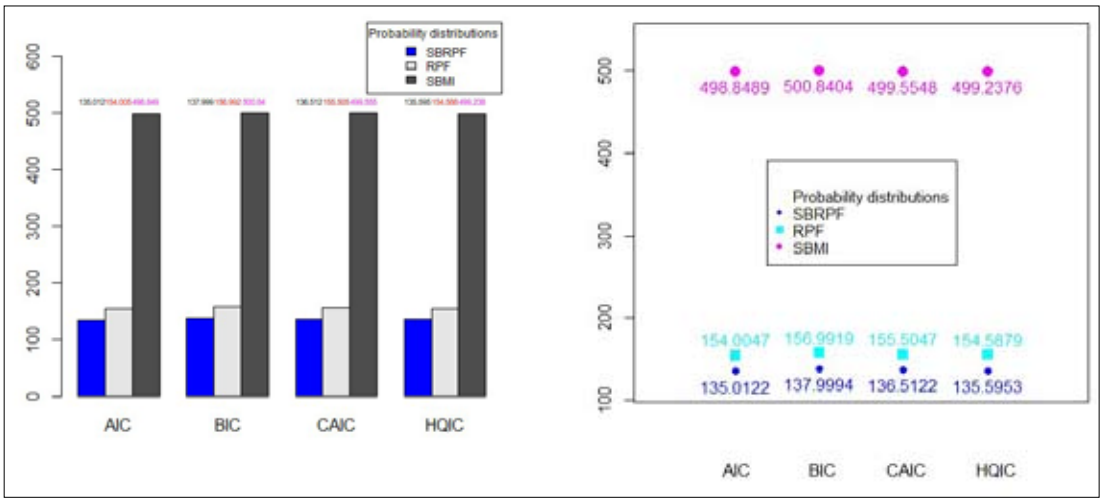


Figure 9: Evaluation of AIC, BIC, CAIC, HQIC for the real dataset.

Two proposed control monitoring

We introduce EWMA and EEWMA control charts, for MLEs of the SBRPF distribution. We utilized the R programming language to generate the results for these new control charts.

The EWMA control monitoring

The EWMA statistics, say $Z_{(t)}$, based on the MLEs of the SBRPF distribution is given by

$$A^*_{(t)} = \lambda^{**}\hat{\gamma}_{MLE(t)} , +(1-\lambda^{\&*})A^*_{(t-1)},$$

we observe $A^*_{(t)}$ as the EWMA statistic, and $A^*_{(t-1)}$ represents the previous value of the EWMA statistic. Also λ^{**} is smoothing constant and $\hat{\gamma}_{MLE(t)}$ is the maximum likelihood estimator of the shape parameter.

The EWMA control limits based on the MLEs of the SBRPF distribution are given by

$$UCL_{Z^*(t)} = \delta_0 + L\sqrt{Var(\hat{\gamma}_{MLE})\left[\frac{\lambda^{**}}{2-\lambda^{**}}\right][1-(1-\lambda^{**})^{2t}]}, \quad ...(15)$$

$$CL_{Z^*(t)} = \delta_0 \quad ...(16)$$

and

$$LCL_{Z^*(t)} = \delta_0 - L\sqrt{Var(\hat{\gamma}_{MLE})\left[\frac{\lambda^{**}}{2-\lambda^{**}}\right][1-(1-\lambda^{**})^{2t}]}. \quad ...(17)$$

The EEWMA control chart

The EEWMA statistics, say $E_{(t)}$, based on the MLE of the SBRPF distribution has the form $E_{(t)} = \varphi_1\hat{\gamma}_{MLE(t)} - \varphi_2\hat{\gamma}_{MLE(t-1)} + (1-\varphi_1+\varphi_2)\,E_{(t-1)}$, where $0 \le \varphi_1 \le 1,\ 0 \le \varphi_2 \le 1$,..

The EEWMA control limits based on the MLE of the SBRPF distribution are defined by

$$UCL_{E(t)} = \delta_0 + L\sqrt{Var(\hat{\gamma}_{MLE})\left[(\varphi_1{}^2+\varphi_2{}^2)*\left(\frac{1-(1-\varphi_1+\varphi_2)^{2i}}{1-(1-\varphi_1+\varphi_2)^2}\right)-2(1-\varphi_1+\varphi_2)\varphi_1\varphi_2\left(\frac{1-(1-\varphi_1+\varphi_2)^{2i-2}}{1-(1-\varphi_1+\varphi_2)^2}\right)\right]}, \quad ...(18)$$

$$CL_{E(t)} = \delta_0 \quad ...(19)$$

and

$$LCL_{E(t)} = \delta_0 - L\sqrt{Var(\hat{\gamma}_{MLE})\left[(\varphi_1{}^2+\varphi_2{}^2)*\left(\frac{1-(1-\varphi_1+\varphi_2)^{2i}}{1-(1-\varphi_1+\varphi_2)^2}\right)-2(1-\varphi_1+\varphi_2)\varphi_1\varphi_2\left(\frac{1-(1-\varphi_1+\varphi_2)^{2i-2}}{1-(1-\varphi_1+\varphi_2)^2}\right)\right]}. \quad ...(20)$$

Performance evaluation of the control monitoring

We have evaluated the control monitoring charts as follows:

1. Generate random data from the SBRPF distribution.
2. Compute $\hat{\gamma}_{MLE}$, for the data generated in step 1.
3. Repeat the first two steps 2,000 times to calculate ($\hat{\gamma}_{MLE}$).
4. Set control limits for the EWMA monitoring using step 3.
5. Fix ARL_0 for both proposed control charts, and for various shift values, compute ARL_1. Repeat similar steps for the EEWMA monitoring of the MLE.

The numerical results are reported in Table 4 and illustrated in Figure 11. The findings indicate that the EEWMA control monitoring based on $\hat{\gamma}_{MLE(t)}$, outperforms the EWMA control chart, as it detects shifts more quickly compared to the EWMA control monitoring.

Table 4: Evaluation indices for EWMA and the EEWMA monitoring of the MLE with $ARL_0 = 289$.

$ARL_0 = 289$		
Shifts	$EWMA \rightarrow ARL\ (\hat{Y}_{MLE(t)})$	$EEWMA \rightarrow ARL\ (\hat{Y}_{MLE(t)})$
0	289	289
0.05	289	276
0.1	288	74
0.12	284	74
0.14	281	74
0.6	239	17
0.7	237	17
1	196	17
1.3	172	12
1.6	153	10
2	91	10
4	1	1

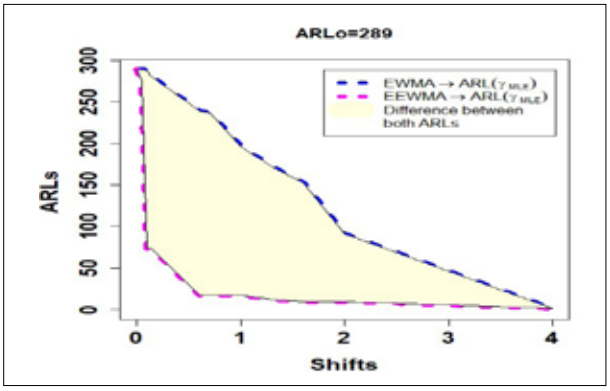


Figure 11: Comparison of EWMA and EEWMA control charts based on the MLE through ARLs with $ARL_0 = 289$.

Simulation results

A simulation study was conducted for the EWMA and EEWMA control charts based on the MLE of the SBRPF distribution. Initially, random observations were generated from the SBRPF distribution with parameters $(\gamma, \beta, \theta) = (1,2.1,5)$ to compute $\hat{Y}_{MLE}$. This process was repeated for 40 samples to obtain the EWMA statistics $Z_{(t)}$ and the control limits for the EWMA control chart. The $Z_{(t)}$ values were plotted against the subgroups (samples) in Figure 12.

The same procedure was followed for the EEWMA monitoring charts. The results are summarized in Table 5 and illustrated in Figure 12.

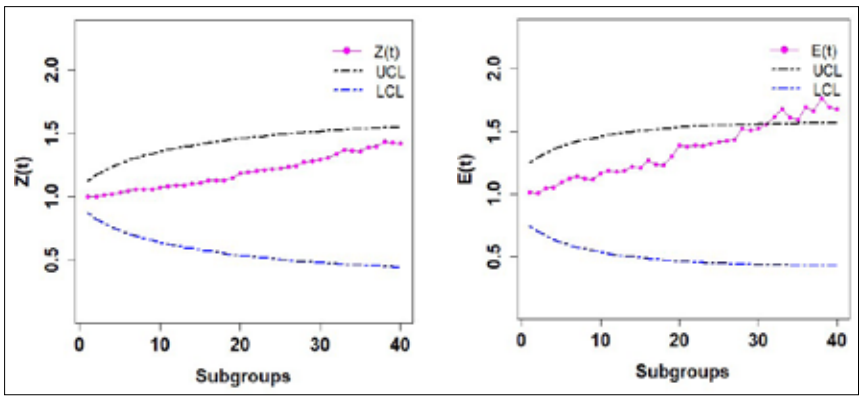


Figure 12: The EWMA and EEWMA monitoring of the MLE.

Table 5: Simulated results of the EWMA and EEWMA monitoring for the MLE.

EWMA $\lambda = 0.02$, L=10			EEWMA $\varphi_1 = 0.065$, $\varphi_1 = 0.02$, L = 6		
Z	UCL	LCL	E	UCL	LCL
1.004759	1.122616	0.877384	1.015465	1.250164	0.749836
1.004711	1.17168	0.82832	1.010166	1.29336	0.70664
1.016776	1.208186	0.791814	1.049179	1.327831	0.672169
1.022649	1.238034	0.761966	1.054982	1.356373	0.643627
1.037417	1.263536	0.736464	1.095768	1.380543	0.619457
1.050256	1.285894	0.714106	1.120398	1.401319	0.598681
1.061347	1.305832	0.694168	1.140704	1.419371	0.580629
1.060174	1.323827	0.676173	1.122451	1.435182	0.564818
1.059182	1.340215	0.659785	1.117574	1.449117	0.550883
1.075557	1.355243	0.644757	1.169138	1.46146	0.53854
1.087668	1.3691	0.6309	1.18824	1.472435	0.527565
1.090424	1.381935	0.618065	1.180801	1.482228	0.517772
1.094004	1.393869	0.606131	1.185671	1.490988	0.509012
1.106353	1.404999	0.595001	1.21817	1.498844	0.501156
1.109638	1.415408	0.584592	1.211713	1.505902	0.494098
1.130199	1.425164	0.574836	1.270725	1.512255	0.487745
1.128672	1.434328	0.565672	1.239288	1.51798	0.48202
1.127428	1.442951	0.557049	1.231763	1.523148	0.476852
1.149125	1.451077	0.548923	1.298802	1.527817	0.472183
1.185884	1.458746	0.541254	1.390271	1.532039	0.467961
1.196176	1.465992	0.534008	1.378498	1.535861	0.464139
1.206009	1.472847	0.527153	1.392166	1.539322	0.460678
1.209137	1.479338	0.520662	1.384317	1.542461	0.457539
1.217346	1.485491	0.514509	1.400047	1.545307	0.454693
1.227212	1.491327	0.508673	1.415848	1.54789	0.45211
1.236997	1.496868	0.503132	1.42949	1.550235	0.449765
1.244325	1.502131	0.497869	1.435054	1.552365	0.447635
1.27621	1.507135	0.492865	1.522916	1.554301	0.445699
1.284237	1.511895	0.488105	1.506657	1.55606	0.44394
1.294075	1.516424	0.483576	1.520755	1.55766	0.44234

While comparing EWMA and EEWMA, Figure 12 indicates that the process is out of control starting from observation 32, leading to the conclusion that corrective actions may be necessary. This detection is given by EEWMA, making it more powerful Control monitoring as compared to EWMA for SRPF distribution.

Real data application

Here, we extend the EWMA and EEWMA with real-life data of "Electronic Component Failure time," sourced from [39]. The resulting plots based on this data are shown in Figure 14.

Table 5: Simulated results of the EWMA and EEWMA monitoring for the MLE (continued).

EWMA λ = 0.02 , L=10			EEWMA λ_1 = 0.065, λ_2 = 0.02, L=6		
Z	UCL	LCL	E	UCL	LCL
1.309651	1.520738	0.479262	1.551534	1.559115	0.440885
1.338554	1.524847	0.475153	1.619321	1.560439	0.439561
1.368199	1.528763	0.471237	1.674705	1.561644	0.438356
1.360935	1.532497	0.467503	1.608254	1.56274	0.43726
1.35676	1.536059	0.463941	1.590673	1.563738	0.436262
1.390283	1.539458	0.460542	1.693191	1.564647	0.435353
1.394637	1.542702	0.457298	1.660857	1.565474	0.434526
1.430629	1.545799	0.454201	1.761584	1.566228	0.433772
1.425408	1.548758	0.451242	1.694449	1.566915	0.433085
1.420938	1.551584	0.448416	1.672931	1.56754	0.43246

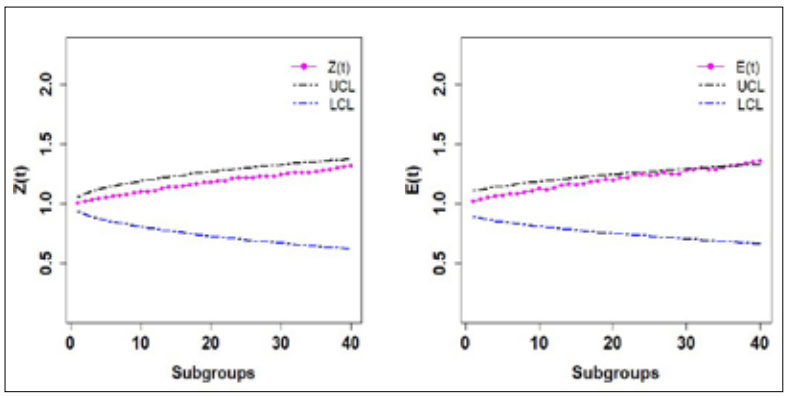


Figure 14: The EWMA and EEWMA monitoring of MLE for the real dataset.

In Figure 14, the EWMA and EEWMA monitoring of MLE are presented. In the right panel of Figure 14 that is the result of EEWMA, indicators show that the process is out of control from observation 36 onward. So once again EEWMA proved itself as best control monitoring to be used for SBRPF distribution.

CONCLUSION

In this research, we initially present the model identification for the proposed distribution. We investigate various properties of this distribution and discuss estimation methods, demonstrating the consistency of the mean squared error (MSE) for all estimators derived from both types of estimation methods. We also apply the proposed distribution to real data, providing a summary of the data, goodness-of-fit measures, and numerous plots. These results illustrate that our proposed distribution outperforms several existing distributions. Additionally, we implement EWMA and EEWMA control charts based on the maximum likelihood estimation (MLE) of the SBRPF distribution, highlighting its practical applications. From both Simulation studies and real life application, we observe that EEWMA is more effective control chart for detecting early change during a process for SBRPF distribution.

Funding

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R735), Princess Nourah bint Abdulrahman University, Riyadh,Saudi Arabia.

REFERENCES

Ahmad, M., Jabeen, R., Zaka, A., Hamdi, W. A. and Alnssyan, B. (2024) A unified generalized family of distributions: Properties, inference, and real-life applications. *AIP Advances 14*, 015043, https://doi.org/10.1063/5.0187188.

Abbasi, W. M., Ahmad, S., Perveen, S. and Rehman, T. (2017). Preliminary phytochemical analysis and invivo evaluation of antipyretic effects of hydro-methanolic extract of Cleome scaposa leaves. *Journal of Traditional and Complementry Medicine, 8*(1) (2017) 147-149. https://doi.org/10.1016/j.jtcme.2017.05.004.

Afify, A. Z., and Mohamed, O. A. (2020). A new three-parameter exponential distribution with variable shapes for the hazard rate: estimation and applications. *mathematics, 8*(1), 135. https://doi.org/10.3390/math8010135.

Brereton, C. F., and Jagals, P. (2021). applications of systems science to understand and manage multiple influences within children's environmental health in least developed countries: a causal loop diagram approach. *International Journal of Environmental Research and Public Health, 18*(6), 3010. https://doi.org/10.3390/ijerph18063010.

Cohen, A. C. (1973). The reflected Weibull distribution. *Technometrics, 15*(4), 867-873. https://doi.org/10.2307/1267396.

Crowder, S. V. and Hamilton, M. D. (2018). An EWMA for monitoring a process Standard Deviation. *Journal of Quality Technology, 24*(1) (2018) 12-21. https://doi.org/10.1080/00224065.1992.11979369.

Dar, A. A., Ahmed, A. and Reshi, J. A. (2017). Transmuted weighted exponential distribution and its application. *Journal of Statistics Applications and Probability, 6*(1), 219-232. https://doi.org/10.18576/jsap/060117.

Dallas, A. C. (1976). Characterizing the pareto and power distributions. *Annals of Institute of Statistical Mathematics 28*, 491–497. https://doi.org/10.1007/BF02504764.

Dennis, B. and Patil, G. (1984). The gamma distribution and weighted multimodal gamma distributions as models of population abundance. *Mathematical Biosciences, 68*(2), 187-212. https://doi.org/10.1016/0025-5564(84)90031-2.

Fisher, R. A. (1934). The effect of methods of ascertainment upon the estimation of frequencies. *Annals of Human Genetics. 6*(1), 13-25. https://doi.org/10.1111/j.1469-1809.1934.tb02105.x.

Grosenbaugh, L. R. (1958). Point-sampling and line-sampling probability theory, geometric implications, Synthesis. USDA Forest Service, Southern Forest Experiment Station, Occasional Paper 160.

Gove, J. H. and Patil, G.P. (1998) Modeling the basal area-size distribution of forest stands: acompatible approach. *Forest Science 44*(2), 285-297.

Huwang, L., Wu, C. H. and Lee, Y. R. (2021). EWMA and adaptive EWMA variable sampling intervals charts for simultaneous monitoring of Weibull parameters. *Quality Technology and Quantitative Management. 18*(5), 552-575. https://doi.org/10.1080/16843703.2021.1918439.

Jabeen, R., Ahmad, M., Zaka, A., Mansour, M. M., Aljadani, A. and Elrazik, E. M. (2024). A new statistical approach based on the access of electricity application with some modified control charts. *Journal of Mathematics*, 2024, Article ID 6584791. https://doi.org/10.1155/2024/6584791.

Jabeen R., Akhtar, A. and Zaka, A. (2023). Modified calibrated control charts for monitoring the population mean under stratified sampling. *Quality and Reliability Engineering International 2023*; *39*, 143–163. https://doi.org/10.1002/qre.3225.

Jabeen, R., Zaka, A. and Khan K. I. (2022). Classical estimator based modified control charts for phase-II monitoring in real life. *Quality and Reliability Engineering International 2022; 38*, 2862–2880. https://doi.org/10.1002/qre.3112.

Kao, J. H. (1958). Computer methods for estimating Weibull parameters in reliability studies. *IRE Transactions on Reliability and Quality Control, 13*, 15–22. https://doi.org/10.1109/IRE-PGRQC.1958.5007164.

Kao, J. H. (1959). A graphical estimation of mixed Weibull parameters in life testing electron tube. *Technometrics, 1* (1959) 389-407. https://www.jstor.org/stable/1266719.

Kranich, G. D. (2016). Inconsistent conceptions of acceleration contributing to formative assessment limitations, Electronic Theses and Dissertations. 2438. https://digitalcommons.library.umaine.edu/etd/2438.

Lappi,J and Bailey,R.L. (1987). Estimation of the diameter increment function or other tree relations using angle-count samples. *Forest Science. 33*(3) (1987) 725-739. https://doi.org/10.1093/forestscience/33.3.725

Mahfoud,M and G.P. Patil,G.P.(1982). On Weighted Distributions, Statistics and Probability, New York, (1982), 479-492.

Meniconi, M., Barry, D. M. (1996). The power function distribution: A useful and simple distribution to assess electrical component reliability. *Microelectronic Reliability, 36*(9), 1207–1212. https://doi.org/10.1016/0026-2714(95)00053-4

Montgomery D.C. (2012). Introduction to Statistical Quality Control, 7th edition. John Wiley and Sons, New York, USA.

Murthy,D.N., Xie,M and Jiang, R. (2004). Weibull Models. Wiley. ISBN: 978-0-471-47327-5.

Mutairi, A. A., Iqbal, M. Z. Arshad, B. Al-Mofleh, A. H. and Afify, A. Z. (2021). A new extended model with bathtub-shaped failure rate: properties, inference, simulation, and applications. *Mathematics. 9*(17) (2021). https://doi.org/10.3390/math9172024.

Naveed, M., Azam, M., Khan, N. and Aslam, M. (2018). Design of a control chart using extended EWMA statistic. *Technologies. 6*(4) (2018) 108. https://doi.org/10.3390/technologies6040108.

Ozsan, G., Testikb, M. C. and Weiß, H. (2010) Properties of the exponential EWMA chart with parameter estimation. *Quality and Reliability Engineering International, 26* (2010) 555-569. https://doi.org/10.1002/qre.1079.

Pascual,F. (2010). EWMA charts for the Weibull shape parameter. *Journal of Quality Technology, 42*(4) (2010)

400-416. https://doi.or\g/10.1080/00224065.2010.11917836.

Patil,G.P and Ord,J.K. (1976). On Size-Biased Sampling and Related Form-Invariant Weighted Distributions. *Industrial Journal of Statistics. 38*(1) (1976) 48- 61.

Patil,G.P.(1981). Studies in statistical ecology involving weighted distributions. In Proceedings of the Indian statistical Institute, Calcutta. (1981) 478-503.

Rao, C. R. (1965). On discrete distributions arising out of methods of ascertainment. In classical and contagious discrete distributions. Pergamon Press and Statistical Publishing Society, Calcutta, 320-332. https://www.jstor.org/stable/25049375

Roberts, S. W. (2000). Control chart tests based on geometric moving averages. *Technometrics, 42*(1), 97–101. https://doi.org/10.1080/00401706.2000.10485986

Siddiqui, S. A., Dwivedi, S., Dwivedi, P. and Siddiqui, I. (2016). Development of Size-Biased Mukherjee-Islam Distribution. *International Journal of Pure and Applied Mathematics, 107*(2) (2016) 505-515. https://doi.org/10.12732/ijpam.v107i2.18.

Taillie, C., Patil, G.P. and Hennemuth, R. (1995) Modeling and analysis of recruitment distributions. *Ecological and Environmental Statistics 2*(4), 315-329.

Warahena-Liyanage, G. B., Oluyede, T. Moakofi, T and W. Sengweni, W. (2023). The new exponentiated half logistic-Harris-G family of distributions with actuarial measures and applications. *Stats. 6,* (2023) 773–801, https://doi.org/10.3390/stats6030050.

Yu,D., Jin,L.L., Li,J., Qin,X., Z. Zhu,Z.and Zhang,J.(2023) Monitoring the Weibull scale parameter based on type I censored data using a modified EWMA control chart. *Axioms, 12*(5) (2023) 487.https://doi.org/10.3390/axioms12050487.

Zaka A., Akhter A.S. & Jabeen R. (2020). The new reflected power function distribution: theory, simulation and application. *AIMS Mathematics 5*(5): 5031–5054. https://doi.org/10.3934/math.2020323.

Zaka, A., Jabeen, R and Khan, K.I. (2021). Control charts for the shape parameter of skewed distribution. *Intelligent Automation & Soft Computing, 30*(3), 1007–1018. https://doi.org/10.32604/iasc.2021.016491.

Zeineldin ,R.A., Chesneau ,C., Jamal ,F and Elgarhy,M,(2019). Different estimation methods for type i half-logistic topp–leone distribution. *Mathematics, 7*(10) (2019) 985, https://doi.org/10.3390/math710098.

J.Natn.Sci.Foundation Sri Lanka 2025 54 (1): 31 - 43
DOI: http://dx.doi.org/10.4038/jnsfsr.v54i1.12587

RESEARCH ARTICLE

Deep Learning

Fast and reliable identification of abnormal crowd behaviour in surveillance footage

MM Krishnan[1] , ST Ramya[2] and K Ramar[3]
[1] *Department of Computer Science and Engineering. Francis Xavier Engineering College, Vannarpettai, Tirunelveli, Tamilnadu, India.*
[2] *Department of Computer Science and Design, RMK Engineering College, Kavarapettai, Tiruvallur, Tamilnadu, India.*
[3] *Department of Computer Science and Engineering, RMK College of Engineering and Technology, Pudovoyal, Tiruvallur, Tamilnadu, India.*

Submitted: 07 March 2025; Revised: 23 February 2026; Accepted: 07 March 2026

Abstract: Detecting unusual crowd events in surveillance video is crucial due to increased crime rates in recent years. However, automatically identifying these events is challenging because feature comparison across the training and test phases is time-consuming. To address this, we introduce the Markov-nearest transition unusual event classifier (MTUEC), which classifies input frames as either normal or unusual events. The MTUEC algorithm comprises several modules: the spatial slice model, static object removal computation, spatio-temporal estimation, and the Markov-nearest transition based unusual events classifier. The MTUEC algorithm focuses on comparing features between immediate training frames and the testing frame. If the immediate training frame does not match the testing frame, the algorithm considers other training frames for comparison. This approach significantly reduces the time needed to detect unusual crowd events. To examine the effectiveness of the proposed method, we used two benchmark datasets for unusual crowd events: UMN and UCSD Ped1 and Ped2. We also compared the performance of the proposed approach with several existing algorithms for unusual event detection.

Keywords: Events classification, feature comparison, Markov-nearest transition, slice model, unusual events detection.

INTRODUCTION

In the digital era, surveillance cameras are ever-present in areas where security is needed. Surveillance camera footage is used for post-event analysis (Kim et al., 2011), such as event recognition, action recognition, unusual crowd event detection, and abnormal event detection. These analyses can be applied across various surveillance domains, such as traffic surveillance environments, institutions, and hospitals. To conduct post-event analysis, many approaches, such as machine learning (Paul et al., 2023), deep learning (Lohithashva et al., 2018), clustering (Anjum & Cavallaro, 2008), and feature-based (Gnouma et al., 2018), have been widely used. Nowadays, anomaly detection is widely used in surveillance settings, as crime rates have increased in recent years. Therefore, an anomaly detection system is needed to alert people and rescue them from abnormal humans or situations. Unusual crowd behaviour analysis is also a promising application in video technology, as crowds can pose unwanted problems for the public. Though many research works have addressed unusual events and crowd problems, those studies have not achieved high accuracy, since analyzing crowd behaviour is a challenging task (Al-Khazaleh et al., 2026; Beigh et al, 2025; Jadhav & Bartere, 2025; Biswas & Babu, 2017). Developing crowd patterns and extracting features from density content are difficult tasks, since humans may be located at various locations, split and move in different directions, or all be located in the same place. Even if many algorithms try to detect objects or crowds, they are mostly developed for common scenes with a minimum number of people. In this situation, when these algorithms are applied to

* Corresponding author (mukeshkrishnan91.m@gmail.com; https://orcid.org/0009-0003-6401-9552)

crowd event frames, the unusual crowd event problem cannot be handled, since a large number of humans are present in the frame sequence in reality. Therefore, those algorithms cannot identify the effective features in the frame sequence (Chen et al., 2016).

Abnormal actions can be categorized into two types in the literature review (Chong et al., 2017; Chu et al., 2018; Cong et al., 2013): dense and sparse. Sparse abnormal actions involve only a few individuals and are characterized by conditions such as loitering, intrusion, falling, fighting each other, etc. On the other hand, groups of people may be involved in anomalous behaviour, such as unusual crowd activities, people trying to escape during a natural disaster, etc. Detecting abnormal crowd behaviour in a surveillance environment is a crucial challenge, as analysing large volumes of video data can be expensive and time-consuming. To reduce computational complexity, this paper introduced the Markov-transition based unusual events classifier (MTUEC) using the following four modules: spatial slice model, removal of static objects computation, spatio-temporal estimation, and unusual events classifier. The paper's main contributions are summarized here.

- Spatio-temporal estimation (STE) is introduced to estimate temporal information between moving objects for efficient analysis of unusual event detection.
- MTUEC is introduced to classify instances as normal or abnormal. Also, this classifier reduces feature comparison between the training and testing phases, since it mainly uses the previous instance for comparison.
- Various unusual event approaches, such as feature-based, machine learning, and deep learning-based methods, are used to analyze the results in different ways.

Related works

To capture intricate spatiotemporal patterns in video and trajectory data, deep learning-based models have become increasingly important for anomaly and uncommon event identification. The modelling of long-term dependencies in motion behaviour has been investigated using transformer designs and recurrent neural networks. To avoid bridge collisions, Mou et al. (2026) suggested a Transformer–BiLSTM hybrid model for identifying anomalous ship trajectories. Although the method works well for learning sequential dependencies, it is not suitable for real-time deployment in large-scale surveillance systems due to its expensive computational resources and reliance on massive amounts of training data.

Another strategy to lessen reliance on labelled data is self-supervised learning. Yang et al. (2025) presented a paradigm for video anomaly detection based on self-supervised spatio-temporal proxy tasks, enabling reliable feature learning without explicit labels. Low-latency applications may be hampered by this method's long temporal sequences and computationally intensive training process, despite its enhanced generalization.

With learning-based pattern modelling, trajectory-focused anomaly detection continues to develop. A thorough analysis of trajectory anomaly detection techniques was given by Li et al. (2025), who also highlighted new issues such scene reliance, occlusion, and trajectory fragmentation. Cho and Kang (2025) further address these problems by proposing a context-aware variational autoencoder for detecting pedestrian trajectory anomalies in metropolitan settings. VAE-based techniques are sensitive to reconstruction thresholds and require careful adjustment to prevent excessive false-positive rates, even with improved contextual modelling.

Liang et al. (2025) introduced an interaction–scene collaborative framework that represents object interactions and scene context together for traffic anomaly identification, going beyond trajectory-only representations. Although relational dynamics are well captured by this method, it is highly dependent on precise item detection and interaction modelling, both of which may suffer in cluttered or crowded environments. Similarly, Sundaram et al. (2025) used deep learning methods to analyze human behaviour in intelligent video surveillance systems, achieving excellent results at the expense of higher memory and processing demands.

In this case, spatial location information is erased by computing a histogram. Furthermore, most methods include a single mechanism to extract spatial or temporal information (Sodemann et al., 2021; Vallejo et al., 2009; Ahmed et al., 2018; Lin et al., 2015), because those algorithms work based on anyone's perception of objects, such as direction, appearance, motion, etc. The end result failed in many places due to illumination changes, which occluded the foreground object. It tends to have poor accuracy. To avoid these problems, STE is introduced to extract temporal information between moving objects for efficient detection of unusual events. Also, to reduce the computational and comparison complexity, this paper introduced the MTUEC technique by using the following four modules: spatial slice model, removal static objects computation, spatio-temporal estimation, and unusual deep learning-based techniques have been increasingly used in anomalous event classifiers in recent

years to overcome the drawbacks of manually created features and explicit object tracking. Spatiotemporal autoencoders, which learn compact representations via frame reconstruction or future frame prediction, have been widely used to model typical crowd behaviour. The identification of abnormal occurrences is based on high prediction or reconstruction mistakes. Despite their effectiveness, these techniques often struggle with complex motion patterns and require substantial training data to generalize effectively across other scenarios.

This concept is further developed by GAN-based video synthesis methods, which produce realistic future frames conditioned on historical observations (Jency & Ramar, 2025). When there is a substantial difference between the synthesized frames and the real observations, abnormal events are identified. GAN-based techniques are vulnerable to mode collapse, computationally intensive, and challenging to train, which reduces their resilience in real-world surveillance settings, even with improved motion dynamics models.

In recent years, transformer-based anomaly detection (Mancy & Naith, 2025) models have attracted attention for their ability to leverage self-attention mechanisms to capture long-range temporal relationships. When modelling intricate spatiotemporal interactions across lengthy video sequences, these techniques work admirably. However, they are less appropriate for real-time or resource-constrained surveillance systems due to their huge memory requirements and quadratic computational complexity with respect to sequence length.

By building spatiotemporal graphs, graph-based models have also been proposed to explicitly depict interactions among objects, regions, or motion patterns (Ilyas & Bawany, 2025). Relational dynamics in crowded scenes can be effectively modeled using graph convolutional networks. However, these methods usually depend on precise object tracking and detection, and they perform worse in crowded settings with a lot of occlusions and fragmented observations.

The majority of current methods have one or more of the following drawbacks, notwithstanding the success of these sophisticated models: they are computationally intensive, require extensive annotated data, are sensitive to changes in illumination and occlusion, or depend on long-term trajectory consistency. Furthermore, many approaches focus on either temporal dynamics or spatial appearance separately, leading to poor robustness in real-world surveillance scenarios.

Research gap

A number of important issues still need to be addressed, despite the significant advances in unusual event detection enabled by trajectory-based techniques and more recent deep learning approaches such as spatiotemporal autoencoders, GAN-based video synthesis, transformer-based anomaly detection, and graph-based models. While deep learning methods frequently require large-scale training data, significant computational resources, and complex model architectures, trajectory-based approaches struggle in scenarios with many people due to occlusion and fragmented tracking. The application of transformer-based and graph-based techniques in real-time surveillance scenarios is further limited by their reliance on lengthy temporal sequences or dependable object interactions. Furthermore, the majority of current approaches focus on either temporal dynamics or spatial appearance separately, resulting in reduced resilience to partial occlusions, background motion, and changes in lighting. As a result, there is a research need for a reliable and effective framework for detecting abnormal events that can preserve low computational and comparison complexity appropriate for real-world surveillance systems while simultaneously capturing spatiotemporal information without explicit trajectory dependency.

MATERIALS AND METHODS

The ultimate aim of the proposed method is to reduce the computational and feature comparison complexity by using the following four modules: 1. The spatial slice model is used to estimate the frame sequence using colour, texture, and edge features. 2. Removal static objects computation: it is used to identify the foreground objects in order to reduce the feature comparison while removing unwanted static objects. 3. Spatio-temporal estimation is employed to extract the motion of foreground objects. 4. Marko-nearest transition-based unusual events classifier is introduced to reduce the feature comparison between the training and testing phases. An overview of the proposed methodology is shown in Figure 1. The proposed method used the frame sequence as the input for analyzing unusual events.

More clarity can be obtained by clearly outlining the data pretreatment and module integration pipeline, even though the suggested methodology describes the main functional modules. To capture localized movement patterns, the suggested architecture first breaks down raw surveillance footage into individual frames, which are then spatially segmented using the spatial slice model.

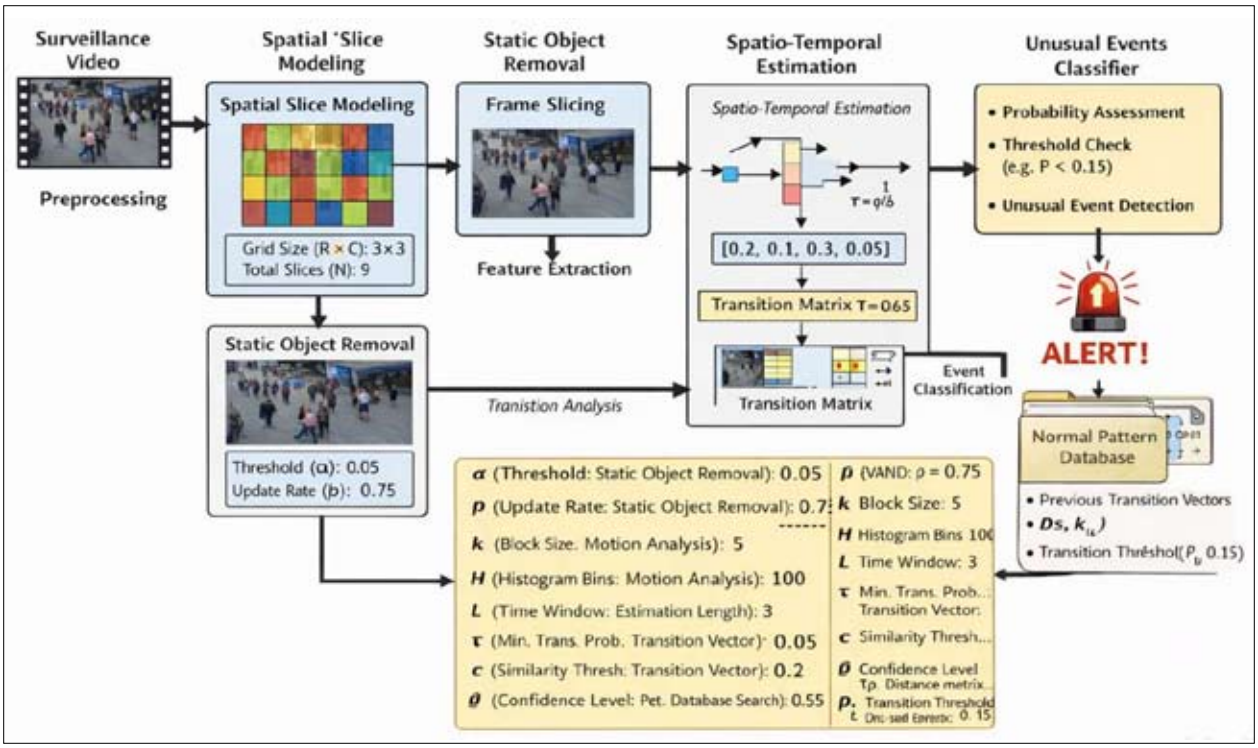


Figure 1: Proposed architecture

After that, immobile elements and background regions are removed using static object removal, ensuring that only significant dynamic regions are analyzed further. The analysis of frame-to-frame transitions from these processed slices yields spatiotemporal features, which are then modeled using a Markov-based nearest-neighbor transition technique to create compact feature vectors that depict local temporal dynamics. To differentiate between regular and abnormal occurrences, these feature vectors are then passed to the unusual occurrences classifier, which calculates transition probabilities and confidence scores. To balance detection accuracy and false alarm rates, rule-based thresholds, such as transition probability limits and confidence levels, are selected via empirical tuning on validation subsets of benchmark datasets. Future developments might include adaptive or data-driven threshold optimization, even if fixed thresholds guarantee computational effectiveness and real-time applicability. It would be easier to comprehend and give a clear visual grasp of module interactions if a flowchart showed this entire processing pipeline, from raw input frames to the final anomaly judgment.

Spatial slice model

To reduce feature comparison across frames, the frame sequence is segmented (Bird *et al.,* 2005) into equally spaced k vertical slices, where k = 1 to N, and then slice-wise feature-based normalization is carried out. The features such as colour c_i, texture t_i and edge e_i are derived on each slice, where $i = 1$ to n. The measurement of c_i, t_i, and e_i are used to efficiently describe the spatial information whether the foreground region moves slightly or abruptly. To estimate the data point d_p, where p = 1 to N, the slices are used to extract the c_i, t_i, and e_i features. Then, the slice-wise mean m is computed by dividing each data point value by the sum of the slice-wise data point feature values, since each data point summarizes the slices. A d_p can reflect the centre of the c_i, t_i, and e_i using the m parameter. The d_p will be derived on each slice using all three features c_i, t_i, and e_i. The resultant d_p can be explained by using the eqn. (1). The frame slices are shown in Figure 2.

$$m \text{ of } d_p = \frac{c_1+c_2+\ldots\ldots,+c_n}{number of\ c_i\ in the\ k^{th}\ slice}, \quad \frac{t_1+t_2+\ldots\ldots,+t_n}{number of\ t_i\ in the k^{th}\ slice}, \quad \frac{e_1+e_2+\ldots\ldots,+e_n}{number of\ e_i\ in the k^{th}\ slice}, \qquad \ldots(1)$$

where $c_1, c_2, \ldots c_n$, t_1, $t_2, \ldots t_n$, and e_1, $e_2, \ldots, e_n$ are feature values of data points.

Figure 2: The Spatial Slice Model

Removal static objects

To reduce computational and feature-comparison complexity, the static objects are located on the slice and indicated as '0' in the further process, since the moving objects can only express event types such as normal and unusual events (Vennila & Balamurugan, 2023). In general, the density based unusual events detection method cannot efficiently locate the foreground objects. To overcome this challenge, this study used a foreground detection method for identifying moving objects' foreground slices, using d_p of c_i, t_i, and e_i. Let d_q be the data points of the consecutive frame where $q = 1$ to N. To identify the moving foreground regions, the d_{pi}'s were compared with the corresponding d_q. The threshold values are applied to the difference between d_p and d_q. Let α, β, and γ be the threshold values on the difference values of colour, texture, and edge respectively. If the difference meets all three thresholds α, β, and γ, it is considered a moving object slice. Let d_j be the moving object slice, where j = 1 to K. These are explained using the Eqns. (2), (3), and (4).

$$d_j \text{ for color } = \begin{cases} \text{Yes,} & \text{if } (d_p - d_q) < \alpha \\ & \text{where } p \text{ and } q = 1 \text{ to } N \\ \text{No} & \text{Otherwise} \end{cases} \qquad \ldots(2)$$

$$d_j \text{ for texture } = \begin{cases} \text{Yes,} & \text{if } (d_p - d_q) < \beta \\ & \text{where } p \text{ and } q = 1 \text{ to } N \\ \text{No} & \text{Otherwise} \end{cases} \qquad \ldots(3)$$

$$d_j \text{ for edge } = \begin{cases} \text{Yes,} & \text{if } (d_p - d_q) < \gamma \\ & \text{where } p \text{ and } q = 1 \text{ to } N \\ \text{No} & \text{Otherwise} \end{cases} \qquad \ldots(4)$$

To efficiently detect the moving foreground region, the slices with colour, texture, and edge features should have their thresholds set using α, β, and γ. The moving foreground slices d_j can be selected using three threshold values and three features.

Spatio-temporal estimation on spatial slice

The moving foreground region slices with d_j of the frame sequence can be completely extracted by removing static object computation. To find the temporal information between frames, the orientation is used to explain the moving objects since the speed of the d_j can have a significant influence on calculating the displacement between the d_j of consecutive frames. The process of orientation is to estimate the position of the d_j. In general, the orientation of the d_j value does not change, even when the intensity values change due to external defects. Based on this assumption, this paper used Angular Velocity (AV) to obtain the efficient speed of the d_j. It refers to how fast the orientation of the *dj* changes over time. AV is obtained using gradients in the x and y directions. It is detailed using Eqn. (5).

$$tan\ \theta = \frac{G_y}{G_x} \qquad \ldots(5)$$

where G_x and G_y are the gradient values on x and y directions, respectively. Then, each orientation of the d_j is computed by subtracting the corresponding slice's orientation in the current frame *(f_s)* and the consecutive frame (f_{s+1}). It is represented as $d\theta$. AV needs time information to calculate the change in the d_j's orientation over a given period. The time t is computed using optical flow computation. It is obtained by estimating

the difference between the corresponding slice t in the current frame and the consecutive frame; it is denoted as *dt*. The AV is explained in Eqn. (6).

$$AV = \frac{d\theta}{dt} \quad ...(6)$$

The collection of *AV* of slices on the frame sequence is explained in the Eqn. (7). Let av_h be the collections of *AV* where $h = 1$ to *N*.

$$f_s = \{av_1, av_2, \ldots\ldots av_N\} \quad ...(7)$$

Unusual events detection

Let *A* be all the learned frames. *A* consists of normal and unusual events frames. The traditional classifier usually finds the distance between the new input frame and all learned frames (Song et al., 2019). Then the classifier finds the nearest training frame. Finally, the algorithm decides whether an event is normal or unusual based on the event's nearest distance value. By considering all learned frames for estimating the nearest distance, the classifier will incur greater computational complexity. To avoid this problem, this paper proposed MTUEC to efficiently reduce feature comparison between the training and testing phases to detect normal or unusual events using f_s.

In general, normal or abnormal events occur in continuous frames within a certain range. Based on this assumption, the MTUEC is developed to analyze the immediately prior learned frame. In order to firstly select the immediate previous learned frame to find the nearest distance, Markov-nearest transition classifier is used, in which Markov transition (Vennila & Balamurugan, 2020) is a process such that the choices of learned frames is only made on the previous learned frame. The MTUEC analyzes the distance of each d_j between training and testing instance. If the distance is less than σ, the certain d_j is considered as similar values and it is stored in the *B* vector. After finding all the distances d_j between training and testing phases, the MTUEC analyses the *B*. Let Ω be the variation value of distance among the incoming testing frame and the previous learned frame. If the distance variation is less than the Ω value in *B*, the MTUEC needs to check all learned frames until it gets Ω, since the normal or unusual events type may change into another type from the previous learned frame. The MTUEC procedure is discussed in the Algorithm 1.

Algorithm 1: MTUEC

Input: N frames, σ, and Ω
Output: normal or unusual events
For *s = 1* to *N*
Develops Spatial Slice Model
Computes Removal Static Objects
Estimates Spatio-Temporal on Slice
For *j = 1* to *K*
If the distance <σ
B = 1
Else
B = 0
End
End
Find sum(*B*)
If sum(*B*)>Ω
s is the unusual events frame
r = 1(where *r* is the first learned frame)
For *r = 1* to *N*
If the distance <σ
B = 1
If sum(*B*) >Ω
s is the unusual events frame
Else
B = 0
End
k = k+1
End
End
End
End

The proposed MTUEC-based odd-event detection system's entire processing pipeline is depicted in the flowchart in Figure 3. To correct illumination effects, the input surveillance video is first split into individual frames and preprocessed. To record localized motion patterns, each frame is then separated into uniform spatial slices. To ensure that only dynamic forefront information is preserved, adaptive thresholding is used in conjunction with static object removal to remove background regions. Compact transition vectors are produced by applying a Markov-nearest transition model to analyze frame-to-frame transitions within a predetermined temporal window to perform spatiotemporal estimation. A transition matrix of probabilities is created using these vectors and compared to previously learn normal patterns. Lastly, an alarm is triggered when anomalous behaviour is recognized by a threshold-based decision system that classifies events as either normal or abnormal.

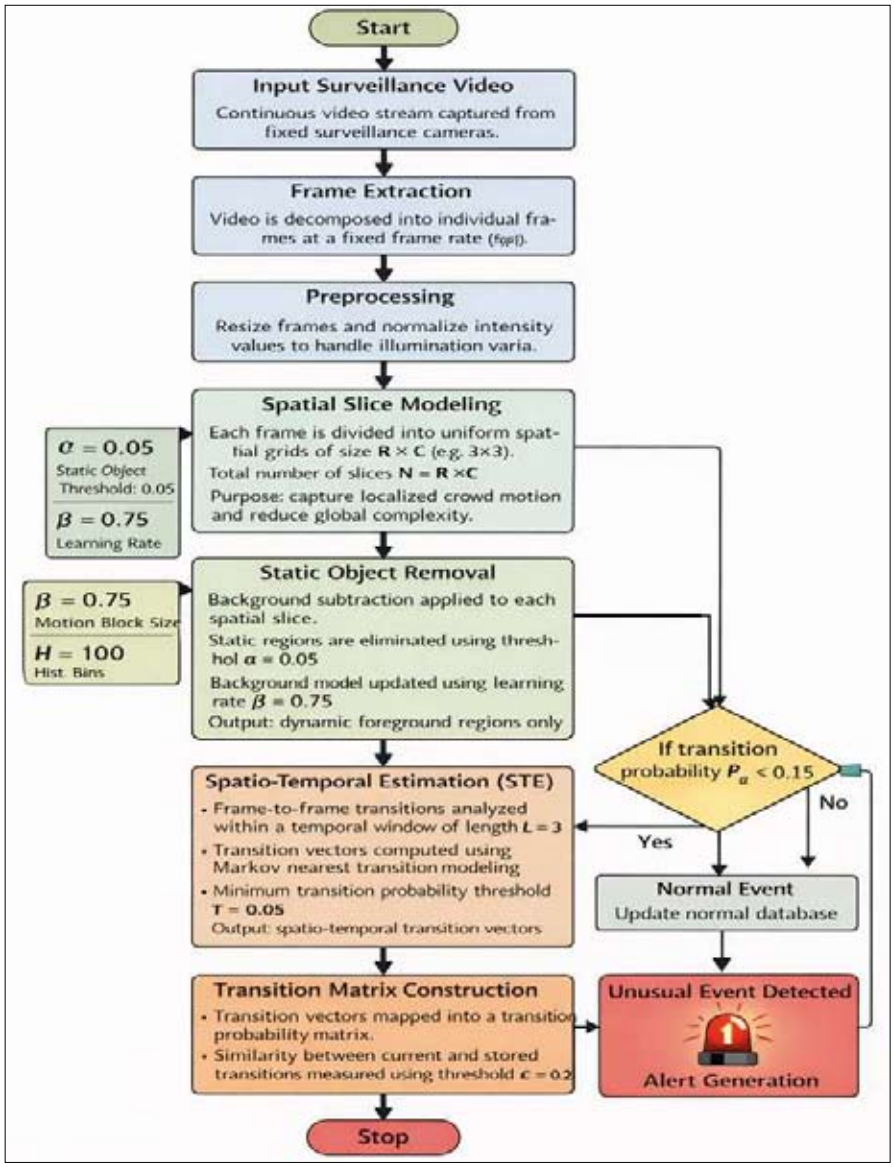


Figure 3: Flowchart of proposed model

RESULTS AND DISCUSSION

Based on the above constraints, the proposed MTUEC work was implemented, and the results were verified using MATLAB 13a. Unusual event detection benchmark datasets, such as UMN (https://UMN-dataset-anomaly-in-crowd_fig3_270790763) and UCSD Ped1 and UCSD Ped2 (http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm), were used to evaluate the performance of the proposed MTUEC. Also, the results of several unusual event detection algorithms were used to compare with the proposed work's results to estimate the effectiveness of the proposed MTUEC work.

Dataset description

UMN dataset: the dataset consists of three scenes. Different environments are captured all three scenes. All three scenes feature crowd density, followed by unusual crowd activities such as people running from the crowd, people merging, and people spreading.

USD dataset: the dataset was created using a stationary camera placed at an elevation, capturing pedestrian walkways. In the dataset, most frames contain crowd density information. The crowd density occurs across the frame sequence, spanning from sparse to dense crowds. Also, the dataset contains people in normal scenarios. On the other hand, people with high motion and non-people entries in the pedestrian walkways are observed. Non-people anomalies, such as small carts, bikers, and skaters, appear in the frame sequence. The normal and unusual events were naturally occurring during the creation of the dataset. The Ped1 consists of 34 training and 36 testing videos and Ped2 consists of 16 training and 12 testing videos. Each video consists of 200 frames. The ground-truth frames are available in the dataset, which consists of unusual events.

Additional real-world data: To further validate the robustness of MTUEC, we tested the algorithm on 10 real-world surveillance videos collected from publicly available urban surveillance footage. These videos, totaling 20 minutes, included a variety of crowd behaviours such as normal pedestrian movement, sudden dispersal, and sparse-to-dense transitions.

Performance evaluation

The frame-level evaluation is used during both normal and unusual events. The Precision, Recall, F1-score, AUC, and EER measures are used to evaluate the proposed MTUEC work performance. F1-score is used to measure the test accuracy. An F1-score of 100% indicates the best test accuracy. An F1-score of 0% indicates the worst test accuracy. These measures are estimated using true positive (TP)**,** false positive **(**FP)**,** false negative **(**FN), true positive rate (TPR), and false positive rate (FPR). TP denotes correctly detected unusual crowd activities frames, FP denotes incorrectly detects the normal frame as unusual crowd activities frame, FN denotes the incorrectly detected unusual crowd activities frame as normal frame, the true positive rate (TPR) is defined as the proportion of correctly identified positive frames relative to the total number of actual positive frames. The false positive rate (FPR), on the other hand, represents the proportion of falsely identified positive frames out of the total number of actual negative frames. They are described in the Eqns. (12), (13), (14), (15) and (16).

$$Precision = \frac{TP}{TP + FP}, \quad ...(12)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad ...(13)$$

$$F1 - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad ...(14)$$

$$FPR = \frac{\text{Number of False positive Frames}}{\text{Number of negative Frames}} \quad ...(15)$$

$$TPR = \frac{\text{Number of True positive Frames}}{\text{Number of positive Frames}} \quad ...(16)$$

Table 1: Experimental result on UCSD (Ped1 and Ped2) dataset

Video No.	UCSD Ped 1			UCSD Ped 2		
	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)
Video 1	93.00	60.00	73.00	91.00	55.00	69.00
Video 2	87.00	64.00	74.00	89.00	60.00	72.00
Video 3	90.00	74.00	81.00	92.00	71.00	80.00
Video 4	93.00	71.00	80.00	96.00	74.00	84.00
Video 5	89.00	79.00	83.00	93.00	75.00	83.00

Table 2: Experimental results on UMN dataset

Scene No.	Precision (%)	Recall (%)	F1-score (%)
Scene 1	88.00	66.00	72.00
Scene 2	86.00	74.00	80.00
Scene 3	91.00	88.00	89.00

Comparative analysis

The results on UCSD Ped1, Ped2, and UMN datasets are summarized in Tables 1 and 2, showing that MTUEC outperforms existing methods in precision and recall while reducing computational complexity. Existing unusual event detection algorithms, such as IGMM

(Gnouma *et al*., 2018), PNN (Lohithashva *et al.,* 2018), CNN (Direkoglu, 2020), GLCM (Lalit & Purwar, 2022), and OPLKT-EMEHO (Rajasekaran & Sekar, 2023), are used for comparison with the proposed MTUEC. To evaluate the effectiveness of the proposed MTUEC algorithm, various approaches to unusual event detection, including machine learning-based methods, are used. The comparative analyses are shown in Tables 3, 4, and 5.

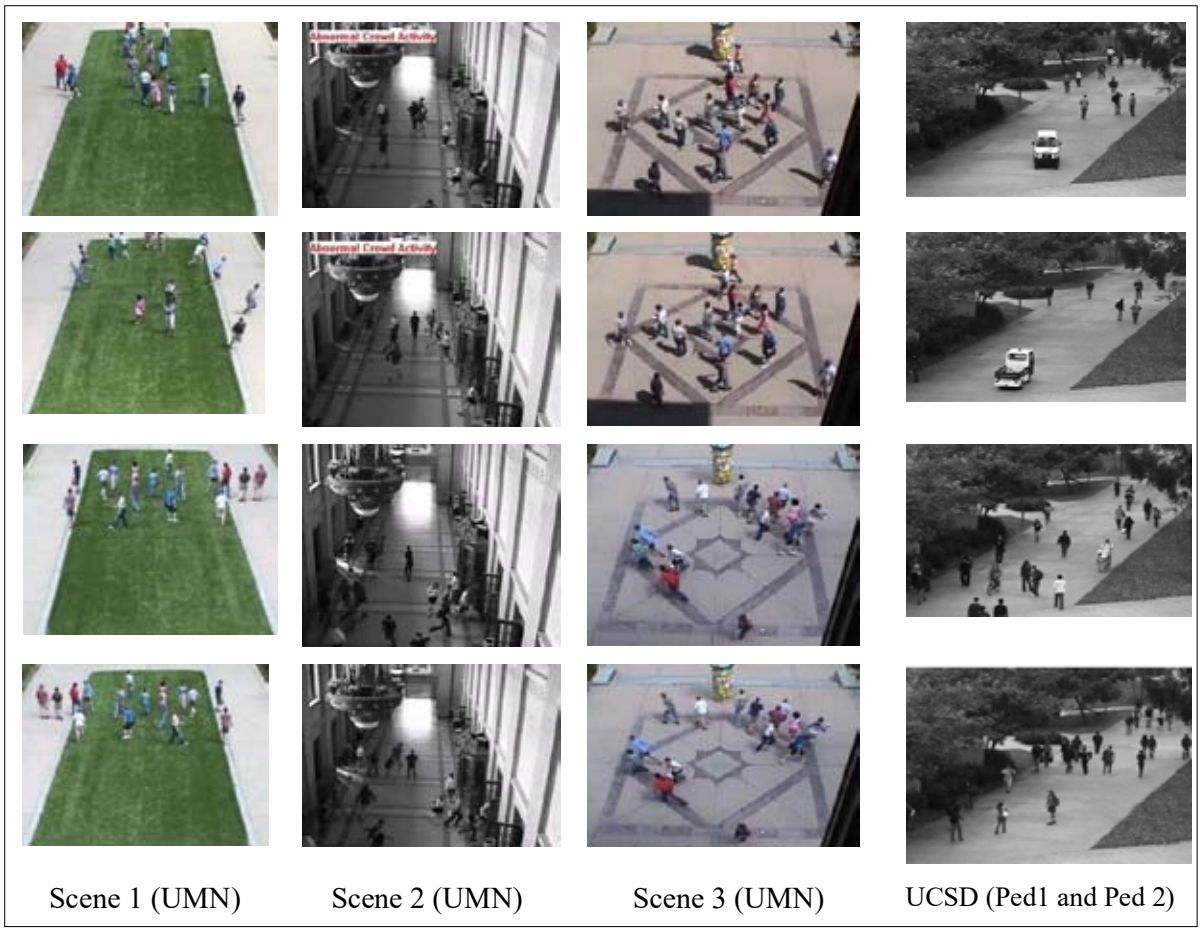


Figure 4: Examples of unusual crowd events and unusual moving objects

Table 3: Comparative analysis on UMN (Scene 1) dataset

Algorithms	Years	AUC	EER
IGMM	2018	79.20%	41.01%
PNN	2019	82.40%	15.00%
CNN	2020	85.00%	17.00%
GLCM	2022	90.00%	12.00%
OPLKT-EMEHO	2023	92.00%	13.00%
Proposed MTUEC		93.50%	14.00%

Table 4: Comparative analysis on UMN (Scene 2) dataset

Algorithms	Years	AUC	EER
IGMM	2018	81.25%	24.00%
PNN	2019	84.20%	13.50%
CNN	2020	86.00%	14.00%
GLCM	2022	91.10%	13.20%
OPLKT-EMEHO	2023	93.40%	13.10%
Proposed MTUEC		94.00%	15.00%

Table 5: Comparative analysis on UMN (Scene 3) dataset

Algorithms	Years	AUC	EER
IGMM	2018	85.00%	18.20%
PNN	2019	87.10%	14.20%
CNN	2020	88.20%	13.10%
GLCM	2022	92.00%	12.00%
OPLKT-EMEHO	2023	94.00%	12.00%
Proposed MTUEC		95.10%	13.30%

To further validate the robustness of MTUEC, we tested the algorithm on a small set of real-world surveillance footage (outside the training datasets). The model successfully detected anomalies in most cases, particularly sudden, unusual movements and small-object anomalies. However, some limitations were observed in scenarios where crowd movement was minimal or when a sparse-to-dense transition occurred, leading to a slight drop in recall performance. The detailed evaluation of MTUEC on real-world surveillance footage is presented in Table 6, highlighting both strengths and areas for improvement.

Table 6: Performance on real-world surveillance Footage

Scenario	Precision (%)	Recall (%)	F1-score (%)	Observations
Normal crowd movement	92.0	87.0	89.4	High accuracy in detecting normal events
Sudden unusual movement	90.5	85.0	87.6	Successfully detected running, dispersal, and merging
Sparse-to-dense transition	85.0	72.0	78.0	Some missed anomalies during gradual crowd build-up
Minimal crowd movement	80.0	68.0	73.2	Difficulty in detecting anomalies in static crowd
Small object anomalies (*e.g.*, bikers, carts)	88.0	75.0	81.0	Some misclassification due to size and speed

From Tables 3, 4, and 5, the proposed MTUEC algorithm yields better performance than the existing algorithms across measures such as precision, recall, F1-score, AUC, and EER. To reduce computational and feature-comparison complexity, the proposed MTUEC algorithm is developed using the following modules: spatial slice model, removal of static objects computation, spatio-temporal estimation, and Marko-nearest transition based unusual events classifier. The above modules help reduce computational and feature-comparison complexity because the features of the frame sequence are computed on the slice, not on each frame element. Then, the MTUEC analyzes the feature comparison between the immediate training frames with the testing frame. If the immediate training frame does not match with testing frame only, the MTUEC algorithm considers the other training frame for comparison. The MTUEC algorithm helps detect normal or unusual events effectively at the first iteration, since the frame sequence contains them continuously within a specific range. Though the MTUEC algorithm works in almost all cases, such as unusual crowds with more moving objects and unusual activities with more multiple objects suddenly moving from the crowd, it does not work well in cases of crowds with fewer moving objects and crowd spreading with fewer moving objects. The example of unusual events in the frame is shown in Figure 4.

To further validate the robustness of MTUEC, we tested the algorithm on 10 real-world surveillance videos collected from websites. These 20-minute videos spanned a variety of crowd behaviours, allowing us to assess the model's adaptability beyond standard benchmark datasets. The model successfully detected anomalies in most cases, particularly sudden, unusual movements and small-object anomalies. However, some limitations were observed in scenarios with minimal crowd movement or during sparse-to-dense transitions, resulting in a slight drop in recall. The detailed evaluation is presented in Table 6, highlighting both strengths and areas for improvement.

Real-time feasibility

The suggested approach significantly reduces the computational burden of exhaustive frame-wise feature evaluations by introducing a novel and useful Markov-nearest frame comparison algorithm. The suggested modular design, which includes spatial slicing and lightweight spatio-temporal estimation using angular velocity, provides an innovative and effective solution ideal for real-time surveillance applications, in contrast to deep learning-centric approaches that rely on intricate architectures and substantial training. Low-latency processing is enabled by a focus on local frame transitions and low memory utilization, making the method especially appealing for contexts with limited resources. A thorough examination of deployment-related aspects, such as per-frame processing time, memory footprint, adaptability to camera resolution, and robustness under changing illumination and crowd density, remains a crucial area for future research, even though the system is theoretically applicable to public surveillance scenarios like transit hubs and congested public areas. Including these assessments would increase the suggested framework's viability and preparedness for the real world.

Positive impacts

Compared with current monitoring systems, the proposed MTUEC framework offers several operational advantages that enhance its practicality. The method greatly reduces computational load by using Markov-nearest transitions to reduce exhaustive frame-wise feature comparisons. This allows for faster incident identification and minimizes delay in real-time monitoring applications. Instead of continuously monitoring several video feeds, security staff can be instantly alerted to unexpected events through reduced processing complexity, thereby directly reducing operator workload. The suggested method can be incorporated into existing CCTV and smart surveillance systems that use conventional CPU-based infrastructure, unlike deep learning-based surveillance software, which requires expensive GPUs, substantial memory footprints, and offline training. Additionally, the system's modular design and lightweight spatio-temporal estimation enable scalability across multiple cameras, making it suitable for deployment in expansive public areas such as shopping malls, train stations, and airports. Together, these features strengthen the suggested method's deployment potential and practical significance by positioning it as an affordable, low-latency substitute for existing surveillance technologies.

Ethical considerations and privacy implications

There are significant privacy, ethical, and societal issues with the use of automated systems for anomalous event detection in public monitoring settings. Although the suggested framework uses motion-based spatiotemporal features rather than biometric identifiers, caution must be taken to ensure compliance with data protection laws such as the general data protection regulation (GDPR). Video data should be anonymised, access-controlled, and retained for a limited period in practical deployments, in compliance with institutional and regulatory standards. Furthermore, automated anomaly detection systems are prone to false positives, which can lead to erroneous warnings or skewed perceptions of typical crowd behaviour across diverse cultural and environmental contexts. These biases may result from scene-dependent dynamics or motion patterns unique to a dataset. The suggested system should be used as a decision-support tool rather than as a fully autonomous authority to reduce these risks, and the alarm verification process should include human oversight. To improve ethical robustness and responsible deployment, future research will concentrate on integrating adaptive thresholding, bias-aware evaluation across various datasets, and privacy-preserving processing techniques.

CONCLUSION

This paper introduced the Markov-nearest transition unusual event classifier (MTUEC) algorithm, aimed at reducing the computational complexity and feature comparison time during the detection of unusual crowd events. The MTUEC algorithm was developed using several integrated modules, including the spatial slice model, static object removal, spatio-temporal estimation, and the Markov-nearest transition-based classifier. Our experimental results demonstrate that the MTUEC algorithm performs effectively in most scenarios, especially in cases involving unusual crowds with numerous moving objects and sudden changes in crowd movement patterns.

However, the algorithm exhibits limitations in scenarios with sparse crowds or fewer moving objects, where detecting unusual events becomes more challenging. Additionally, its performance may be affected by noisy or low-quality video inputs. Future work could focus on enhancing the algorithm's robustness to sparse crowds and its adaptability across diverse surveillance settings. Furthermore, the proposed method shows potential for application in various real-world surveillance tasks,

such as monitoring public spaces to help prevent crowd collisions and ensuring safety during mass gatherings.

REFERENCES

Ahmed, S. A., Dogra, D. P., Kar, S., & Roy, P. P. (2018). Trajectory-based surveillance analysis: A survey. *IEEE transactions on circuits and systems for video technology*, *29*(7), 1985-1997. https://doi.org/10.1109/TCSVT.2018.2857489

Al-Khazaleh, S., Badwan, N., & Saleh, Q. (2026). Digital economy prosperity analysis based on the Grey–Markov model: Empirical evidence from MENA region. *International Journal of Islamic and Middle Eastern Finance and Management, 19*, 269–300. https://doi.org/10.1108/IMEFM-07-2024-0348

Anjum, N., & Cavallaro, A. (2008).Multifeature object trajectory clustering for video analysis. *IEEE Transactions on Circuits and Systems for Video Technology*, *18*(11), 1555-1564. https://doi.org/10.1109/TCSVT.2008.2005603

Beigh, T. M., Venkatesan, V. P., & Arumugam, J. (2025). Predictive summarization framework for resource-constrained device surveillance videos. *Pattern Recognition*, 112274. https://doi.org/10.1016/j.patcog.2025.112274

Bird, N. D., Masoud, O., Papanikolopoulos, N. P., & Isaacs, A. (2005). Detection of loitering individuals in public transportation areas. *IEEE Transactions on Intelligent Transportation Systems*, *6*(2), 167-177. https://doi.org/10.1109/TITS.2005.848370

Biswas, S., & Babu, R. V. (2017). Anomaly detection via short local trajectories. *Neurocomputing*, *242*, 63-72. https://doi.org/10.1016/j.neucom.2017.02.058

Chen, H., Zhao, X., Wang, T., Tan, M., & Sun, S. (2016, June). Spatial-temporal context-aware abnormal event detection based on incremental sparse combination learning. In *2016 12th World Congress on Intelligent Control and Automation (WCICA)* (pp. 640-644). IEEE. https://doi.org/10.1109/WCICA.2016.7578533

Cho, J., & Kang, Y. (2025). Context-Aware Anomaly Detection of Pedestrian Trajectories in urban back streets using a variational autoencoder. *ISPRS International Journal of Geo-Information*, *14*(11), 438. https://doi.org/10.3390/ijgi14110438

Chong, Y. S., & Tay, Y. H. (2017, May). Abnormal event detection in videos using spatiotemporal autoencoder. In *International symposium on neural networks* (pp. 189-196). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-59081-3_23

Chu, W., Xue, H., Yao, C., & Cai, D. (2018). Sparse coding guided spatiotemporal feature learning for abnormal event detection in large videos. *IEEE Transactions on Multimedia*, *21*(1), 246-255. https://doi.org/10.1109/TMM.2018.2846411

Cong, Y., Yuan, J., & Liu, J. (2013).Abnormal event detection in crowded scenes using sparse representation. *Pattern Recognition*, *46*(7), 1851-1864. https://doi.org/10.1016/j.patcog.2012.11.021

Direkoglu, C. (2020). Abnormal crowd behavior detection using motion information images and convolutional neural networks. *IEEE Access*, *8*, 80408-80416. https://doi.org/10.1109/ACCESS.2020.2990355

Gnouma, M., Ejbali, R., & Zaied, M. (2018). Abnormal events' detection in crowded scenes. *Multimedia Tools and Applications*, *77*(19), 24843-24864. https://doi.org/10.1007/s11042-018-5701-6

Ilyas, A., & Bawany, N. (2025). Crowd dynamics analysis and behavior recognition in surveillance videos based on deep learning. *Multimedia Tools and Applications*, *84*(23), 26609-26643. https://doi.org/10.1007/s11042-024-20161-7

Jadhav, S., & Bartere, M. (2025).Generalizable and privacy-preserving multimedia forgery detection via semantically disentangled and temporally-aware deep learning architectures. *International Journal of Information Technology, 1*, 1–15. https://doi.org/10.1007/s41870-025-02978-z

Jency, A., & Ramar, K. (2025). A Review of abnormal behaviour detection in crowd for video surveillance: advances and trends, datasets, opportunities and prospects. *Expert Systems*, *42*(4), e70013. https://doi.org/10.1111/exsy.70013Digital Object Identifier

Kim, J. S., Yeom, D. H., & Joo, Y. H. (2011). Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems. *IEEE Transactions on Consumer Electronics*, *57*(3), 1165-1170. https://doi.org/10.1109/TCE.2011.6018870

Lalit, R., & Purwar, R. K. (2022). Crowd abnormality detection using optical flow and GLCM-based texture features. *Journal of Information Technology Research (JITR)*, *15*(1), 1-15. https://doi.org/10.4018/JITR.2022010110

Li, C., Wang, X., Wang, X., Chen, S., Ye, J., & Zhao, S. (2025, June). Pattern Learning for Trajectory Anomaly Detection: Recent Developments and Emerging Challenges. In *2025 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-10).IEEE.DOI: 10.1109/IJCNN64981.2025.11227964

Liang, R., Li, Y., Wu, Z., & Li, X. (2025). An interaction-scene collaborative representation framework for detecting traffic anomalies in driving videos. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/TITS.2025.3553830

Lin, W., Zhang, Y., Lu, J., Zhou, B., Wang, J., & Zhou, Y. (2015). Summarizing surveillance videos with local-patch-learning-based abnormality detection, blob sequence optimization, and type-based synopsis. *Neurocomputing*, *155*, 84-98. https://doi.org/10.1016/j.neucom.2014.12.044

Lohithashva, B. H., Manjunath Aradhya, V. N., Basavaraju, H. T., & Harish, B. S. (2018, December). Unusual crowd event detection: an approach using probabilistic neural network. In *Information Systems Design and Intelligent Applications: Proceedings of Fifth International Conference INDIA 2018 Volume 1* (pp. 533-542). Singapore: Springer Singapore. https://doi.org/10.1007/978-981-13-3329-3_50

Mancy, H., & Naith, Q. H. (2025). SwinIoT: A hierarchical transformer-based framework for behavioral anomaly

detection in IoT-Driven smart cities. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3551207

Mou, H., Rong, H., & Teixeira, A. P. (2026). Detecting abnormal ship trajectory to avoid bridge collisions via a Transformer-BiLSTM model. *Ocean Engineering*, *343*, 123232. https://doi.org/10.1016/j.oceaneng.2025.123232

Paul, M. A., Kumar, K. S., Sagar, S., & Sreeji, S. (2023). LWDS: lightweight DeepSeagrass technique for classifying seagrass from underwater images. *Environmental Monitoring and Assessment*, *195*(5), 614. https://doi.org/10.1007/s10661-023-11183-z

Rajasekaran, G., & Sekar, J. R. (2023). Abnormal Crowd Behavior Detection Using Optimized Pyramidal Lucas-Kanade Technique. *Intelligent Automation & Soft Computing*, *35*(2). https://doi.org/10.32604/iasc.2023.029119

Song, H., Sun, C., Wu, X., Chen, M., & Jia, Y. (2019).Learning normal patterns via adversarial attention-based autoencoder for abnormal event detection in videos. *IEEE Transactions on Multimedia*, *22*(8), 2138-2148. https://doi.org/10.1109/TMM.2019.2950530

Sundaram, D. M., Mannam, S., & Jayachandran, S. (2025, August). Intelligent Video Surveillance for Human Behavior Analysis and Anomalous Activity Detection using Deep Learning. In *2025 5th International Conference on Soft Computing for Security Applications (ICSCSA)* (pp. 950-955). IEEE. https://doi.org/10.1109/ICSCSA66339.2025.11171248

Vallejo, D., Albusac, J., Jimenez, L., Gonzalez, C., & Moreno, J. (2009).A cognitive surveillance system for detecting incorrect traffic behaviors. *Expert Systems with Applications*, *36*(7), 10503-10511. https://doi.org/10.1016/j.eswa.2009.01.034

Vennila, T. J., & Balamurugan, V. (2020, February).A stochastic framework for keyframe extraction. In *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)* (pp. 1-5). IEEE. https://doi.org/ 10.1109/ic-ETITE47903.2020.294.

Vennila, T. J., & Balamurugan, V. (2023). A rough set framework for multihuman tracking in surveillance video. *IEEE Sensors Journal*, *23*(8), 8753-8760. https://doi.org/10.1109/JSEN.2023.3242007

Yang, Q., Wang, C., Liu, P., Jiang, Z., & Li, J. (2025). Video anomaly detection via self-supervised and spatio-temporal proxy tasks learning. *Pattern Recognition*, *158*, 111021. https://doi.org/10.1016/j.patcog.2024.111021

J.Natn.Sci.Foundation Sri Lanka 2026 54 (1): 45- 61
DOI: http://dx.doi.org/10.4038/jnsfsr.v54i1.12512

RESEARCH ARTICLE

Biodiversity Conservation

A holistic conservation approach for threatened filmy and grammitid ferns of Sri Lanka

PGSA Kumarasinghe[1], RHG Ranil[2*], RMSR Chamara[3], UGTN Janaranjana[4], GMWL Gunawardena[5], and B Parris[6]

[1] *School of Natural Science, University of Chester, CH1 4BJ, United Kingdom.*
[2] *Department of Crop Science, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka.*
[3] *Department of Plantation Management, Faculty of Agriculture and Plantation Management, Wayamba University of Sri Lanka, Makandura, Gonawila, Sri Lanka.*
[4] *Postgraduate Institute of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka.*
[5] *Typhoon Science and Technology Research Center, Yokohama National University, 79-1 Tokiwadai, Hodogaya Ward, Yokohama, Kanagawa, Japan.*
[6] *Fern Research Foundation, 21 James Kemp Place, Kerikeri, Bay of Islands, 0230, New Zealand.*

Submitted: 03 January 2025; Revised: 28 September 2025; Accepted: 30 January 2026

Abstract: Isolated tropical islands support unique biodiversity and high endemism but face high extinction risks for some plant species due to small, fragmented populations and increasing anthropogenic pressures. Sri Lanka, a continental island in the Western Ghats–Sri Lanka biodiversity hotspot, hosts remarkable pteridophyte flora, including two highly threatened groups: the filmy ferns (Hymenophyllaceae) and grammitid ferns (Grammitidoideae). Owing to their narrow ecological niches and dependence on humid, shaded microhabitats, these taxa are highly sensitive to microclimatic shifts and habitat disturbance. This study proposes a holistic conservation approach to safeguard these fern groups by integrating historical and contemporary data, herbarium records, geospatial analyses, hotspot identification, and expert knowledge. Using 752 herbarium specimens from 22 global herbaria, together with recent field collections, we mapped species distributions, examined temporal collection patterns, and identified priority regions and climatic envelopes essential for their survival. Results reveal that high-elevation and southwestern wet-zone forests are key hotspots. Additionally, the analysis identified three hotspots for grammitid ferns and two for filmy ferns. The study outlines a roadmap incorporating species prioritization, targeted field surveys, microhabitat restoration, *in situ* conservation, *ex situ* propagation via high-tech greenhouses, community engagement, and collaborative research. Emphasis is placed on strengthening national herbarium resources, integrating citizen science, applying genomic tools, and aligning conservation actions with global frameworks. This holistic, multidisciplinary strategy provides a comprehensive foundation for conserving Sri Lanka's threatened filmy and grammitid ferns and offers a model applicable to other island flora worldwide.

Keywords: Conservation, grammitidoideae, hymenophyllaceae, island populations, threatened species.

INTRODUCTION

Isolated islands are characterized by rich biological diversity and high endemicity, but the unique biodiversity of many islands is experiencing higher extinction rates than that of mainland populations (Frankham, 1997; Woinarski, 2010; Rønsted et al., 2022). Hamabata et al. (2019) claimed that endangered island endemic plants have vulnerable genomes and face a higher extinction risk than non-endangered species. Also, global change agents often have a greater impact on the dynamics of island ecosystems (Oro et al., 2022). The small and isolated populations of many island taxa are particularly vulnerable to human activities (Nogué et al., 2017), and

* Corresponding author (rrajapaksha@agri.pdn.ac.lk; https://orcid.org/0000-0003-2880-5428)

the rapid increase in natural resource exploitation on tropical islands, often pursued to address rural poverty and stimulate economic growth, has already driven numerous species to extinction or to an endangered status, a trend that is further intensified by the impacts of climate change (Jaisankar et al., 2018). Therefore, island populations demand greater conservation priority compared to mainland populations.

Sri Lanka is a tropical continental island characterized by rich floristic diversity and high endemism, primarily due to its long-term isolation from mainland India and its unique climate, soils, and geography (Gunatilleke et al., 2008; Pethiyagoda & Sudasinghe, 2017). Despite the collective importance of the Western Ghats–Sri Lanka biodiversity hotspot, the island of Sri Lanka has distinct and exceptional biodiversity that differs from the Ghats and southern India (Sarathchandra et al., 2021). Pteridophytes are a vital component of biodiversity and have great potential as ecological indicators due to the correlation between their geographic distribution and abiotic variables (Della, 2022). Sri Lanka is home to 389 pteridophyte taxa with 42 endemic species, which is a considerable number compared to the size of the island (Ranil et al., 2022). According to the National Red List 2020, 25 species of 350 native pteridophyte taxa are classified as critically endangered (possibly extinct), including four endemics. Additionally, 207 taxa (59%) are considered threatened, and 16 are listed as data deficient. Of particular concern is that 28 out of 42 Sri Lankan endemic taxa (67%) have been classified as threatened, highlighting the urgent need for conservation efforts.

From a conservation point of view, Sri Lankan ferns, specifically filmy and grammitid ferns, hold significant importance as they face a considerably higher risk of extinction than other pteridophyte species. In fact, 95% of filmy ferns and 68% of grammitid ferns in Sri Lanka have been classified under the threatened category, according to the National Red List 2020. Filmy and grammitid ferns predominantly thrive in epiphytic or rupestral habitats, with terrestrial occurrences being rare, and their epiphytic and epilithic nature makes them highly sensitive to microclimatic changes due to their close dependence on host trees. Consequently, they are largely restricted to the humid, shaded environments of Sri Lanka's rainforest ecosystems, which provide the stable microhabitats essential for their survival. Among the various factors, the moisture content of their immediate environment plays a critical role in determining their survival. A notable example is the family *Hymenophyllaceae*, which exhibits poikilohydric characteristics (Aros-Mualin & Kessler, 2024).

By the dawn of the nineteenth century, Sri Lanka's forest cover was estimated at around 70% of the total land area, but it has progressively decreased over time. These facts highlight the critical need for the conservation of these particular fern species against the changing climate and unsustainable land use. Furthermore, their vulnerable status emphasizes the urgency of implementing immediate conservation measures to safeguard their natural habitats while enhancing our understanding of their ecological and biological significance. Thus, assessing the impact of threats, identifying and prioritizing vulnerable taxa, and predicting conservation pockets are vital for the conservation and management of threatened species (Mehta et al., 2020).

Given the high extinction risk and significant ecological value of filmy and grammitid ferns in Sri Lanka, the primary research question guiding this study is: What comprehensive conservation measures and strategies can be implemented to effectively protect and sustain the populations of grammitid and filmy ferns in Sri Lanka? Therefore, the main objective of this study is to propose conservation measures and a roadmap for these two highly threatened fern groups in Sri Lanka. To accomplish this goal, we utilized a holistic conservation approach and developed conservation strategies for species within these two groups. Additionally, this study highlights the areas that need to be strengthened and protocols that should be implemented for the successful conservation of filmy and grammitid ferns in Sri Lanka. The proposed conservation measures and roadmap aim to serve as a comprehensive framework for developing conservation strategies to protect island populations of endangered species worldwide.

MATERIALS AND METHODS

The study was primarily based on a combination of existing literature and the firsthand experiences of the researchers. Additionally, the study used various web-based data repositories and employed several scientific tools and techniques (including GIS applications) together with researchers' opinions. This holistic approach enabled a comprehensive analysis of both qualitative and quantitative data, ensuring a thorough assessment of the current status and future prospects of these highly vulnerable groups of species.

Selection of candidate families and species

Our analysis was specifically directed toward the family Hymenophyllaceae (filmy ferns) and the subfamily Grammitidoideae (grammitid ferns). The grammitid

ferns were earlier treated as a separate family, the Grammitidaceae, and based on the recent molecular work, they were nested within the family Polypodiaceae (PPG I, 2016). The selection of these two groups was based on the level of threat they face, as determined by the proportion of species classified as threatened in the National Red List (2020). We used Rajapaksha et al. (2022) as a base document to extract the 42 species belonging to two candidate families.

Data collection and categorization

We primarily considered the locality information given in "*A Handbook to the Flora of Ceylon*" (Volume 16A & B) for the family Hymenophyllaceae (Jayasekara, 2006) and the subfamily Grammitidoideae (Parris, 2006). It consists of 752 herbarium records based on the specimens deposited at 22 worldwide herbaria. Additionally, the collections made by this publication's corresponding author from 2003 to 2021 were also considered while preparing spatial distribution maps. The virtual herbaria (B, BR, BM, E, K, G, P, W) were also consulted for Sri Lankan ferns deposited there. The information on the date of collection, collector name, locality details (province, district, village, and elevation), and habitat characteristics was extracted from the above-mentioned literature, herbarium records and personal collections. Information on historical herbarium records of all candidate species was organized into 50-year intervals to examine changes in collection intensity over time. All available records from 1800 to 2020 were compiled and categorized into the following periods: 1800-1850, 1850-1900, 1900-1950, 1950-2000, and 2000-2020. The number of specimens representing Grammitidoideae and Hymenophyllaceae in each interval was then quantified. These data were used to construct a temporal bar chart illustrating variations in collection effort and the historical progression of pteridophyte documentation in Sri Lanka. The National Red List (2020) was used to identify the conservation status of each species.

Construction of phylogenetic tree

Phylogenetic analysis was conducted to determine the placement of Sri Lankan pteridophyte families within the global pteridophyte phylogeny, with particular focus on the Grammitidoideae and Hymenophyllaceae. Chloroplast *rbcL* gene sequences were used for this analysis. Representative *rbcL* sequences from selected species of each family were obtained from GenBank and incorporated into the dataset. This phylogenetic tree was used to illustrate the approximate placement of Sri Lankan pteridophyte families within the broader global pteridophyte phylogeny, providing readers with an overview of where the Sri Lankan families, particularly Grammitidoideae and Hymenophyllaceae, fit within the global context. It also summarized their total, endemic, and threatened taxa in Sri Lanka.

Mapping of spatial distribution

Locality information for Hymenophyllaceae and Grammitidoideae was obtained from *A Revised Handbook to the Flora of Ceylon* (Jayasekara, 2006; Parris, 2006), which includes 752 herbarium records from 22 international herbaria. Additional records were incorporated from collections made by the corresponding author between 2003 and 2021. All locality descriptions were checked against original herbarium labels and updated, when necessary, using topographic maps and online gazetteers. Records that could not be georeferenced with reasonable accuracy (±5 km), were duplicates, or represented cultivated plants were excluded. When multiple specimens of the same species came from the same locality, only one representative record was retained. After screening, a total of 471 unique and reliably georeferenced localities were selected for mapping.

Spatial analyses were carried out in QGIS (v. 3.28.3) and ArcGIS (v. 10.3), with all coordinates standardized to the WGS 84 coordinate system. Administrative boundary layers for Sri Lanka's 25 districts and three major climatic zones (wet, intermediate, dry) were obtained from the Survey Department of Sri Lanka and reprojected to the same coordinate system. Each locality was assigned to a district and climatic zone through spatial overlay. Species distribution maps were generated separately for Hymenophyllaceae and Grammitidoideae. Locality points were symbolized by species presence, and district-level species richness was calculated by counting the number of species recorded within each district. Maps were exported using standardized scale, layout, legends, and scale bars to allow easy comparison among figures.

Identification of hotspots

Species occurrence hotspots were identified using the Hot Spot Analysis (Getis–Ord Gi*) tool in ArcGIS. Each species locality point was assigned a constant weight of one. Spatial relationships were conceptualized using the default fixed-distance band, with the threshold distance selected based on the tool's recommendations for the dataset. Local Gi* statistics were computed, and clusters with Gi_Bin ≥ 2 (≥ 95% confidence) were interpreted as significant hotspots. These clusters were compared with known forested regions and major topographic features to confirm their biological significance for conservation planning.

Climatic preferences and habitat characteristics

Climatic preferences for hotspots and species-rich areas were assessed using gridded climate data from the NASA POWER database (1981–2021). For each hotspot and locality, mean annual temperature, rainfall, and relative humidity were extracted from the NASA rasters. Climate layers were resampled, where necessary, to a common resolution and processed using raster calculations and zona statistics in ArcGIS. The resulting values were summarized into ecologically meaningful ranges, representing the preferred climatic envelopes for filmy and grammitid ferns. No additional interpolation or species distribution modelling was performed; analyses were based solely on observed localities and hotspot areas.

A roadmap and proposed conservation measures

Future conservation measures were proposed based on gathered information from past herbarium records, relevant literature, developed distribution maps, and the authors' personal observations. To effectively achieve the conservation goals, we have extensively discussed the importance of accommodating additional supportive services and technological tools. A flowchart was developed to visually represent the diverse fields, elements, and tools to develop these conservation guidelines and the roadmap to achieve future conservation targets. Furthermore, results highlight areas that require further strengthening and attention.

RESULTS AND DISCUSSION

Here, we provide detailed guidelines for developing a conservation plan for filmy and grammitid ferns of Sri Lanka, aligned with the proposed schematic diagram. In addition, we highlight conservation gaps and provide supportive information for bridging these gaps to facilitate the development of a comprehensive strategic plan for the candidate species.

Understanding of overall diversity

Understanding the overall diversity of the target taxa or group within a specific geographic unit and their past and current status is the initial step in developing any conservation plan. Figure 1 illustrates the placement of pteridophyte families of Sri Lanka within the global pteridophyte phylogenetic tree, along with their total number of taxa and number of endemics and threatened species. This information serves as a foundational resource for the selection and prioritization of families in developing conservation and management guidelines.

The Hymenophyllaceae is represented by 20 species, including one endemic to Sri Lanka. Jayasekara (2006) has provided a description of each species in the enumeration of the Hymenophyllaceae for "*A revision of the flora of Ceylon*", Volume 15(A). His enumeration included recent collections of some taxa from a few selected localities across the wet and intermediate zones of Sri Lanka. The subfamily Grammitidoideae is represented by 22 species in eight genera, 36% (8 species) of which are endemic to the country. Parris (2006) has provided detailed descriptions of all species for "*A revision of the flora of Ceylon*, Volume 15(A)". Such enumeration was based on the specimens available at the 22 worldwide herbaria. However, having comprehensive and detailed botanical descriptions greatly facilitates the process of setting conservation guidelines.

Figure 2 illustrates the number of specimens used in the present study, collected between 1800 and 2020 and currently housed in 22 herbaria. This reveals that the highest number of specimens have been collected between 1950-2000, followed by 1850-1900. W.A. Sledge's specimen collections from his enumeration of Sri Lanka's fern flora (1950-1951 and 1954) significantly contributed to the increase in recorded specimens during this period, followed by additional collections made by Robert B. Faden in 1976. Most Sri Lankan fern specimens were collected in the 19th century during the British Colonial period (Rajapaksha *et al*., 2022). It is the main reason for the deficit in the second-largest collection from 1850 to 1900. The recent explorations reveal that some species are overlooked and demonstrate a wider range of distribution than previously assumed (Figure 3). However, future collections should be expedited to bridge research gaps and set more effective conservation priorities.

Mapping of spatial distribution pattern

Mapping geographical distribution patterns provides valuable information for understanding the ecological requirements and identifying the most preferable habitats. This knowledge forms the fundamental basis for developing effective conservation plans. Spatial distribution maps depicting the distribution of each group across the administrative districts and the three major climatic zones are provided herein (Figure 4). It is evident from these maps that the vast majority of the recorded locations are positioned within the wet zone of Sri Lanka (Figures 4A and 4B). This aligns with the general distribution pattern of the fern flora of Sri Lanka. As Jayasekera & Wijesundara (1993) reported, approximately 81% of Sri Lankan lycophytes and ferns

are confined to the wet zone ecosystems. Moreover, the wet zone, which accounts for only one-third of the country's total land area, contains all currently known endemic pteridophytes (Ranil *et al*., 2008).

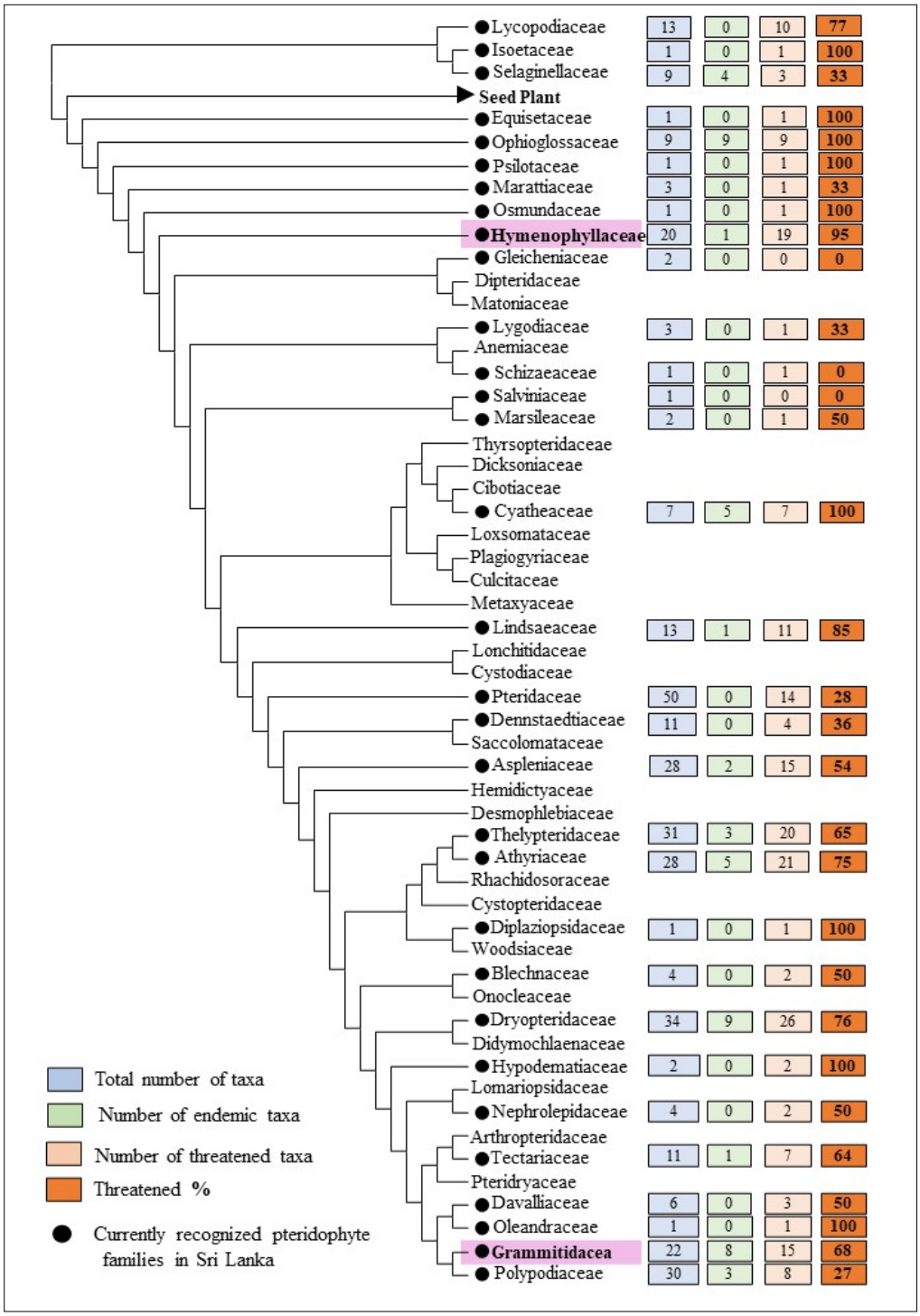


Figure 1: Global pteridophyte phylogenetic tree showing the currently recognized pteridophytes families in Sri Lanka, including the number of threatened and endemic species, with emphasis on the Grammitidaceae and Hymenophyllaceae families.

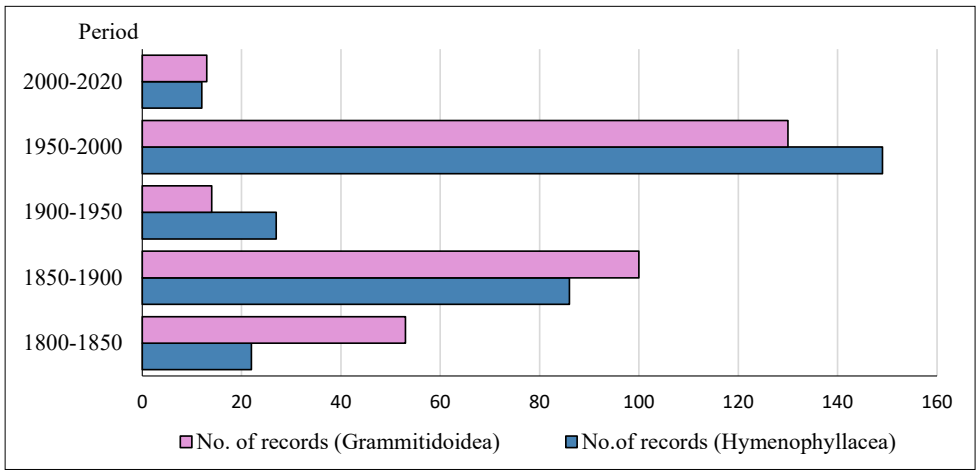


Figure 2: Temporal distribution of herbarium records of Grammitidoideae and Hymenophyllaceae in Sri Lanka from 1800 to 2020.

Figure 3: A few rare filmy and grammitid ferns recorded from Sri Lanka. **A**. *Crepidomanes minutum* (Blume) K. Iwats, a rare and endangered filmy fern species reported from the Ritigala Strict Nature Reserve. **B**. *Hymenophyllum macrogrossum* Bosch, an endemic and highly microhabitat-specific filmy fern species recorded from the Adam's Peak Forest Reserve. **C**. *Didymoglossum bimarginatum* (Bosch) Ebihara & K. Iwats, a simple-leaved and one of the smallest fern species in Sri Lanka growing on moist rocks in the Kanneliya Forest Reserve. **D**. *Prosaptia obliquata* (Blume) Mett., a locally common epilithic grammitid fern observed in the Dothalugala Forest Reserve. **E**. *Oreogrammitis medialis* (Baker) Parris, a grammitid fern species mostly restricted to moist rocks in the spray zones of fast-flowing streams in the south-western forests. **F.** *Ctenopterella blechnoides* (Grev.) Parris, an endangered grammitid fern species reported from the Kanneliya Forest Reserve.

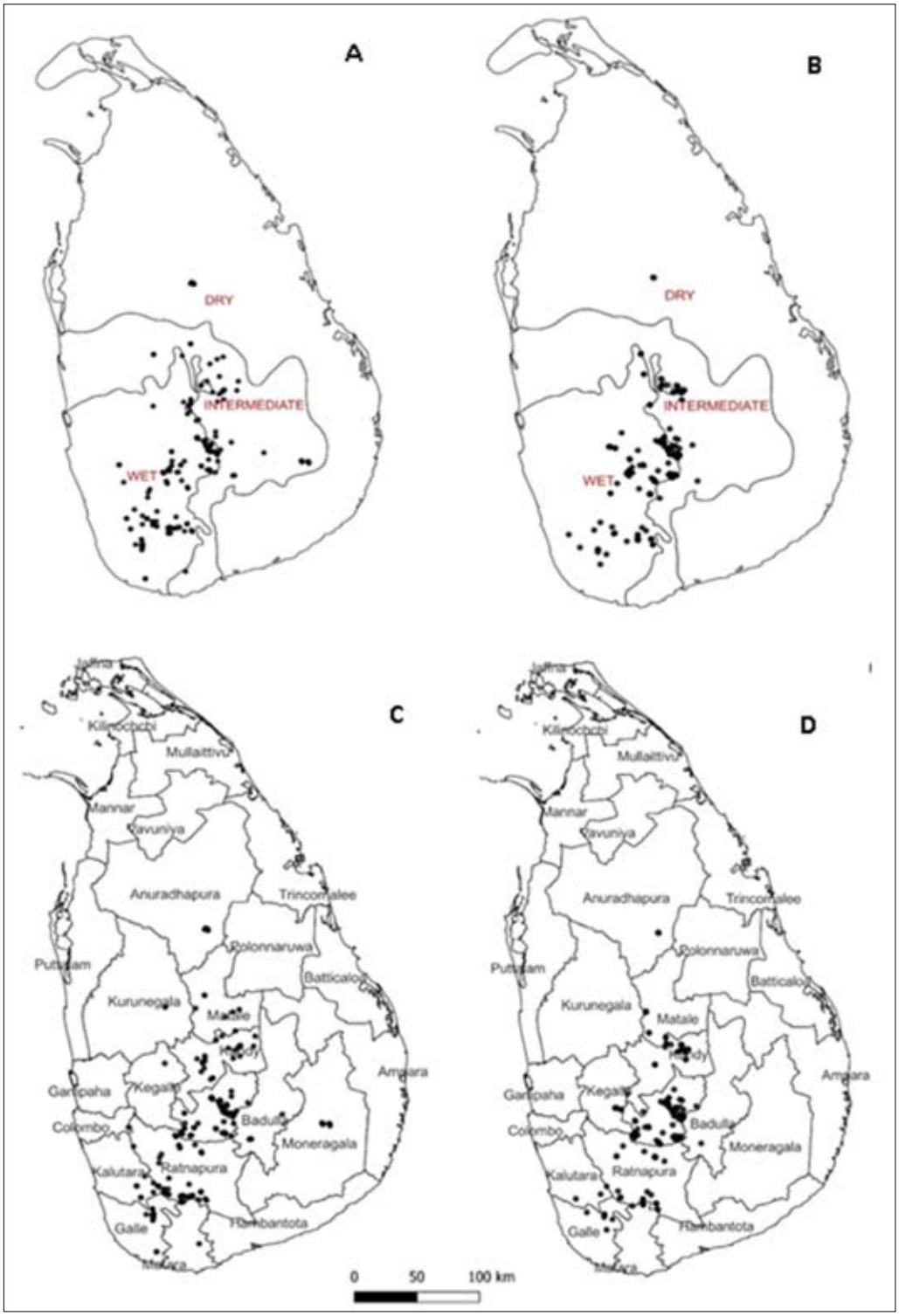


Figure 4: Spatial distribution patterns of filmy and grammitid ferns in Sri Lanka, based on 752 herbarium records housed across 22 herbaria worldwide. Map **A**. Distribution of filmy ferns in the three major climatic zones of Sri Lanka. Map **B**. Distribution of grammitid ferns in three major climatic zones of Sri Lanka. Map **C**. Spatial distribution pattern of filmy ferns in 25 administrative districts of Sri Lanka. Map **D**. Spatial distribution patterns of grammitid ferns in 25 administrative districts of Sri Lanka.

The distribution of filmy and grammitid ferns reveals some exceptional localities that are situated in the isolated and biologically diverse mountains within the dry zone and intermediate zones of Sri Lanka (Figures 4A and 4B). These distinct areas are unique habitats for these fern groups, showcasing their presence outside the typical wet zone regions. However, except *Abrodictyum obscurum* (Blume) Ebihara & K. Iwats, all other taxa of both groups thrive well in epiphytic and epilithic substrates of humid and shady environments. Moreover, according to the analysis, the *Nuwara Eliya* district offers habitats for 23 species (55%), with five species exclusively confined to this particular district (Figures 4C and 4D). Among the nine administrative provinces, the Central Province records the presence of 35 species (83%), followed by 21 species (50%) distributed in the *Sabaragamuwa* Province. The combination of factors such as the highest annual rainfall, elevated relative humidity, and favorable soil and elevational gradients explains the occurrence of the highest number of species in both provinces. This spatial distribution pattern provides insight into the rediscovery of possibly extinct species and find new locations for other endangered species to facilitate future studies on various aspects of filmy and grammitid ferns of Sri Lanka.

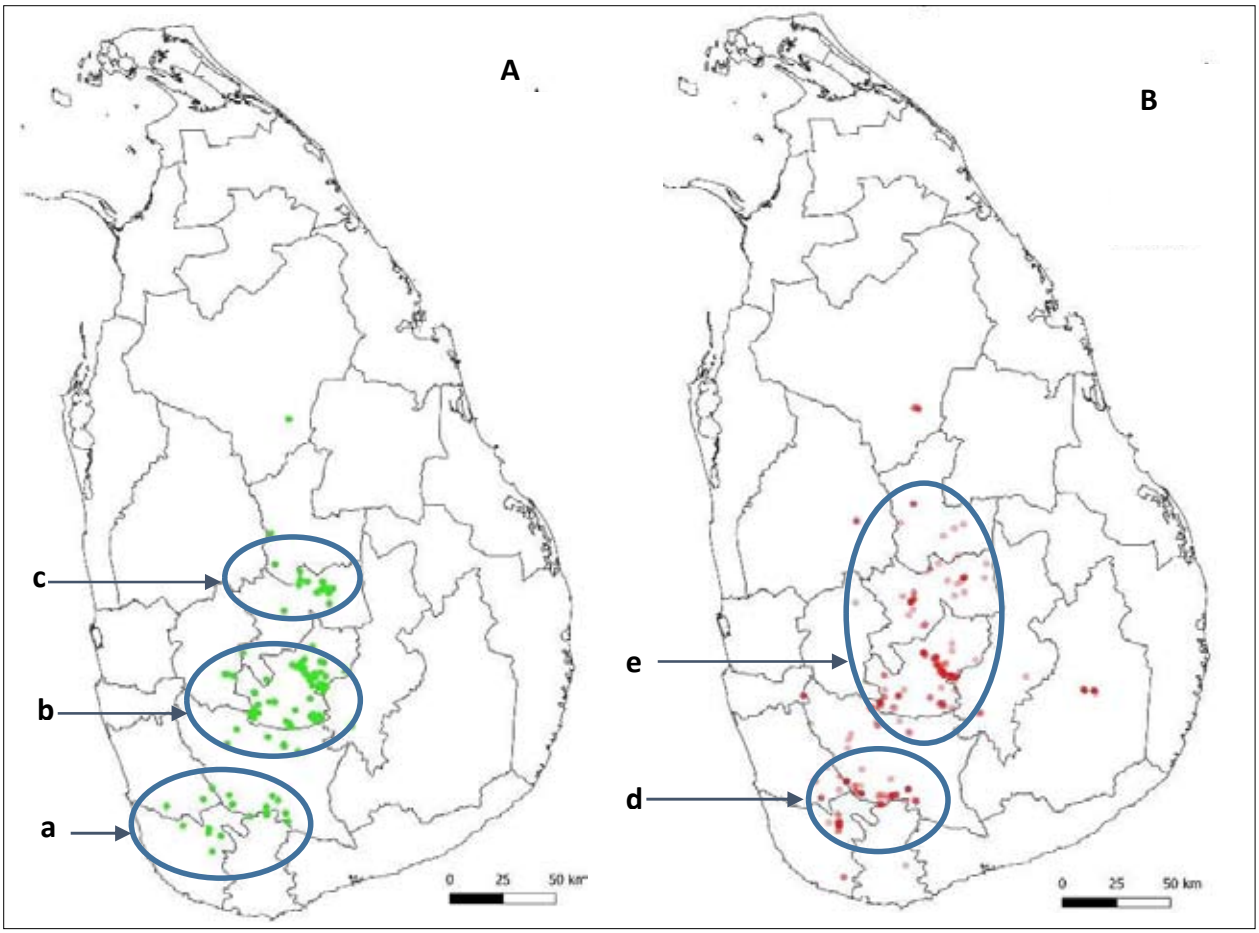


Figure 5: Map **A**. Grammitid ferns hotspots (a- lowland forests in southwest Sri Lanka: Sinharaja World Heritage Site, Kanneliya-Dediyagala-Nakiyadeniya forest complex, Gilimale forest reserve, Morapitiya-Roonakanda forest reserve, Beraliya forest reserve and Hiniduma forest reserve. b- forests of Kandy and Nuwara Eliya districts: Horton Plains National Park, Piduruthalagala mountain, Ambagamuwa forest reserve, Hakgala strict nature reserve, Hantana mountain range; Adam's Peak mountain range. c- Knuckles Mountain Range in Sri Lanka). Map **B.** Filmy ferns hotspots (e-lowland forests in southwest of Sri Lanka: Sinharaja world heritage site, Kanneliya forest reserve, Gilimale forest reserve, Morapitiya- Roonakanda forest reserve, Beraliya forest reserve and Hiniduma forest reserve). d- high altitude forests in central highlands: Knuckles Mountain Range, Adam's Peak mountain range, Horton Plains National Park, Piduruthalagala mountain, Ambagamuwa forest reserve, Hakgala strict nature reserve in Sri Lanka).

Over time, due to changes in regional landscaping for agricultural and industrial development and rapid urbanization, some previously known localities of certain species no longer exist. For example, the possibly extinct species *Chrysogrammitis glandulosa* (J.Sm.) Parris and *Calligrammitis beddomeana* (Alderw.) Parris, Sundue, Li Bing Zhang, X.M. Zhou & Ralf Knapp have been collected only from the Nuwara Eliya district, and such previous localities have undergone cultivation and various human activities. Furthermore, the endemic *Prosaptia ceylanica* Parris has only been reported from its type locality, which is now significantly disturbed. Utilizing the generated maps, revisiting previously documented localities, and exploring adjacent areas can provide valuable information and help expand these species' known areas of occurrence.

Identification of hotspots

Species distribution maps are widely used to identify priority areas based on occurrence data and environmental factors (Lannuzel *et al*., 2021). Identifying hotspots holds significant value for conservationists, policymakers, and researchers engaged in biodiversity conservation. These hotspots serve as repositories of valuable information. In this study, we utilized the available occurrence data to identify specific areas of high importance for the conservation of filmy and grammitid ferns. Based on the available data, we identified three hotspots for grammitid ferns (Figure 5A) and two hotspots for filmy ferns (Figure 5B). The extent of hotspots can be further expanded by incorporating new data on species occurrences.

Identification of preferable climatic requirements and habitats

Identifying appropriate climatic requirements is essential for determining suitable habitats and predicting future population size and distribution changes. It is also necessary to understand appropriate ecological requirements for *ex situ* conservation. Identifying suitable climates ensures successful reintroduction into their natural range or similar areas. Moreover, understanding the specific conditions a plant requires (*e.g.*, temperature, rainfall, soil type) helps prioritize the protection and restoration of suitable habitats. In this study, we have identified the preferred climatic requirements by analyzing their natural distribution patterns. Our results are primarily based on three major climatic factors: temperature, relative humidity, and annual rainfall (Figure 6). To validate these findings, screening the natural population of both plant groups across the country is essential. In the meantime, it is essential to develop probability area maps based on past and future observations because probability area maps offer a wealth of information that can help to expand the occurrences of species or groups by revealing areas with similar habitats and specific ecological niches. They offer further support to find if previously undiscovered populations or suitable habitats for target species or groups may exist. Thus, developing such probability area maps for future investigation of both plant groups is a major requirement.

Assessment of conservation status

Assessing species for their conservation needs is a vital first step in identifying and prioritizing species for *ex situ* and *in situ* conservation actions (Johnson *et al*., 2020). The first list of conservation status for the pteridophyte flora of Sri Lanka was published in 2012, followed by the second list in 2020.

The conservation status of each taxon remained relatively the same due to the lack of collections during the period between the two assessments, although conducting an island-wide survey was a key recommendation of the 2012 assessment. Moreover, both assessments were largely based on herbarium specimens of 34 worldwide herbaria (Ranil *et al*., 2022). Therefore, here we considered the 2020 list the most updated document concerning species conservation status (Figure 7).

Accordingly, 19 species of filmy ferns (95%) are treated under the threatened category. Out of 22 grammitid species, 17 (68%) are also in the same category, indicating that both groups are at a high risk of extinction. However, it is essential to emphasize the importance of conducting comprehensive field surveys across the priority areas for these species prior to compiling the next Red List. Without updated occurrence data, any conservation status assessment is worthless and likely to produce similar results as before. Comprehensive analysis and a systematic island-wide field survey are essential for policymakers and conservationists in decision-making. Such updated assessments also provide researchers with guidance for setting future research goals, specifically in prioritizing species for conservation within these two highly threatened plant groups of Sri Lanka. This prioritization is necessary because resources for nature conservation are limited (Arponen, 2012). However, the conservation status given in the 2020 list provides adequate support to prioritize the species for future conservation efforts.

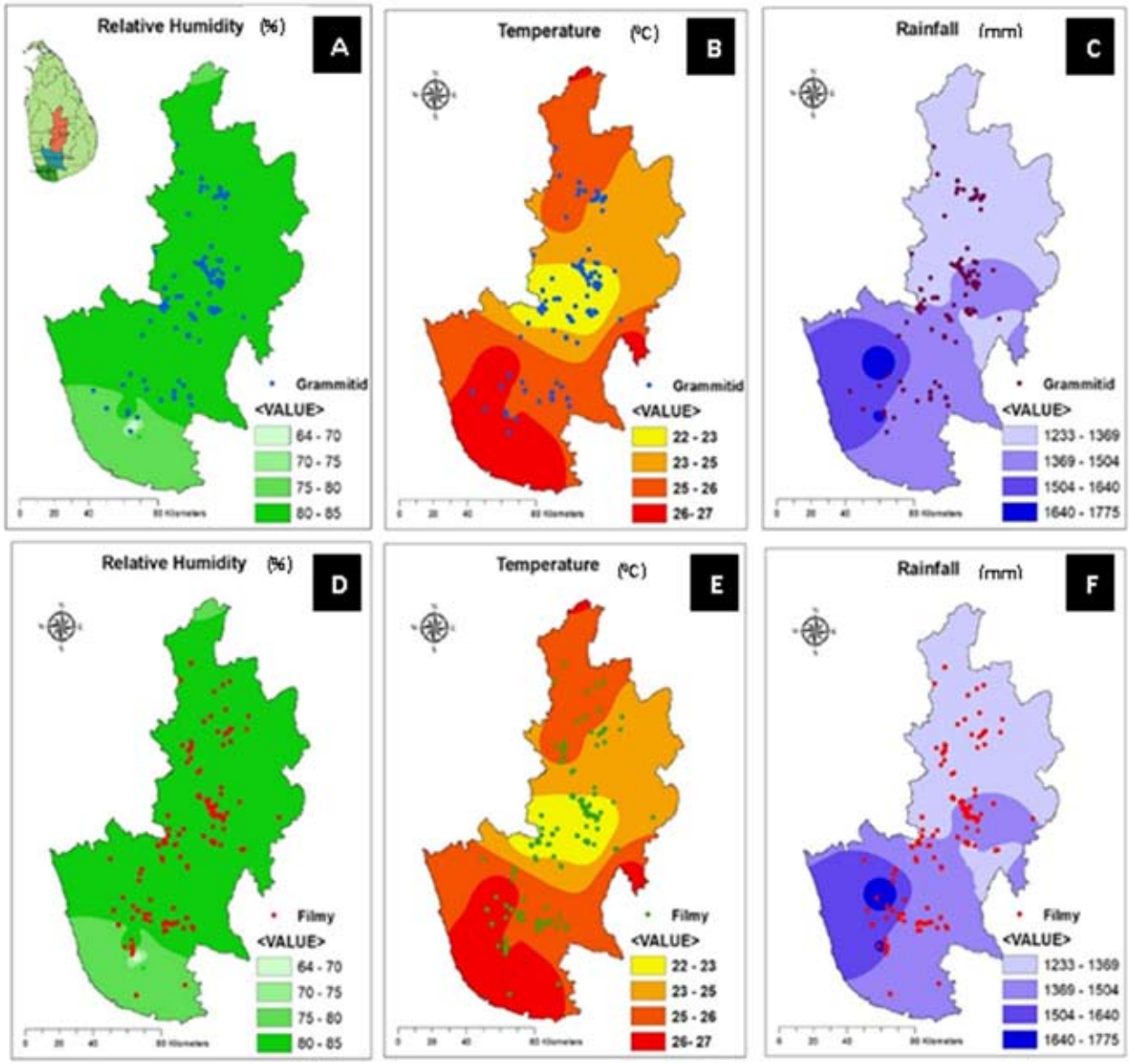


Figure 6: Maps A–C show the most preferred relative humidity, temperature, and rainfall ranges for the occurrence of grammitid ferns. Maps D–F show the corresponding preferred ranges for filmy ferns. These patterns are illustrated for the districts of Galle, Matara, Ratnapura, Kandy, and Matale in Sri Lanka.

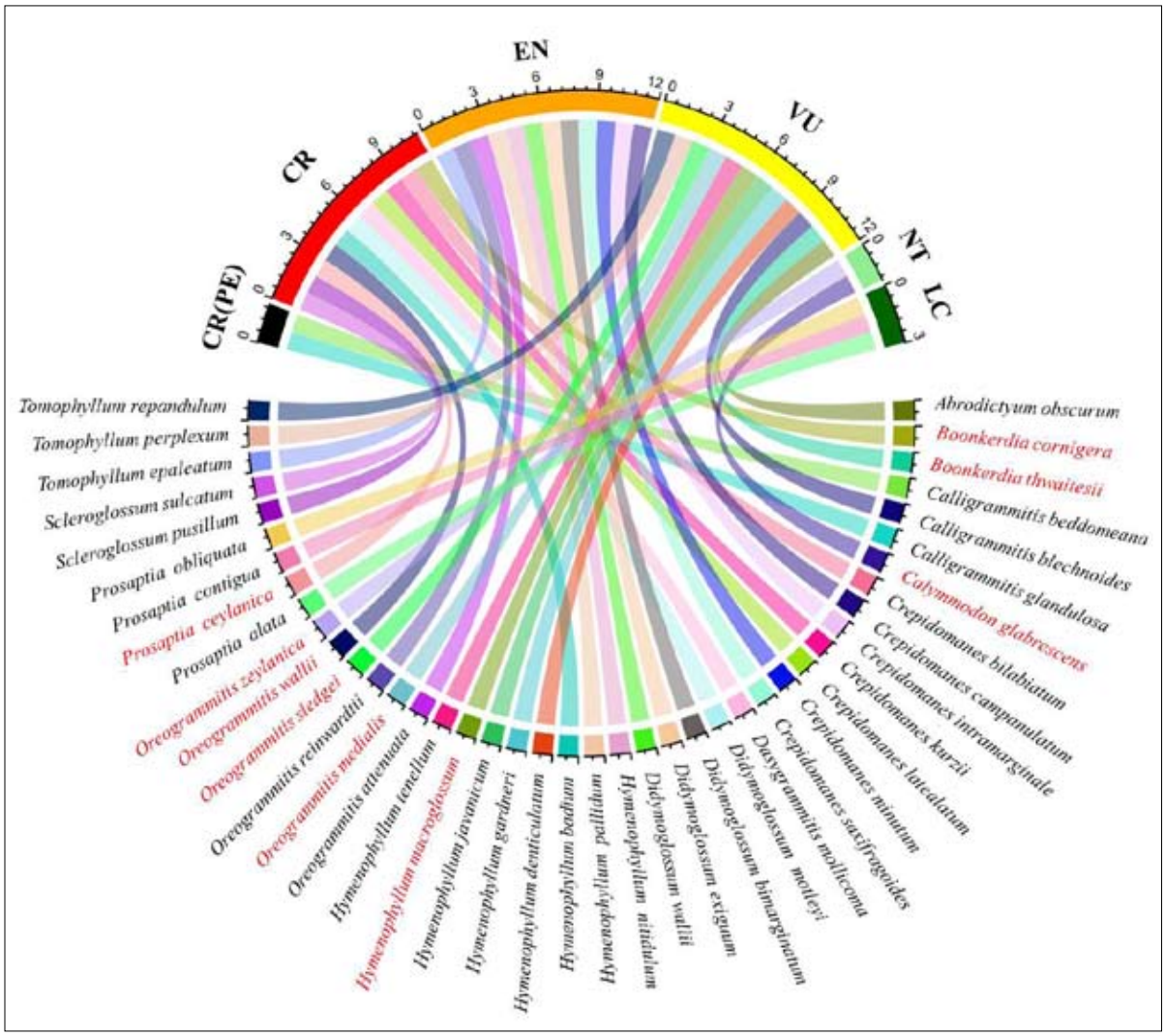


Figure 7: Conservation status of 42 species in the subfamily Grammitidoideae and the family Hymenophyllaceae (based on the 2020 National Red List). The botanical names highlighted in red indicate that they are endemic to Sri Lanka.

Species selection and prioritization for developing a conservation plan

The criteria used for species-level conservation prioritization are multidimensional and range from aesthetic to evolutionary considerations (Arponen, 2012; Mee *et al*., 2015). Given that both filmy and grammitid ferns in Sri Lanka have yet to be extensively studied regarding their ecology, reproductive biology, phylogeny, and population dynamics, the IUCN conservation status is the sole criterion for prioritizing species for future conservation efforts. We highly recommend implementing urgent conservation measures for all critically endangered and potentially extinct species, considering them as top priorities for intervention. Table 1 provides a list of high-priority species for setting future conservation guidelines.

Hereafter, we discuss the areas that need to be integrated into conservation planning after selecting specific species. Although Sri Lanka lags behind in adopting the fields mentioned below due to a lack of general interest, financial limitations, inadequate technical expertise, and dedicated personnel, it is crucial to include these areas in the country's future research mandate to safeguard these highly endangered species.

Table 1: The species required immediate attention for conservation.

Species	Conservation status	Last record	Herbaria
**Boonkerdia cornigera*	CR	1950	K, L, PDA,
Calligrammitis beddomeana	CR (PE)	1849-1888	BO, BM, K, PDA
Chrysogrammitis glandulosa	CR (PE)	1849	B, BM, CGE, K, PDA
Crepidomanes bilabiatum	CR	1996	BM, K, L, PDA
Crepidomanes kurzii	CR	1954	BM, K, L
Crepidomanes latealatum	CR	-	K
Dasygrammitis mollicoma	CR	1843-1849	K, GH, UC, US
Didymoglossum motleyi	CR	1996	PDA, BM, K
Hymenophyllum badium	CR	-	K
**Oreogrammitis wallii*	CR	1954	B, BM,GH, K,L,PDA, US
**Prosaptia ceylanica*	CR	1954	K
Scleroglossum pusillum	CR	1870-1872	E
Scleroglossum sulcatum	CR	1889	B, BM, CGE, E, GH, K, GH, P, PDA, US

Note: *species endemic to Sri Lanka

Detailed and systematic studies of selected and prioritized species

After selecting a species or species group, conducting in-depth, systematic study of various aspects will guide the accurate application of effective conservation measures. Conducting comprehensive studies in these areas is essential to implement an effective and efficient conservation approach. Accommodating recent knowledge such as biotechnology, GIS applications, and mathematical models is also important when necessary. Apart from having a taxonomically well-focused list and old herbarium specimens, particularly in European herbaria, Sri Lanka has not demonstrated significant interest in studying species within these two groups thus far. Therefore, the highest priority should be given to surveying the country's natural ecosystems to discover new localities and expand the areas of occurrence, followed by strengthening the national herbarium collections.

The family Hymenophyllaceae and subfamily Grammitidoideae have been extensively studied for global phylogenies. Some Sri Lankan non-endemic grammitid species have been included in these studies: *Calligrammitis beddomeana, Chrysogrammitis glandulosa, Ctenopterella blechnoides, Dasygrammitis mollicoma, Prosaptia alata, P. contigua, P. obliquata, Scleroglossum pusillum,* and *S. sulcatum* (Zhou et al., 2023), although Sri Lankan material has not been sampled. However, Sri Lankan collections of the endemic grammitids *Oreogrammitis medialis* and *Tomophyllum perplexum* have been sampled, together with *T. repandulum*, a species described from Sri Lanka (Zhou et al., 2023). As an isolated gene pool in the Indian Ocean, it is important to elucidate the position of all Sri Lankan Hymenophyllaceae and Grammitidoideae species in the global phylogeny, utilizing Sri Lankan material to ascertain the relationships of the endemics and to confirm the identification of non-endemic species typified from elsewhere. Therefore, future studies should also focus on genomic analysis and biogeographic aspects as well. Given the current scenario, the only feasible ways to advance the study and conservation approach of fern species in Sri Lanka are through networking with experts in pteridology and establishing effective collaborations with international institutions.

In addition to the areas discussed above, we emphasize the significance of ethnobotanical knowledge associated with these species. According to Ranil and Bussmann (2021), traditional uses of filmy and grammitid ferns have not been reported in Sri Lanka. However, the communities residing in the vicinity of the wet zone forests possess a deep understanding and knowledge of the habitats, particularly the ecological niches of wild species. Their expertise in ethnoforestry and ethnoecology can therefore be integrated to explore new localities for these species. Today, the use of citizen science in biodiversity conservation projects is growing

in number, size, and scope, and it is gaining recognition as a valuable data source that builds public engagement (Burgess et al., 2017). In Sri Lanka, there is a growing interest among the younger generation, particularly those involved in local environmental organizations and societies, in exploring plant diversity, discovering rare or new species, and engaging in activities such as nature photography. Moreover, they extensively utilize social media platforms to share their knowledge and experiences while networking with professionals and the international community. Since community engagement is a key component of effective species conservation, policymakers and conservationists also need to pay special attention to integrating community knowledge, participation, and citizen science initiatives into future conservation strategies.

Mass propagation of selected species for conservation

Mass propagation of endangered fern species has been practiced for several decades using both conventional and micro-propagation techniques (Fay, 1992; Amoroso & Amoroso, 2003; Manickam et al., 2003; Chiou et al., 2006; Magrini et al., 2020; Suneetha & Hegde, 2022). Although micro-propagated plants may exhibit reduced genetic variability within their homogeneous populations, this approach can still be considered a viable solution for species with limited population sizes. However, the importance of *in vitro* tools to complement other *ex situ* methods for saving plants from extinction is more relevant than ever before (Barnicoat et al., 2011). *In vitro* propagation of the endangered filmy fern, *Trichomanes punctatum* subsp. *floridanum* revealed that *in vitro* propagation methods could be used to provide plants for restoration (Pence, 2014). Considering the widespread popularity of Polypodiaceae species among plant growers and commercial nurseries, grammitid ferns, being part of the same family clade, emerge as promising candidates for various horticultural and commercial endeavors.

However, both *in situ* and *ex situ* conservation pose a significant challenge for this specific group of ferns. Both fern groups studied here have chlorophyllous spores, making their cultivation particularly complex (Lloyd & Klekowski, 1970). These spores are typically adapted to specific ecological niches where they can access light and moisture immediately after dispersal, conditions that are essential for their survival and germination. Chlorophyllous spores are short-lived (Mellado-Mansilla et al., 2022), a characteristic that has puzzled many pteridologists. Their short lifecycle ensures rapid germination in suitable conditions, providing a competitive advantage in dynamic environments like forest canopies. However, this trait also limits their viability under less favorable circumstances. Further research in this area is essential, as it could significantly contribute to developing localized culture protocols for propagating filmy and grammitid ferns. Such advancements would play a vital role in their conservation and sustainable management.

***In situ* conservation**

The *in situ* conservation approach can be implemented through the reintroduction or translocation of species to suitable habitats and the conservation of existing populations, along with the safeguarding of the entire habitat, as explained below. However, these methods have yet to be applied in Sri Lanka for any fern species.

(a) Reintroduction

There is substantial evidence supporting the successful reintroduction of endangered ferns into suitable habitats (McHaffie, 2006; Aguraiuja, 2011; Baker et al., 2014). For example, Baker et al. (2014) provide valuable insights into the rescue, ecology, and conservation of the rediscovered island endemic *Anogramma ascensionis* through *ex situ* approaches, and propose a roadmap for species reintroduction and habitat restoration. However, evidence for the successful reintroduction of filmy and grammitid ferns in particular is limited and remains poorly documented. Their epiphytic and epilithic growth forms, narrow habitat preferences, and strong sensitivity to microclimatic variation make reintroduction particularly difficult and technically demanding. Consequently, any reintroduction programme involving these groups must emphasize careful site selection, microhabitat restoration, and long-term monitoring of population dynamics, habitat conditions, and adaptive responses. Despite these limitations, well-designed reintroduction efforts still have the potential to make meaningful contributions to the conservation of rare and threatened filmy and grammitid ferns, particularly those with restricted distributions and small population sizes.

(b) Conservation of existing populations

The outcome of hotspot analysis provides valuable insights in identifying population hotspots for individual species or groups. To achieve this objective, innovative approaches like mathematical modeling can also be employed. These methods hold significant potential for enhancing our understanding and identification of hotspots for candidate species. If these areas are not currently part of the country's protected area network, declaring them as conservation areas should align

with the policy and legislation of the country, with the involvement of responsible government institutions.

However, some of the sensitive areas and hotspots identified in this study directly overlap with the existing protected area network; unless influenced by the impacts of climate change, they are not highly subject to other external pressures. This indicates that they benefit from sufficient legal protection. However, future attention and conservation priorities should focus on hotspots outside the country's protected area network. Successful conservation plans are not solely achieved by acquiring optimally designed reserves, and ongoing monitoring and managing the biodiversity in those reserves are an equally important, but often neglected or poorly executed, part of the conservation process (Regan *et al*., 2008). Therefore, conducting scientific investigations, accompanied by continuous and long-term monitoring of these identified hotspots, will help minimize the risk of extinction for these particular groups of plants in Sri Lanka.

Apart from accommodating academic disciplines, coordinating government authorities aligned with the country's biodiversity conservation effort is a major requirement. Since the country's protected area network belongs to the central government, direct government involvement is a determining factor for the effectiveness of any conservation plan. Even though public involvement is a major element of the integrative conservation approach, both filmy and grammitid ferns do not show community engagement due to their low economic potential and use value. However, it is mandatory to be aware of the communities around the hotspots and sensitive habitats for these two plant groups to ensure the conservation of the entire ecosystem.

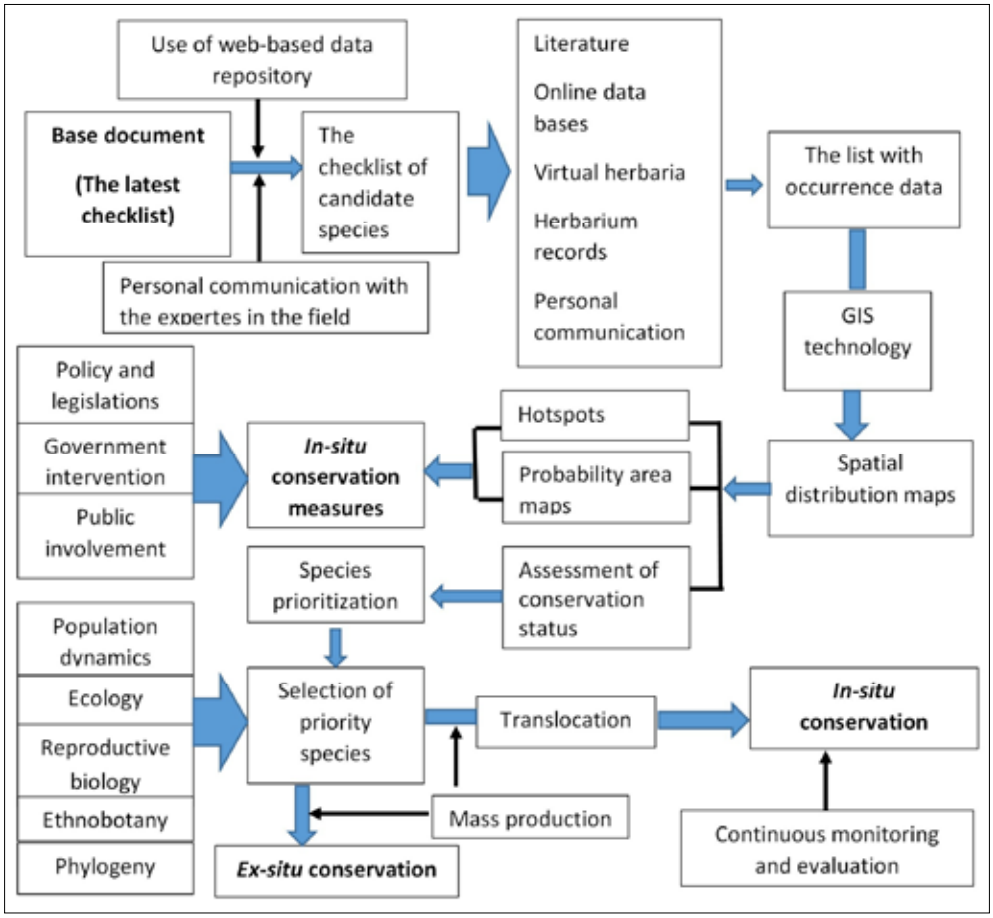


Figure 8: A schematic representation of the diverse fields, elements, and tools employed in a conservation plan.

Ex situ conservation

Although *ex situ* conservation is challenging for both plant families, it may still be possible to cultivate them in high-tech greenhouses where their preferred microclimatic conditions can be externally provided and maintained. Apart from its conservation significance, *ex situ* conservation is a key element for raising public awareness about such rare and endangered taxa, though it is costly. Since both filmy and grammitid ferns naturally grow in more humid and shaded conditions, it is a challenge for countries like Sri Lanka to conserve such species in an *ex situ* environments for a long time, particularly due to the high establishment and maintenance costs. Therefore, rather than prioritizing *ex situ* conservation of filmy and grammitid ferns, *in situ* conservation is generally more practical and effective.

CONCLUSION

A holistic conservation approach is essential to conserve both filmy and grammitid ferns, particularly given their climatic sensitivity, habitat specificity, and narrow distribution. However, Sri Lanka has made limited application of holistic conservation methods for these two plant groups due to the lack of recent studies on their distribution patterns, ecological aspects, population dynamics, and phylogeny. This analysis relied heavily on historical literature and herbarium records from the colonial era. This study presents detailed information on filmy and grammitid ferns and proposes conservation options, serving as a valuable resource for conservationists and policymakers to develop new conservation plans or implement conservation actions. The schematic diagram (Figure 8) illustrates the pathway for conserving these two fern families. The developed distribution maps and highlighted areas requiring further research provide valuable insights for effective conservation strategies. This conservation protocol could also be applied to other threatened plant groups. Strengthening institutional capacities and garnering support from government authorities will establish a strong foundation for conserving Sri Lanka's unique biodiversity. Engaging and educating local communities about hotspots and sensitive habitats are crucial for conserving the entire ecosystem. Aligning conservation research and projects with global strategic plans and fostering long-term collaborations with international organizations and research institutes are essential for successful conservation efforts.

REFERENCES

Aguraiuja, R. (2011). Reintroduction of the endangered fern species *Woodsia ilvensis* to Estonia: a long-term pilot study. *Biodiversity Conservation*, *20*, 391-400. https://doi.org/10.1007/s10531-010-9970-2

Amoroso, C.B., & Amoroso, V.B. (2003). Plantlet production of the Philippine giant staghorn fern [*Platycerium Grande* (Fee) C. Presl] through spore culture. In: *Pteridology in the New Millennium* (eds. S. Chandra, M. Srivastava), pp. 491-495. Springer, Dordrecht, Netherlands. https://doi.org/10.1007/978-94-017-2811-9_33

Aros-Mualin, D. & Kessler, M. (2024). Untangling poikilohydry and desiccation tolerance: evolutionary and macroecological drivers in ferns, *Annals of Botany*, *134* (7), 1139–1150. https://doi.org/10.1093/aob/mcae167

Arponen, A. (2012). Prioritizing species for conservation planning. *Biodiversity Conservation*, *21*, 875–893. https://doi.org/10.1007/s10531-012-0242-1

Baker, K., Lambdon, P., Jones, E, Pellicer, J., Stroud, S., Renshaw, O., Niissalo, M., Corcoran, M., Clubbe, C. & Sarasan, V. (2014). Rescue, ecology, and conservation of a rediscovered island endemic fern (*Anogramma ascensionis*): *ex situ* methodologies and a roadmap for species reintroduction and habitat restoration. *Botanical Journal of the Linnean Society*, *174*, 461-477. https://doi.org/10.1111/boj.12131

Barnicoat, H., Cripps, R., Kendon, J. & Sarasan, V. (2011). Conservation *in vitro* of rare and threatened ferns—case studies of biodiversity hotspot and island species. *In Vitro Cellular Developmental Biology-Plant*, *47*, 37–45. https://doi.org/10.1007/s11627-010-9303-x

Burgess, H., DeBey, L., Froehlich, H., Schmidt, N., Theobald, E., Ettinger, A., HilleRisLambers, J., Tewksbury, J. & Parrish, J. (2017). The science of citizen science: Exploring barriers to use as a primary research tool. *Biological Conservation*, *208*, 113-120. https://doi.org/10.1016/j.biocon.2016.05.014

Chiou, W.L., Huang, Y. M. & Chen, C.M. (2006). Conservation of two endangered ferns, *Archangiopteris somae* and *A. itoi* (Marattiaceae: Pteridophyta), by propagation from stipules. *Fern Gazette*, *17*(5), 271-278.

Della, A.P. (2022). Ferns as ecological indicators. In: *Ferns* (eds. J. Marimuthu, H. Fernández, A. Kumar, S. Thangaiah), pp. 587–601. Springer, Singapore. https://doi.org/10.1007/978-981-16-6170-9_25

Fay, M.F. (1992). Conservation of rare and endangered plants using *in vitro* methods. *In Vitro Cellular Developmental Biology-Plant*, *28*, 1-4. https://doi.org/10.1007/BF02632183

Frankham, R. (1997). Do island populations have less genetic variation than mainland populations? *Heredity*, *78*, 311–327. https://doi.org/10.1038/hdy.1997.46

Jaisankar, I., Velmurugan, A. & Sivaperuman, C. (2018). Biodiversity conservation: issues and strategies for the tropical islands. In: *Biodiversity and climate change adaptation in tropical islands* (eds. C. Sivaperuman, A. Velmurugan, I. Jaisankar), pp. 525-552. Academic Press. https://doi.org/10.1016/B978-0-12-813064-3.00019-3

Gunatilleke, S., Gunatilleke, N. & Pethiyagoda, R. (2008). Biodiversity of Sri Lanka. *Journal of the National Science Foundation of Sri Lanka*. *36*, 25-62.

Hamabata, T., Kinoshita, G., Kurita, K., Cao, P.L., Ito, M., Murata, J., Komaki, Y., Isagi, Y. & Makino, T. (2019). Endangered island endemic plants have vulnerable genomes. *Communication Biology*, *2*, 1-10. https://doi.org/10.1038/s42003-019-0490-7.

Jayasekara, P. (2006). Hymenophyllaceae. In: *A revised handbook to the Flora of Ceylon: Pteridophyta (Ferns and Fern allies)* (ed. M. Shaffer-Fehre), pp. 214-235. Science Publishers, New Hampshire, USA.

Jayasekera, P.W.B. & Wijesundara, D.S.A. (1993). A herbarium survey of Pteridophytes of Sri Lanka. *Proceedings of the 49th Annual Session of the Sri Lanka Association for the Advancement of Science*, Vidya Mandiraya, Vidya Mawatha, Colombo, Sri Lanka, p. 66.

Johnson, K., Baker, A., Buley, K., Carrillo, L., Gibson, R., Gillespie, G.R., Lacy, R.C. & Zippel, K. (2020). A process for assessing and prioritizing species conservation needs: going beyond the Red List. *Oryx*, *54*, 125-132. https://doi.org/10.1017/S0030605317001715

Lannuzel, G., Balmot, J., Dubos, N., Thibault, M. & Fogliani, B. (2021). High-resolution topographic variables accurately predict the distribution of rare plant species for conservation area selection in a narrow-endemism hotspot in New Caledonia. *Biodiversity Conservation*. *30*, 963–990. https://doi.org/10.1007/s10531-021-02126-6

Lloyd, R.M. & Klekowski, E.J. (1970). Spore germination and viability in Pteridophyta: evolutionary significance of chlorophyllous spores. *Biotropica*, 129-137. https://doi.org/10.2307/2989770

Manickam, V.S., Vallinayagam, S. & Johnson, M. (2003). Micropropagation and conservation of rare and endangered ferns of the Southern Western Ghats through *in vitro* culture. In: Pteridology in the new millennium (eds. S. Chandra, M. Srivastava). pp. 497-504. Springer, Dordrecht, Netherlands. https://doi.org/10.1007/978-94-017-2811-9_34

Magrini, S., Azzella, M.M., Bolpagni, R. & Zucconi, L. (2020). *In vitro* propagation of *Isoëtes sabatina* (Isoetaceae): A key conservation challenge for a critically endangered quillwort. *Plants*, *9*(7), 887. https://doi.org/10.3390/plants9070887

McHaffie, H. (2006). A reintroduction programme for *Woodsia ilvensis* (L.) R. Br. in Britain. *Botanical Journal of Scotland*, *58*(1), 75-80. https://doi.org/10.1080/03746600608685109

Mee, J.A., Bernatchez, L., Reist, J.D., Rogers, S.M. & Taylor, E.B. (2015). Identifying designable units for intraspecific conservation prioritization: a hierarchical approach applied to the lake whitefish species complex (Coregonus spp.). *Evolutionary Applications*, *8*(5), 423-441. https://doi.org/10.1111/eva.12247

Mehta, P., Sekar, K.C., Bhatt, D., Tewari, A., Bisht, K., Upadhyay, S., Negi, V.S. & Soragi, B. (2020). Conservation and prioritization of threatened plants in Indian Himalayan Region. *Biodiversity Conservation*, *29*, 1723-1745. https://doi.org/10.1007/s10531-020-01959-x

Mellado-Mansilla, D., Testo, W., Sundue, M.A., Zotz, G., Kreft, H., Coiro, M. & Kessler, M., (2022). The relationship between chlorophyllous spores and mycorrhizal associations in ferns: evidence from an evolutionary approach. *American Journal of Botany*, *109*(12), 2068-2081.

Nogué, S., de Nascimento, L., Froyd, C.A., Wilmshurst, J.M., de Boer, E.J., Coffey, E.E., Whittaker, R.J., Fernández-Palacios, J.M. & Willis, K.J. (2017). Island biodiversity conservation needs palaeoecology, *Nature ecology & evolution*, *1*(7), 0181. https://doi.org/10.1038/s41559-017-0181

Oro, D., Pueyo, Y., Bauzà, J., Errea, M.P. & Arroyo, A.I. (2022). Long transient response of vegetation dynamics after four millennia of anthropogenic impacts in an island ecosystem. *Global Change Biology*, *28*(21), 6318-6332. https://doi.org/10.1111/gcb.16363

Parris, B. (2006). Grammitidaceae. In: *A revised handbook to the Flora of Ceylon: Pteridophyta (Ferns and Fern allies)* (ed. M. Shaffer-Fehre), pp. 214-235. Science Publishers, New Hampshire, USA.

Pence, V.C. (2014). *In vitro* propagation and cryopreservation of the endangered filmy fern, *Trichomanes punctatum* subsp. *floridanum* (Hymenophyllaceae). *Fern Gazette*, *19*, 307-317.

Pethiyagoda, R. & Sudasinghe, H. (2017). The development of Sri Lankan biogeography in the colonial period. *Ceylon Journal of Science*, *46*(5), 5-18. http://doi.org/10.4038/cjs.v46i5.7451

PPG I. (2016). A community-derived classification for extant lycophytes and ferns. *Journal of systematics and evolution*, *54*(6), 563-603.

Ranil, R.H.G., Pushpakumara, D.K.N.G. & Wijesunda, D.S.A. (2008). Present status of taxonomic research and conservation of endemic pteridophytes in Sri Lanka. *Proceedings of the 4th Symposium on Asian Pteridology and Garden Show* (ed. V.B Amoroso), Central Mindanao University, Musuan, Bukidnon, Philippines, pp. 84–93.

Ranil, R.H.G. & Bussmann, R.W. (2021). Potential uses of lycophytes and ferns in Sri Lanka: an ethnopteridological perspective. *Ethnobotany Research and Applications*, *21*(36), 1-11. https://ethnobotanyjournal.org/index.php/era/article/view/2761

Ranil, R.H.G., Fraser-Jenkins, C.R., Ebihara, E, *et al.* (2022). An updated checklist of the pteridophyte flora of Sri Lanka. Gampaha, Sri Lanka: ECERO Publishers and Distributors.

Rajapaksha, R., Zhang, L.B., Pushpakumara, G. & Wijesundara, S. (2022). An analysis of the current status and future prospects of Sri Lankan pteridophytes towards a new dimension. *Biologia*, *77*(12), 3333–3351. https://doi.org/10.1007/s11756-022-01139-y

Regan, H.M., Hierl, L.A., Franklin, J., Deutschman, D. H., Schmalbach, H.L., Winchell, C.S. & Johnson, B.S. (2008).

Species prioritization for monitoring and management in regional multiple species conservation plans. *Diversity and Distributions*, *14*(3), 462-471. https://doi.org/10.1111/j.1472-4642.2007.00447.x

Rønsted, N., Walsh., S.K., Clark, M., Edmonds, M., Flynn, T., Heintzman, S., Loomis, A., Lorence, D., Nagendra, U., Nyberg, B. & Opgenorth, M. (2022). Extinction risk of the endemic vascular flora of Kauai, Hawaii, based on IUCN assessments. *Conservation Biology*, 36(4), 13896. https://doi.org/10.1111/cobi.13896

Sarathchandra, C., Abebe, Y.A., Wijerathne, I.L., Aluthwattha, S.T., Wickramasinghe, S. & Ouyang, Z. (2021). An Overview of Ecosystem Service Studies in a Tropical Biodiversity Hotspot, Sri Lanka: Key Perspectives for Future Research. *Forests*. *12*(5), 540. https://doi.org/10.3390/f12050540

Suneetha, C. & Hegde, S. (2022). Micropropagation of Pteridophytes. In: *Ferns: Biotechnology, Propagation, Medicinal Uses and Environmental Regulation* (eds. J. Marimuthu, H. Fernández, A. Kumar, S. Thangaiah), pp. 201-242, Springer, Singapore. https://doi.org/10.1007/978-981-16-6170-9_9

Woinarski, J.C.Z. (2010). Biodiversity conservation in tropical forest landscapes of Oceania. *Biological Conservation, 143*(10), 2385-2394. https://doi.org/10.1016/j.biocon.2009.12.009

Zhou, X.M., Yang, J.J., Liang, Z.L., Pollawatn, R., Knapp, R., Parris, B., Sundue, M., Ranker, T. A., Zhou, L., Lu, N.T. & Luong, T.T. (2023). A global phylogeny of grammitid ferns (Polypodiaceae) and its systematic implications. *Taxon*. *72(5),* 974-1018. https://doi.org/10.1002/tax.12992

J.Natn.Sci.Foundation Sri Lanka 2026 54 (1): 63 - 75
DOI: http://dx.doi.org/10.4038/jnsfsr.v54i1.12563

RESEARCH ARTICLE

Animal Science

Kinniya cattle: A locally bred draught cattle population in Sri Lanka

KGCB Wijebandara[1*], M Hans[2], I Ratheefa[2], MS Kurukulasuriya[3] and GLLP Silva[4]

[1] *Postgraduate Institute of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka.*
[2] *Government Veterinary Surgeon's Office, Kinniya, Sri Lanka.*
[3] *Department of Animal Science, Faculty of Animal Science and Export agriculture, Uva Wellassa University of Sri Lanka, Badulla, Sri Lanka.*
[4] *Department of Animal Science, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka.*

Submitted: 14 February 2025; Revised: 19 November 2025; Accepted: 27 January 2026

Abstract: Indigenous cattle in Sri Lanka consists of several geographically isolated populations. Due to lack of descriptive information, they remained non-descriptive. With isolated breeding with directional selection, some of these populations have developed unique features allowing the opportunity to establish indigenous breeds. The Kinniya cattle population is one such breed, which is featured by unique selection and breeding system under specific management circumstances. This study was carried out to identify the unique characteristics in Kinniya cattle, while unveiling the associated social, cultural and economic background. Comparison of the distribution and characteristics of Kinniya cattle of those with different indigenous cattle populations in Sri Lanka was done using descriptive statistics. The results revealed that the Kinniya cattle are reared extensively in large herds, and bull calves are selected for well-defined draught characters at three months of age. Having undergone directional selection for generations, the population has now stabilized to feature well developed horns with uniform orientation, prominent hump and dewlap. The comparison with other populations revealed that, despite the ancestral similarities that could exist with the Thamankaduwa cattle, Kinniya cattle showed a distinct morphometric attribute. The study disclosed the directional selection and breeding strategies adopted by the communities over generations, and the distinct herd management strategies adopted have helped developing the uniqueness of Kinniya cattle as a separate indigenous population in Sri Lanka and the only population bred for draught purpose.

Keywords: Draught animals, indigenous cattle, Kinniya cattle, phenotypic characterization.

INTRODUCTION

Sri Lanka is an agricultural country possesing a variety of natural resources and a rich biodiversity. Crop and livestock farming is one of the oldest economic activities in Sri Lanka and it has been the main livelihood since ancient civilizations in the country. Livestock production performs a key role in national agricultural sector, within which cattle are the prominent contributor (Chandrasiri, 2002). The total cattle population in Sri Lanka in 2023 has been estimated as 1.2 million, where majority of them (0.83 million) are indigenous (Department of Census and Statistics, 2023). Sri Lankan indigenous cattle are of the Zebu type, which being raised over generations and are well adapted to harsh rural environments, especially with heat tolerance and disease resistance (Silva et al., 2010). Even though these cattle are triple purpose (milk, draught and meat) animals, they are poor milk producers. Irrespective of the purpose of keeping, adaptation to the indigenous harsh environment, especially under the rural smallholder system, is a significant feature (Silva et al., 2008).

* Corresponding author (chamodbhashitha138@gmail.com; https://orcid.org/0009-0003-6684-6244)

Some Sri Lankan cattle, are differentiated as distinct indigenous types as Batu, Thawalam, Thamankaduwa and Cape/Hatton cattle. Of these, Cape cattle is now considered as an extinct (Chandrasiri, 2002). These indigenous cattle types can be identified as geographically distinct populations such as North Central, Northern, Southern, Thamakaduwa and Thawalam which represent the general indigenous cattle populations. Despite geographical separations and phenotypic variations, indigenous cattle remain as non-descript populations owing to gaps in information generated through systematic investigation (Silva et al., 2010). In addition to the clearly differentiated indigenous types, there are isolated groups in Sri Lanka, selected over generations. These populations are managed for specified attributes, and are found in limited numbers in isolated pockets. Kinniya cattle are one such population, which are specially used for draught purpose. Kinniya cattle, regionally called 'Naattu maadu' (Vernacular Tamil name for local cattle) are found in the Kinniya, Muthur and Verugal regions, where Kinniya serves as the predominant distribution center. Kinniya cattle are distinguished from other indigenous cattle populations mainly by the specified morphological features; especially the prominent and long horns, well developed hump and dewlap of the males, and the white coat colour. A local pastoral community in the Kinniya region has reared cattle over generations. This suggests that Kinniya cattle have evolved through the selection of early Indian White cattle which existed in the region. Still, selection can be observed in the Kinniya population, where cows are bred with selected bulls which have prominent draught morphology (Wijebandara & Silva, 2024).

Analyzing the indigenous cattle and their associated sociocultural environment is important for understanding the dynamics of the production environment. This will also support the global effort in describing indigenous farm animal genetic resources aiming to provide society with a greater range of options to meet future challenges (FAO, 2007). Hence, this work was carried out to identify the population specific phenotypic characteristics of the Kinniya cattle, and their uniqueness compared to other indigenous cattle populations in Sri Lanka, while exploring the social, cultural and economic background associated with the indigenous cattle production in Kinniya region of Sri Lanka.

MATERIALS AND METHODS

Sampling location and selection of Kinniya cattle

Kinniya cattle for phenotypic characterization were selected based on the population distribution data gathered from the regional office of Kinniya Veterinary Division, Sri Lanka. There were 20–30 isolated breeding herds in the region with 50–70 animals per herd. Thirty mature animals including six males and 24 females of

Table 1: Description of morphometric measurements

Measurement	Description
1. Head length	Distance from the nape to the rostral end of the muzzle
2. Face length	Distance from the widest part of head to rostral end of muzzle
3. Face width	Distance of the widest points of the head
4. Ear length	Distance from the root to the end point of ear
5. Ear width	Distance of the largest points of the ear
6. Horn length	Distance from the base of horn to the tip
7. Muzzle circumference	The circumference of the muzzle over the nostrils
8. Body length	The horizontal distance from the point of shoulder to pin bone
9. Rump height	Distance from the bottom of the hind foot to the highest point between hooks
10. Heart girth	Circumference just behind the forelegs
11. Whither height	Distance from the bottom of the front foot to the highest point over wither
12. Tail girth	The biggest circumference at the base of the tail
13. Rump width	The distance between the points on either side of the animal located at one half of the distance measured from ventral point
14. Tail length	Distance from the base of the tail proximal end of the first coccygeal bone to the distal end of the last coccygeal bone

4-8 years of age were subjected to morphological and morphometric characterization from six different herds (an average of five animals per herd) which represented approximately 2.5% of the total Kinniya cattle population in Kinniya veterinary division.

Morphological and morphometric data collection

Six descriptive traits, *viz*., hair coat colour, dewlap development, hump development, horn morphology, hoof colour and tail switch colour were considered for morphological traits along with morphometric measurements recommended for phenotypic characterization by the Food and Agriculture Organization (FAO, 2012). Accordingly, fourteen morphometric traits, *viz*., head length, face length, face width, ear length, ear width, horn length, muzzle circumference, body length, rump height, heart girth, whither height, tail girth, rump width, and tail length were measured. A full description of each morphometric measurement is given in Table 1. All measurements were taken using a scaled band and a scale ruler as illustrated in Figure 1.

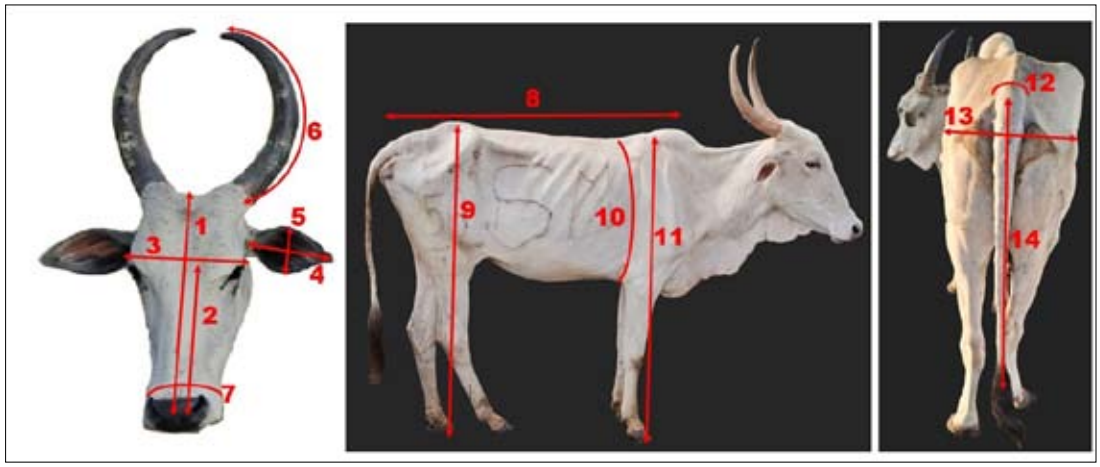


Figure 1: Measurement of morphometric indices of Kinniya cattle

Data analysis

Descriptive statistics for all morphometric measurements of Kinniya cattle were calculated using R 4.4.0 software (R core team 2023). To compare the morphometric characteristics of Kinniya cattle (both bulls and cows) with those of other indigenous cattle population which have the closest similarity, an independent two-sample t-test was performed using summary statistics (means, standard deviations, and sample sizes). This test was selected because the available reference data provided sample means and variances rather than population parameters, making a one-sample *t*-test inappropriate. Since the population standard deviations were unknown, the t-statistic (rather than a z-statistic) was used for hypothesis testing. Morphometric traits compared included heart girth, height at withers, rump height, rump width, body length, head length, face width. Reference data for these traits were obtained from previously published studies on Sri Lankan indigenous cattle (Silva et al., 2010; Wijebandara et al., 2022). Additionally, the observed morphometrics of Kinniya cattle were compared with the regional Zebu cattle populations to identify the phenotypic similarities and possible genomic inheritances.

RESULTS AND DISCUSSION

Distribution of Kinniya cattle and the system of operation

Kinniya cattle are natively distributed in mainland and peninsula (predominantly Kinniya, Muthur and Verugal divisions) of Trincomalee district of Sri Lanka where Kinniya serves as the predominant distribution center. This region is located at 8°20’ - 8°60’N and 81°10’ - 81°40’ E in the Low Country Dry Zone (Figure 2).

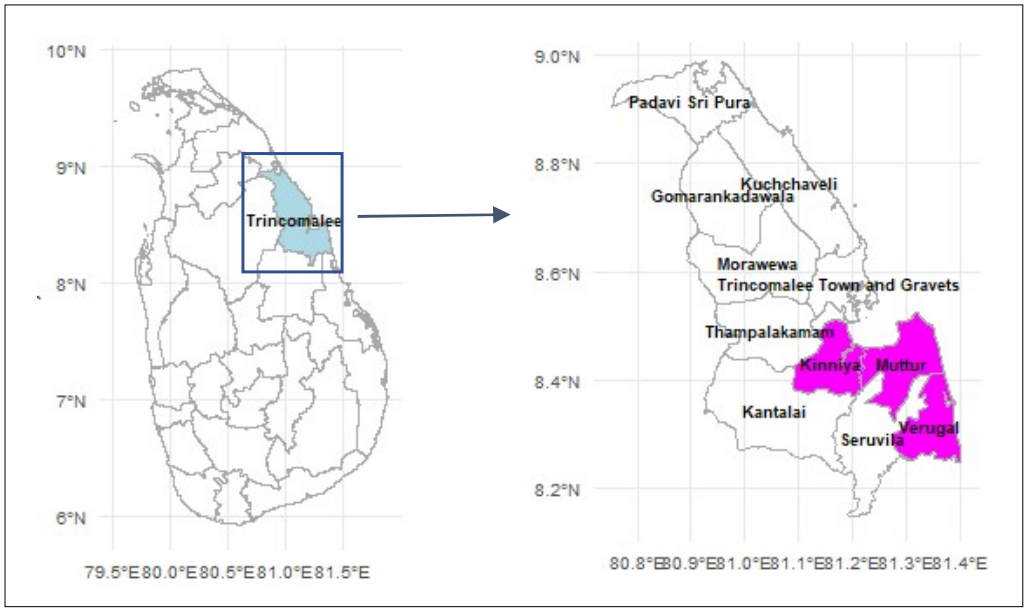


Figure 2: Distribution of Kinniya cattle

The main economic activity in the area is fishing and related industries like dry fish production. Rain fed crop production is also a major activity in the backland area. Although there are no proper irrigation channels in the area, small tanks are used to store rainwater. Only one season of rain fed paddy cultivation (September to January: north-east monsoon/Maha season) and the irrigated seasonal crop (Maize, chilli) production are the agricultural activities. Salt production and animal farming are other income generating activities together with trading, especially in the commercial areas.

Kinniya cattle are bred naturally. In a herd of about 70 animals, one or two breeding bulls with well-developed body structure with desired morphologies such as well-developed horns, hump and dewlap (Figure 3) are kept for breeding purposes. A minimum number of bulls are kept in a herd to minimize management difficulties caused by aggressive behaviour.

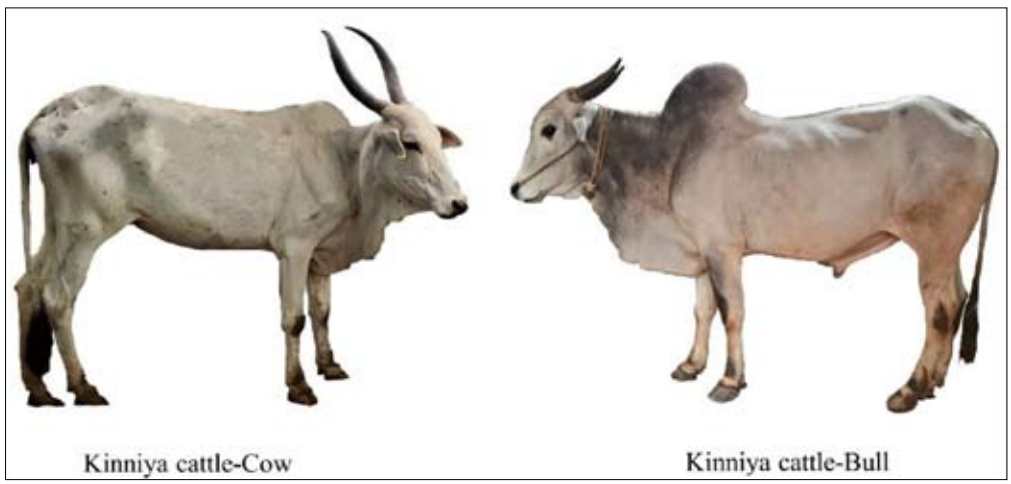


Figure 3: Male and female animal of Kinniya cattle

Rotating of bulls among herds is practiced to prevent inbreeding. Around three months of age, the appropriate males for draught purpose, especially for pulling carts (Figure 4), are chosen based on their morphological characteristics by visual observations, particularly the horns and hump. The bulls used for human transportation (Figure 4B) are subjected to castration to ensure the tamed behavior whereas others remain intact.

Figure 4: Pulling carts used mostly for the delivery of firewood and goods in the surrounding areas of Kinniya (A) and for tourism purpose (B) in Sigiriya and Habarana areas

Kinniya cattle are managed under semi-nomadic management strategies. Bulls which are selected and raised for pulling carts are managed intensively at the stalls established in the backyards of cattle owner. There are two to three such bulls in a stall where they are fed mainly with hay and concentrates. Those which are not selected for draught activities are reared in extensively managed herds with 50 – 70 animals on average (Figure 5). A herdsman who is employed by cattle owners looks after the animals under extensive conditions. Skin branding is the only identification method used.

Figure 5: Extensively manages Kinniya cattle herds

During the paddy cultivating season, cattle herds are moved to the uncultivated, mainly forest areas around 20 to 30 km away from the Kinniya region. During the day, animals in the forest are allowed to graze freely, and at night they are kept in temporarily established paddocks next to the forest area (Figure 6-A). During the rainy seasons, if the surrounding forest areas get flooded, animals are fed with tree fodder inside the paddocks. A

herd of 50 – 70 animals is kept in a roughly 50 m × 60 m paddock. Animals are brought back to the villages during the harvesting season and kept in temporarily built paddocks at night on rental basis. The paddocking area is an elevated area of a harvested paddy field (Figure 6-B). The land rent is settled with milk, manure, calves, or rental fees for the landowner. During the day, the animals are let loose to graze freely in the harvested fields. Rain-fed tanks next to the paddy fields provide water during the day. However, animals have no access to food or water during the night.

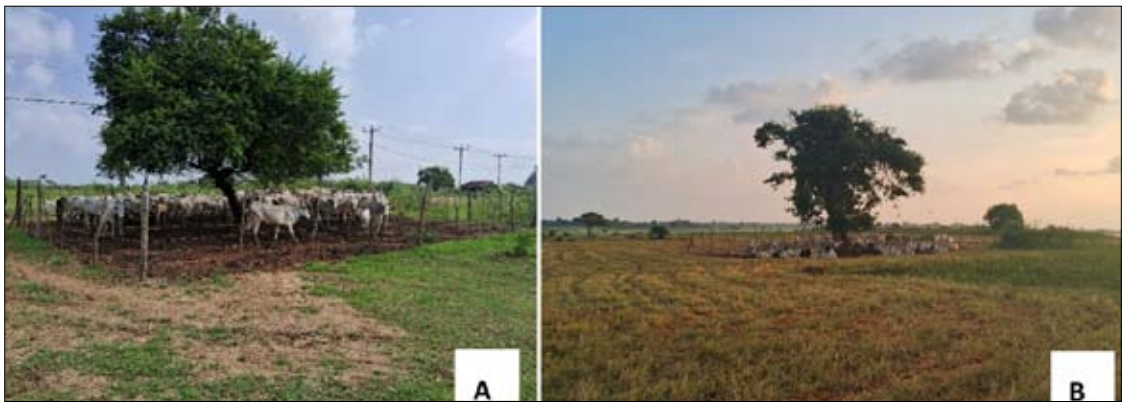


Figure 6: Temporarily established paddocks; A: paddock next to forest area, B: paddock on an elevated land of a paddy field

Abeygunawardena et al. (1997) have identified two distinct local cattle production systems based on the agro-ecology, rainfall pattern and cropping system in the dry zone of Sri Lanka, namely, dry zone traditional village system and dry zone irrigated settlement system. The dry zone traditional village system is characterized by extensive management, communal grazing during day time, keeping in paddock at night, and milking in the morning. Primary source of income of about 92% of the traditional dry zone villagers is cattle rearing. Although most of the operations within Kinniya cattle farming are aligned with the dry zones traditional village system. Management practices such as movement of herds regularly and cyclically based on the seasonal availability of pasture and water resources caused the Kinniya cattle farming to emerge as a semi-nomadic management system with some deviation from it. These specified management practices are essential for the animals' survival and well-being, especially under arid or semi-arid conditions.

Meat is the primary production associated with Kinniya cattle farming. Mature, well-built animals typically weigh 250-350 kg. Milk production is not a priority as a cow produces only 1-3 L of milk per day. and lactation lasts approximately for six months. There is no formal milk collection chain linked with the farms. Local vendors visit herds to collect milk which is sold within the village. Productive lifespan of an animal is 8-10 years and the older animals that are not productive are culled. According to Abeygunawardena et al. (1997), 45% of the income of the dry zone cattle farming came from selling animals for meat, 34% from selling milk and the balance from selling manure (12%), draught associated services and selling animals for draught. Pagthinathan & Sathiyasegar (2013) reported that in the dry zone indigenous cattle management system, the average milk yield was 1.7 L per animal per day and around 22.5% of farmers kept cattle for meat purposes.

Phenotypic characterization and identification of unique phenotypes of Kinniya cattle

In general, Kinniya cattle are white in their hair coat. But the bulls typically have shaded areas of darker hair, ranging from grey to black, especially on the head, neck, hump, and knees, whereas cows have almost white hair. Dewlap and hump are prominently developed in males. Around 70% of the cows have curved horns while others have straight horns oriented upward. In selected bulls, horns are usually trimmed and polished to maintain their appearance at the carts. Hoof and tail switch color is black in both bulls and cows (Figure 3). Compared to other local cattle in the country, Kinniya cattle show

a relatively large body structure. Table 2 provides the comparison of Kinniya cattle morphometrics together with the values reported for other indigenous cattle in the country. The specific morphometrics of Kinniya cattle are listed in Table 3.

Table 2: Morphometrics [Mean ± SD (cm)] of Kinniya and other indigenous cattle in Sri Lanka.

Morphometric characteristic	Sex	Northern Dwarf [f]	Northern	Southern	North Central	Thawalam	Thamakaduwa	Kinniya [g]
Heart girth	Male	110.0 ± 2.5^{a}	110 ± 19^{ab}	125 ± 14^{ab}	132 ± 9^{bc}	140 ± 12^{bc}	150 ± 16^{cd}	167.6 ± 6.7^{d}
	Female	105.2 ± 3.4^{a}	111 ± 18^{b}	125 ± 13^{c}	132 ± 9^{d}	N/A	131 ± 13^{d}	144.8 ± 4.9^{e}
Height at withers	Male	90.1 ± 3.0^{a}	93 ± 8.7^{ab}	96 ± 9^{ab}	101 ± 7^{b}	112 ± 09.6^{c}	123 ± 16^{cd}	130.4 ± 7.3^{d}
	Female	88.1 ± 1.5^{a}	94 ± 8.5^{b}	96 ± 9^{b}	101 ± 6^{c}	N/A	118 ± 20^{d}	117.0 ± 2.3^{d}
Rump height	Male	97.4 ± 1.5^{a}	N/A	N/A	N/A	N/A	N/A	135.5 ± 1.5^{b}
	Female	94.5 ± 2.2^{a}	N/A	N/A	N/A	N/A	N/A	124.3 ± 3.1^{b}
Rump width	Male	N/A	29 ± 4^{a}	30 ± 5^{a}	32 ± 3^{a}	31 ± 4.6^{a}	33 ± 3^{a}	35.6 ± 2.5^{a}
	Female	N/A	30 ± 4^{a}	30 ± 5^{a}	32 ± 3^{b}	N/A	38 ± 5^{c}	33.9 ± 1.8^{d}
Body length	Male	N/A	97 ± 9^{a}	97 ± 12^{a}	103 ± 3.8^{a}	116 ± 10^{b}	128 ± 33^{abc}	126 ± 3^{c}
	Female	N/A	97 ± 9^{a}	97 ± 11^{a}	106 ± 8.8^{b}	N/A	109 ± 5^{c}	114 ± 3^{d}
Head length	Male	N/A	35 ± 3^{a}	38 ± 5^{abc}	41 ± 3^{bc}	44 ± 4^{b}	41 ± 2^{c}	52.5 ± 1.4^{d}
	Female	N/A	35 ± 3^{a}	38 ± 5^{b}	41 ± 3^{c}	N/A	41 ± 3^{c}	43.8 ± 2.2^{d}
Face width	Male	N/A	12 ± 1^{a}	16 ± 2^{bc}	16 ± 1^{b}	14 ± 3^{c}	16 ± 2^{bc}	19.5 ± 0.7^{d}
	Female	N/A	12 ± 1^{a}	16 ± 2^{b}	16 ± 2^{b}	N/A	14 ± 2^{c}	17.2 ± 0.6^{d}

[a,b,c,d] Values with different superscripts differ significantly ($p < 0.05$). Measurements for indigenous cattle types were based on Silva et al., 2010, except for those values followed by [f] Wijebandara et al., 2022 and [g] from the present study. N/A – Not available

Table 3: Additional morphometric measurements [Mean ± SD (cm)] recorded for Kinniya cattle

Sex	Face length	Ear length	Ear width	Horn length	Muzzle circumference	Tail length	Tail girth
Male	36.5 ± 1.4	18.6 ± 0.7	9.5 ± 0.5	34.7 ± 2.9	42.3 ± 1.4	103.3 ± 5.2	20.3 ± 0.5
Female	34.8 ± 3.6	18.4 ± 0.8	8.7 ± 0.5	44.3 ± 6.1	36.9 ± 1.4	93.9 ± 2.6	18.2 ± 0.9

Morphometric characteristics of Kinniya cattle demonstrate several unique features that distinguish them from other indigenous Sri Lankan cattle, while showing strong similarity to Thamankaduwa cattle. Most of the morphometric measurements of Kinniya cattle exhibit higher values ($p < 0.05$) than those of Southern, Northern, and Thawalam cattle, indicating a comparatively larger body frame and better-developed skeletal structure. For instance, the heart girth of Kinniya cattle bulls (167.6 ± 6.7 cm) exceeds ($p < 0.05$) that of Southern (125 ± 13.7 cm) and Northern cattle (110 ± 18.9 cm). Their body length (126 ± 3 cm in bulls) is also greater ($p < 0.05$) than the corresponding measurements of Northern and Southern population. Height at withers in Kinniya cattle (130.4 ± 7.3 cm) is also higher than ($p < 0.05$) Northern and Southern cattle populations, again positioning Kinniya animals toward the upper end of the national cattle morphometric spectrum. Cranial measurements show the similar trend where Kinniya cattle head length and Face width are higher ($p < 0.05$) than values reported for other indigenous cattle populations in Sri Lanka. Despite these larger overall dimensions, their morphometric profile aligns most closely with Thamankaduwa cattle ($p > 0.05$), as seen in comparable measurements. Collectively, these results suggest that Kinniya cattle are larger than most Sri Lankan cattle types and morphometrically most similar to the Thamankaduwa breed.

The identified morphometrical differences of these indigenous cattle might be due to ancient, demographic related introductions and geographical distributions. These indigenous cattle populations are well adapted and managed for different purposes. Batu cattle are the prominent type of indigenous cattle, reared especially for milk, meat, and draught, which is similar to most Zebu type cattle found in the tropics (Silva et al., 2010). Batu cattle are the descendants of early archaic cattle (since 543 BC) in Sri Lanka with little relation to the Indo-Pakistani zebu. Thawalam cattle is specially a selected cattle population for pack animal morphology, different from other indigenous cattle (specially from Batu cattle), where breeding and rearing are done in isolated highlands in Central and Uva provinces of Sri Lanka. Generations of breeding in isolation has made them to emerge as a separate type of cattle that now possess different body conformation to rest of the indigenous cattle found in Sri Lanka (Silva et al.,2008). The Northern Province dwarf cattle population is phenotypically short compared to the other local cattle populations in Sri Lanka which are mainly reared for milk, meat, draught and manure and also a small proportion of males are reared mainly for breeding and traditional gaming activities. These population has exhibited a close evolutionary divergence with south Indian dwarf cattle populations according to mitochondrial Cytochrome Oxidase Sub-unit I based phylogenetic analysis (Wijebandara et al., 2022). Thamankaduwa cattle, reared for meat and draught, are found predominantly in the Eastern region of the country and it is believed that they originated through the cross breeding of ancient Lankan cattle with Indian white cattle breeds (Nadheer, 2005; DAPH, 2010; Silva et al., 2010; Shanjayan and Lokugalappatti, 2015). Kinniya cattle also have distributed with close geographical proximity to the areas where Thamankaduwa cattle have distributed. In addition to these specified populations, the Northern, North Central and Southern cattle represent the regionally localized Zebu cattle, with close morphology as Batu cattle (Silva et al., 2008). As indicated by Abeygunawardena et al. (1997), the majority (90%) of the indigenous cattle are distributed in the dry and intermediate zones of Sri Lanka. The district wise geographical distribution of the identified indigenous cattle types within the country is shown in Figure 7.

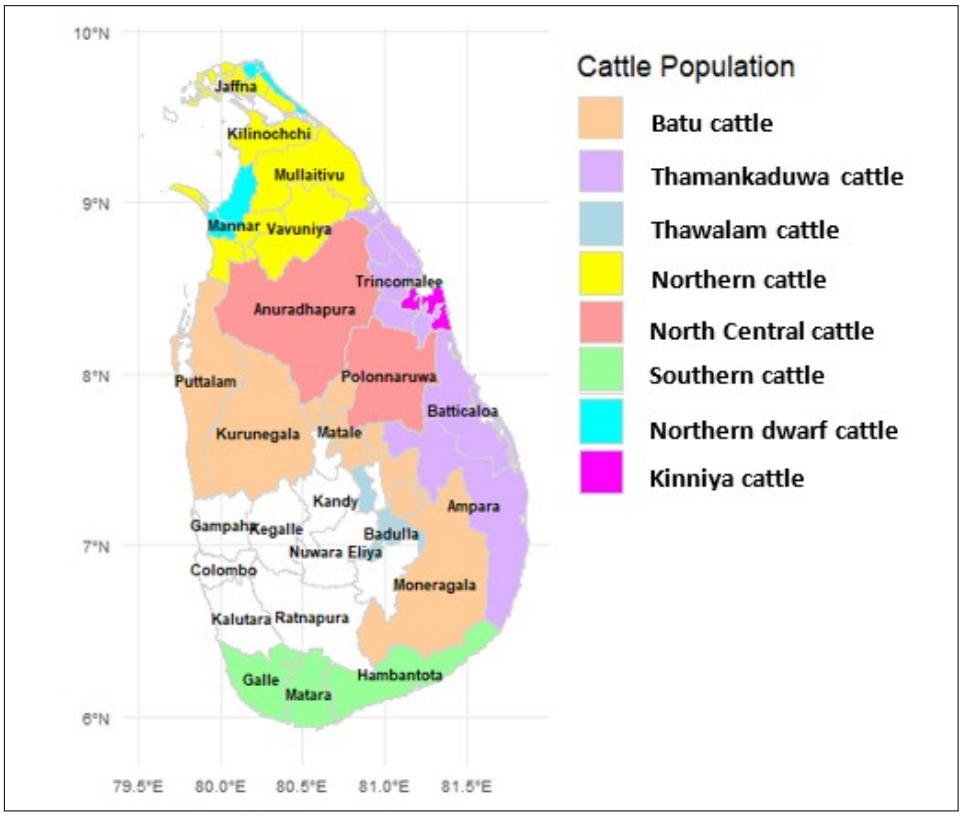


Figure 7: District wise geographical distribution of the indigenous cattle populations in Sri Lanka

Given the morphometric and geographical similarities between Kinniya and Thamankaduwa cattle (Table 2 and Figure 7), the *t* distribution analysis was carried out to assess the phenotypic deviation of the two cattle populations. The local pastoral community has been rearing Kinniya cattle over generations in the Kinniya region. This suggests that most likely Kinniya cattle have been evolved from the Indian white cattle, which were historically present in the region. The Kinniya cattle are being subjected to directional selection for draught morphology. This evolutionary process is similar to the origin of Thamankaduwa cattle, which are considered as the crossbred of ancient Lankan cattle with Indian White cattle breeds such as Tharparkar, Khillari, Hariana and Kangayam (Nadheer, 2005; DAPH, 2010; Silva et al., 2010; Shanjayan and Lokugalappatti, 2015; Lokugalappatti et al., 2023). Therefore, the recorded values for Kinniya and Thamankaduwa cattle were compared with the available reference morphometrics of Indian White cattle breeds (Table 4).

Table 4: Morphometric comparison [Mean ± SD (cm)] between Kinniya, Thamankaduwa and Indian white cattle breeds

Morphometric characteristic	Sex	Kinniya	Thamakaduwa [d]	Khillari [e]
Heart Girth	Male	167.6 ± 6.7[a]	150 ± 16[b]	173.4 ± 0.4[a]
	Female	144.8 ± 4.9[a]	131 ± 13[b]	156.8 ± 0.3[c]
Height at withers	Male	130.4 ± 7.3[a]	123 ± 16[a]	136.5 ± 0.2[a]
	Female	117.0 ± 2.3[a]	118 ± 20[a]	126.5 ± 0.2[b]
Rump width	Male	35.6 ± 2.5[a]	33 ± 3[a]	N/A
	Female	33.9 ± 1.8[a]	38 ± 5[b]	N/A
Body Length	Male	126 ± 3[a]	128 ± 33[ab]	144.1 ± 0.3[b]
	Female	114 ± 3[a]	109 ± 5[b]	132.2 ± 0.3[c]
Head length	Male	52.5 ± 1.4[a]	41 ± 2[b]	53.4 ± 0.3[a]
	Female	43.8 ± 2.2[a]	41 ± 3[b]	49.6 ± 0.1[c]
Face width	Male	19.5 ± 0.7[a]	16 ± 2[b]	20.5 ± 0.1[c]
	Female	17.2 ± 0.6[a]	14 ± 2[b]	18.9 ± 0.1[c]
Ear length	Male	18.6 ± 0.7[a]	N/A	22.1 ± 0.1[b]
	Female	18.4 ± 0.8[a]	N/A	21.5 ± 0.1[b]
Horn length	Male	34.7 ± 2.6[a]	N/A	47.9 ± 2.2[b]
	Female	44.3 ± 6.1[a]	N/A	46.3 ± 0.2[a]
Tail length	Male	103.3 ± 4.7[a]	N/A	111.7 ± 0.2[b]
	Female	93.9 ± 2.6[a]	N/A	105.1 ± 0.2[b]

[a,b,c] Values with different superscripts differ significantly ($p < 0.05$). [d] Silva et al., 2010, [e] Gokhale et al., 2009

Results indicated that, Kinniya cattle bulls are significantly differed from Thamankaduwa cattle bulls when head length and face width ($p<0.05$) were considered, whereas measurements of cows were significantly different in heart girth and face width ($p<0.05$). Additionally, these populations exhibit sex-based differences in their morphometric characteristics. Similar observations of sex-based differences in the morphometric traits within indigenous cattle populations have been previously reported by several studies (Dhal et al., 2007; Kayastha et al., 2011; Chandran et al., 2014; Jain et al., 2018; Mohan et al., 2025). Sex-based differences in morphometric traits in cattle largely reflect sexual dimorphism, driven by both hormonal and genetic mechanisms. In bulls, higher levels of androgens stimulate bone apposition, muscle mass accumulation, and prolonged growth, resulting in greater body size, heart girth, and linear measurements, whereas in females, estrogens contribute to earlier closure of growth plates and reduced periosteal expansion (Gibson et al., 2011). At the genomic level, sexually dimorphic loci associated with size and muscularity have been identified in cattle, demonstrating that male and female

growth trajectories are regulated in part by sex-specific expression of key genes (Doyle et al., 2021). Moreover, phenotypic studies across indigenous cattle populations had consistently found that males significantly exceed females in height, heart girth, and body length, reflecting not only endocrine effects but also evolutionary and functional roles, as example; larger males can confer greater performance for work or competition, while females prioritize reproduction (Fetene et al., 2025).

The morphometric comparison of Kinniya cattle with Thamankaduwa cattle and the reference Indian White cattle breeds further supports the hypothesis of a shared ancestry and parallel evolutionary processes. The height at withers of Kinniya bulls (130.4 ± 7.3 cm) lies between Thamankaduwa (123.0 ± 16.0 cm) and the larger Indian breeds such as Khillari (136.5 ± 0.2 cm) indicating an intermediate body stature consistent with partial Indian White cattle influence. Heart girth values of Kinniya bulls (167.6 ± 6.7 cm) exceed those of Thamankaduwa (150 ± 16 cm) and align more closely with heavier Khillari type (173.4 ± 0.4 cm). Cranial measurements also reinforce this pattern; Kinniya bulls' head length (52.5 ± 1.4 cm) is higher than ($P < 0.05$) Thamankaduwa (41 ± 2 cm) and approaches the longer-headed Khillari breed (53.4 ± 0.3 cm), while face width value (19.5 ± 0.7 cm) is intermediate between Thamankaduwa and Indian White breeds, reflecting a shared *Indicine* craniofacial morphology. The ear length of Kinniya cattle (18.6 ± 0.7 cm in males) exhibit close values for Khillari (22.1 ± 0.1 cm) and falls within the elongated-eared phenotype characteristic of tropical *Bos indicus* cattle (Mohan et al., 2025). Additionally, tail length in Kinniya cattle (103.3 ± 4.7 cm in bulls) is comparable to Khillari (111.7 ± 0.2 cm), suggesting retention of Indian white cattle ancestral morphology.

These observations confirm the morphological deviation of directional selection, despite the possible common ancestral path of the Kinniya and Thamankaduwa cattle populations. As listed in Table 4, Kinniya cattle has responded to the directional selection and breeding strategies adopted. This was further evidenced in Figures 8 and 9 where the visual comparison against body and horn characteristics of Thamankaduwa cattle are shown in detail. Kinniya bull possesses well-developed horns, hump and dewlap as well as the shaded areas with dark hair, ranging from grey to black (Figure 8), and the uniform orientation in horns which are straight and curved (Figure 9). Directional selection and isolated breeding may have helped Kinniya cattle to evolve as a separate population with distinguishable unique phenotypes. Despite the genetic similarities that could exist between the Kinniya cattle and Thamankaduwa cattle populations, the unique morphology and distinct herd management strategies found in the present study could be the evidences for recording the uniqueness of the Kinniya cattle. The morphological features of Kinniya cattle reported in this study could be useful in future referencing to track the deviation of the Kinniya cattle population owing to isolated breeding and directional selection for draught morphology.

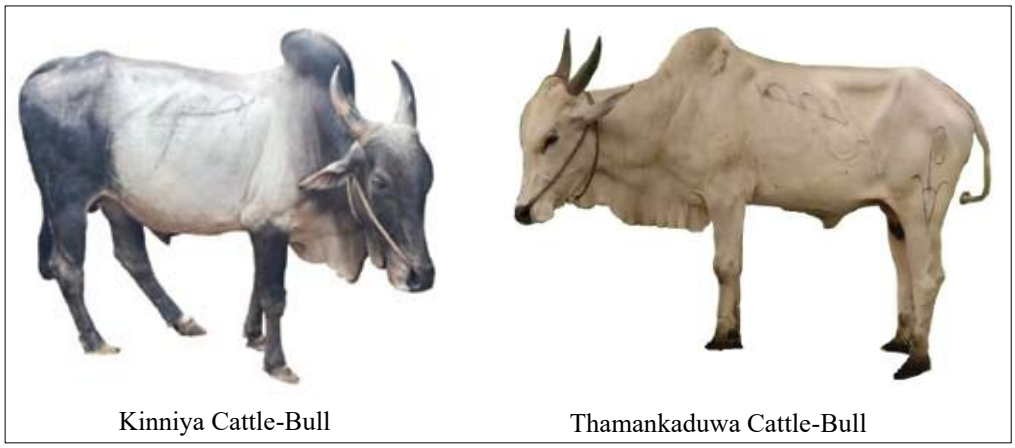


Figure 8: Morphological difference between Kinniya and Thamankaduwa cattle bulls

Thamankaduwa cattle

Kinniya cattle

Figure 9: Uniform orientation in horn morphology (straight or curved) in Kinniya cattle with comparison to Thamankaduwa cattle

Phenotypic characterization plays a crucial role in the breeding management and conservation of indigenous cattle breeds, particularly because distinct breeds are adapted to different production environments and ecological zones, with some demonstrating exceptional resilience to extreme heat and disease outbreaks (Hoffmann, 2010). By documenting morphological, adaptive, and functional traits in the animals' original production environments, characterization provides the foundational information needed to design appropriate and sustainable breeding programs tailored to village conditions (Fetene et al., 2025). Detailed morphometric profiling enables the identification of unique breed-specific attributes, detection of within-breed variation, and recognition of superior individuals for selective breeding, thereby helping to prevent genetic dilution caused by uncontrolled crossbreeding (FAO, 2012). Furthermore, phenotypic characterization highlights adaptive traits such as heat tolerance, disease resistance, and structural conformation suited to local systems which are essential for community-based breeding initiatives aimed at improving productivity without compromising breed integrity (Rege & Okeyo, 2006). In regions where genomic tools are limited, phenotypic data offer a practical and cost-effective means of identifying at-risk populations, prioritizing breeds for conservation, and formulating both in situ and ex situ conservation strategies (FAO, 2013). As many valuable animal genetic resources are linked to marginal environments and pastoral communities (Fetene et al., 2025), phenotypic characterization ensures proper documentation, utilization, and sustainable development of these genetic resources and thereby supporting long-term breed conservation and climate-resilient livestock production (Mustefa, 2023).

Several studies conducted on production system, morphology and morphometrics analysis of local cattle have implicated their importance in identifying unique characteristics to effectively incorporate them into breeding activities. As examples, Islam et al., (2022) has reported the importance of Malaysian Kedah-Kelantan cattle for meat associated breeding through phenotypic and morphometric associated selection. Kumar & Singh (2008), has reported the effects of livestock production systems and their changing dynamics on farmers across different agro-ecoregions of India. This research further revealed the potential benefits that Kinniya cattle could deliver for the production system where it plays a central role as a key genetic resource in the given socio-cultural environment. Stakeholders can make well-informed decisions about resource allocation and development projects that improve the quality of life for local pastoral communities by acknowledging the intrinsic value of this indigenous cattle population. Utilizing the unique traits of Kinniya cattle and adopting sustainable practices can boost the rich cultural heritage embedded within Sri Lankan agriculture. Even though the Kinniya cattle population differed phenotypically from the many other local cattle populations in the country, it is important to uncover their genomic linkages to methodically investigate the origin and evolution as a separate population. Also, further research into phenotypic and genetic differentiation is crucial for their long-term preservation and sustainable utilization in Sri Lanka's agricultural sector.

CONCLUSION

This study identified the population-specific phenotypic characteristics of the Kinniya cattle and demonstrated their distinctiveness compared to other indigenous cattle populations in Sri Lanka. The findings highlight that Kinniya cattle possess a unique phenotype shaped by long-term directional selection and breeding practices associated with the traditional draught-based production system. The study also documents the social, cultural, and economic importance of these cattle to the communities in the Kinniya region, underscoring their integral role in local livelihoods. To further validate and strengthen these findings, future research incorporating genomic analyses is recommended.

REFERENCES

Abeygunawardena H.A., Rathnayaka D. & Jayathilaka W.M.A.P. (1997). Characteristics of cattle farming systems in Sri Lanka. *Journal of the National Science Foundation of Sri Lanka 25*(01): 25–38. http://dx.doi.org/10.4038/jnsfsr.v25i1.5016

Chandran P.C., Dey A., Barari S.K., Kamal R., Bhatt B.P. & Prasad R.E. (2014). Characteristics and performance of Bachaur cattle in the Gangetic plains of North Bihar. *Indian Journal of Animal Sciences 84*(8): 872–875. https://doi.org/10.56093/ijans.v84i8.43271

Chandrasiri A.D.N. (2002). *Country Report on the State of Animal Genetic Resources (Sri Lanka)*. The Food and Agriculture organization of the United Nations, Rome, Italy.

Department of Census and Statistics, Colombo, Sri Lanka (2023). *National Livestock Statistics 2018–2023* (pdf). Department of Census and Statistics, Colombo, Sri Lanka. Available at http://www.statistics.gov.lk/, Accessed 01 May 2024.

Dhal B.K., Patro B.N., Rao P.K. & Panda P. (2007). Khariar cattle – an indigenous germplasm of Nuapada in the undivided Kalahandi district of Orissa. *Indian Journal of Animal Sciences 77*(9): 889–893.

Doyle J.L., Purfield D.C., Moore T., Carthy T.R., Walsh S.W., Veerkamp R.F., Evans R.D. & Berry D.P. (2021). Identification of genomic regions that exhibit sexual dimorphism for size and muscularity in cattle. *Journal of Animal Science 99*(5): skab070. https://doi.org/10.1093/jas/skab070

FAO. (2012). *Phenotypic Characterization of Animal Genetic Resources*. Food and Agriculture Organization, Rome, Italy.

FAO. (2013). *In vivo conservation of animal genetic resources*. Food and Agriculture Organization, Rome, Italy.

FAO. (eds.) (2007). *The State of the World's Animal Genetic Resources for Food and Agriculture*. Food and Agriculture Organization, Rome, Italy.

Fetene G., Alemayehu W.K., Kebede D. & Birara T.D. (2025). Qualitative and quantitative characterization of the Surma cattle population in the Bench Maji Zone, Southwest Ethiopia. *Frontiers in Animal Science 6*: 1565748. https://doi.org/10.3389/fanim.2025.1565748

Gibson M., Hickson R., Back P., Dittmer K., Schreurs N. & Rogers C. (2021). The effect of sex and age on bone morphology and strength in the metacarpus and humerus in beef-cross-dairy cattle. *Animals 11*(3): 694. https://doi.org/10.3390/ani11030694

Gokhale S.B., Bhagat R.L., Singh P.K. & Singh G. (2009). Morphometric characteristics and utility pattern of Khillar cattle in breed tract. *Indian Journal of Animal Sciences* 79(1): 47–51.

Hoffmann, I. (2010). Climate change and the characterization, breeding and conservation of animal genetic resources. *Animal genetics*, *41*: 32-46.

Islam M.S., Yimer N., Abdullah F.F.J., Han M.H.W., Mamat-Hamidi K. & Zawawi H.B.M. (2022). First study on phenotypic and morphological characteristics of Malaysian Kedah-Kelantan cattle (Bos indicus) and method of estimating their body weight. *Veterinary World 15*(03): 728–736. https://doi.org/10.14202/vetworld.2022.728-736

Jain A., Barwa D.K., Singh M., Mukherjee K., Jain T., Tantia M.S., Raja K.N. & Sharma A. (2018). Physical characteristics of the Kosali breed of cattle in its native tract. *Indian Journal of Animal Sciences 88*(12): 1362–1365. https://doi.org/10.56093/ijans.v88i12.85762

Kayastha R.B., Zaman G., Goswami R.N. & Haque A. (2011). Physical and morphometric characterization of indigenous cattle of Assam. *Open Veterinary Journal 1*(1): 7–9. https://doi.org/10.5455/OVJ.2011.v1.i0.p7

Kumar A. & Singh D.K. (2008). Livestock production systems in India: an appraisal across agro-ecological regions. *Indian Journal of Agricultural Economics 63*(04): 1–21. https://doi.org/10.22004/ag.econ.204602

Kurukulasuriya M.S., Chang-Yeon C., Han J. & Silva P. (2023). Genetic variability among indigenous cattle populations in Sri Lanka. In: *International Symposium on Sustainable Animal Production and Health – Current Status and Way Forward* (eds. G. Viljoen, M.G. Podesta & P. Boettcher), Food and Agriculture Organization of the United Nations, Rome, Italy, pp. 134-140.

Lokugalappatti L.G.S., Wickramasinghe S., Abbas P.A.B.D., Hussain T., Ramasamy S., Manomohan V., Tapsoba A.S.R., Pichler R., Babar M.E. & Periasamy K. (2023). Indigenous cattle of Sri Lanka: genetic and phylogeographic relationship with Zebu of Indus Valley and South Indian origin. *PLOS ONE 18*(08): e0282761. https://doi.org/10.1371/journal.pone.0282761

Mohan K., Kumar P. & Kundu A. (2025). Comparative analysis of the physical and phenotypic traits of native cattle (Bos indicus) in the Tarai region of North Bihar for conservation. *Veterinary World 18*(1): 95–101. https://doi.org/10.14202/vetworld.2025.95-101

Mustefa A. (2023). Implication of phenotypic and molecular characterization to breed differentiation of Ethiopian cattle:

A review. *Ecological Genetics and Genomics 29*: 100208. https://doi.org/10.1016/j.egg.2023.100208

Pagthinathan M. & Sathiyasegar K. (2013). Cattle and buffalo farming in the Dry Zone of Sri Lanka: a preliminary survey. In: *International Symposium on Agriculture and Environment 2013*, University of Ruhuna, Matara, Sri Lanka.

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available at https://www.R-project.org/, Accessed 10 March 2025.

Rege J.E.O. & Okeyo A.M. (2006). *Improving our knowledge of tropical indigenous animal genetic resources*. International Livestock Research Institute, Nairobi, Kenya.

Shanjayan N. & Lokugalappatti L.G.S. (2015). A morphometric analysis of indigenous white cattle (Thamankaduwa breed) in the Eastern Province of Sri Lanka with a description of a novel character. In: *Proceedings of the 5th International Symposium – IntSym 2015*, South Eastern University of Sri Lanka (SEUSL), Oluvil, Sri Lanka.

Silva P., Herath H.M.S.P. & Chandrasiri A.D.N. (2008). Indigenous cattle types and production systems in Sri Lanka. *Economic Review 34*(5–6):25–28.

Silva P., Jianlin H., Hanotte O., Chandrasiri A.D.N. & Herath H.M.S.P. (2010). Indigenous cattle in Sri Lanka: production systems and genetic diversity. In: *Sustainable Improvement of Animal Production and Health* (eds. N.E. Odongo, M. Garcia & G.J. Viljoen), pp. 45–48. Food and Agriculture Organization of the United Nations, Rome, Italy.

Wijebandara K.G.C.B., Kurukulasuriya M.S., Bandaranayake P.C.G. & Silva G.L.L.P. (2022). Phylogenetic analysis of local cattle population in Northern Province of Sri Lanka. In: *Proceedings of the 12th World Congress on Genetics Applied to Livestock Production (WCGALP): Technical and Species-Orientated Innovations in Animal Breeding, and Contribution of Genetics to Solving Societal Challenges*, Wageningen Academic Publishers, Wageningen, Netherlands, pp. 1881–1884.

Wijebandara K.G.C.B. & Silva G.L.L.P. (2024). Unique features of 'Kinniya Cattle': a type of cattle bred for draught purpose in Sri Lanka. In: *Proceedings of the 8th International Research Conference of Uva Wellassa University (IRCUWU 2024)*, Uva Wellassa University, Badulla, Sri Lanka, pp. 380.

J.Natn.Sci.Foundation Sri Lanka 2026 54 (1): 77 - 86
DOI: http://dx.doi.org/10.4038/jnsfsr.v54i1.12361

RESEARCH ARTICLE

Deep Learning

Weedex : Deep learning enabled autonomous robot for detection and removal of paddy weeds

CS Silva*, TH Warnakulasooriya, EMUS Bandara, NDP Wanigasuriya and TTGMK Sewmini
Department of Instrumentation and Automation Technology, Faculty of Technology, University of Colombo, Homagama, SriLanka.

Submitted: 21 October 2024; Revised: 11 November 2025; Accepted: 27 November 2025

Abstract: This research addresses the challenge of detecting and removing paddy weeds through the integration of computer vision, deep learning, and robotics. A dataset comprising the five weed species, *Fimbristylis miliacea*, *Panicum repens*, *Cyperus rotundus*, *Commelina diffusa*, and *Limnocharis flava,* was collected using a digital camera. The study focussed on developing a weed removal robot, named *Weedex*, with two core components: hardware design for mechanical weed removal and software model development for automated weed detection. Once weeds are detected by the software model, *Weedex* should be able to autonomously locate and mechanically remove the identified weed plants. To enhance model robustness and address class imbalance, data augmentation techniques such as random flips and color/brightness adjustments were applied. Object detection was performed using the algorithm YOLOv8-s, YOLOv8-x, and YOLOv8-seg. Deep feature extraction from detected objects utilized multiple pretrained architectures, including EfficientNet-B0, EfficientNet-B5, ResNet50, MobileNetV2, and a custom convolutional neural network. Model performance was evaluated using precision, recall, F1-score, and accuracy. Results demonstrate that EfficientNet-B5 consistently outperformed alternative feature extractors, while the combination of YOLOv8-seg with EfficientNet-B5 embeddings achieved the best performance (≈ 99% accuracy, F1 ≈ 0.99) on the test data. This improvement was attributed to pixel-level weed localization, reduced background interference in feature extraction, and robust similarity-based classification. This research provides a practical solution for automated weed management, potentially reducing herbicide use and labor requirements, and promoting sustainable and precision agriculture practices.

Keywords: Agricultural robotics, computer vision, deep learning, paddy weed detection and removal.

INTRODUCTION

Unwanted vegetation, commonly referred to as weeds, presents a significant obstacle in agriculture by competing intensely with cultivated crops for vital resources, including nutrients, sunlight, and water. This persistent competition often results in diminished crop yield and quality, thereby generating considerable economic losses for farmers worldwide. Conventional weed management practices, such as herbicide application and manual removal, have been widely adopted; however, these approaches are associated with limitations, including elevated production costs, human exposure to toxic chemicals, and adverse environmental impacts. Recent advances in robotics, coupled with computer vision and machine learning, offer promising opportunities to enhance weed management. The development of agricultural robots specifically designed for weed detection and eradication has emerged as a rapidly expanding area of research. Furthermore, integrating state-of-the-art machine learning algorithms and computer vision techniques into robotic systems enables effective solutions for automated weed identification and removal.

* Corresponding author (chathurika@iat.cmb.ac.lk; https://orcid.org/0000-0002-3914-5392)

The detection and identification of paddy weeds is very important since paddy provides millions of people with a key source of nutrition, ensuring their survival. Successfully controlling weeds during the vegetative stage can deliver a 95% weed-free yield (Yu & Jin, 2022). This research is expected to remove paddy weeds such as *Limnocharis flava* (Diya Gowa)*, Commelina diffusa* (Gira Pala)*, Cyperus rotundus* (Kaladuru)*, Fimbristylis miliacea* (Kuda matta), and *Panicum repens* (Etora), shown in Figure 1. This research aims to develop a weed removal robot by focusing on hardware implementation and software-based weed detection. Computer vision, together with deep learning, will be used for object detection and feature extraction.

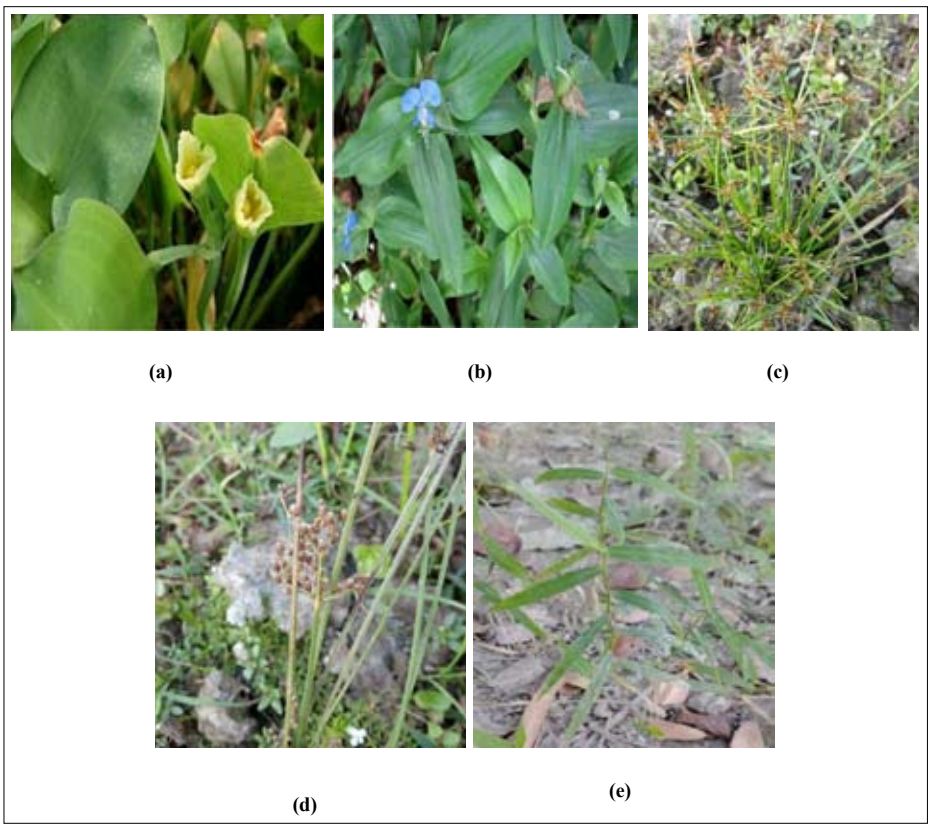


Figure 1: Paddy weeds, (a) *Limnocharis flava* (Diya Gowa), (b) *Commelina diffusa* (Gira Pala), (c) *Cyperus rotundus* (Kaladuru), (d)*Fimbristylis miliacea* (Kuda matta), and (e) *Panicum repens* (Etora)
Source: Ali et al., 2024 (figures b - e)

Literature review

The existing research on weed detection and removal using computer vision and deep learning highlights a rapidly evolving field of technological innovation in contemporary agriculture. In addition, progress in autonomous robotics, particularly the integration of artificial vision and image processing, has further enhanced the precision and efficiency of automated weed management. A weed-removal system based on artificial vision and movement planning by A*(A star) and rapidly exploring random tree (RRT) techniques was proposed by Guzmán et al., (2019). They provided the solution for finding weeds within a crop field using classifiers and the integration of a 3D-vision system that builds a point cloud featuring the plants to safeguard the weeds and the free space using Zed technology, along with search techniques such as A* and RRT to determine the trajectory that the weed-removal tool must follow. An automated weed removal system using a convolutional neural network was proposed by Rodenburg et al., (2010). The model classifies the input into two categories, namely plants and weeds. The centers of all classified weeds are calculated based on the original reference

input image. Later, coordinates of the original image are assigned to the original coordinates in the field. Then the cutting machine moves to the appropriate coordinates and removes weeds, and the whole operation repeats. Sethia et al., (2020) proposed an automated computer vision-based weed removal bot. They proposed to develop a mobile model that can detect weeds in real-time with their position coordinates and scrape them off. Choudhari (2023) proposed an automatic weed detection and removal approach based on a four-wheeled robot. The system is deployed in the field where spinach was cultivated in a row-wise order. The vehicle can move in the real field to collect data. The main advantage of this system is that weeds are identified and can be selectively removed.

Computer vision algorithms quickly evaluate visual data, allowing for efficient and accurate weed detection, which, when fed into machine learning models, continuously improves accuracy and adapts to different weed species and environmental factors. This integration not only streamlines weed management operations but also reduces labor costs, providing farmers with valuable time and contributing to sustainable agriculture (Monteiro & Santos, 2022).

A convolutional neural network is an artificial intelligence technique capable of training millions of images and predicting thousands of different classes from these images. This results in the widespread application of convolutional neural networks (CNN) in most state-of-the-art models for recognition and detection tasks (Zhou & Shao, 2022). Weed detection in paddy fields using an improved RetinaNet network was proposed by Peng *et al.* (Peng, Zhou, & Shao, 2022). This paper proposed a model named WeedDet based on RetinaNet. Det-ResNet reduced the loss of detailed information and improved the feature extraction ability of detailed textures by modifying the initial convolution structure of ResNet. (Let-ResNet and ResNet need a reference.) A unique model for weed and paddy detection using regional convolutional neural networks was proposed by Vaidhehi & Malathy, (2022). The detection of paddy weeds was considered with the aid of an automatic weed predictor model.

The YOLO algorithm's increased popularity can be attributed to its better performance. The abbreviation YOLO stands for 'You Only Look Once'. YOLO employs object detection as a regression problem that provides real-time object detection using neural networks. The original YOLO was written by Joseph Redmon in a custom framework called Darknet (Varghese & Sambathe, 2024) and later several YOLO versions were released: YOLO, YOLOv2, YOLOv3, YOLOv4, YOLOv5, YOLOv6, YOLOv7, YOLOv8. Notably, machine learning, particularly the YOLO algorithm and its versions YOLOv7 and YOLOv8, has emerged as a powerful tool for real-time weed detection. The YOLOv8 architecture comes in multiple scales—such as YOLOv8-small (YOLOv8-s), YOLOv8-large (YOLOv8-l), and YOLOv8-extra-large (YOLOv8-x) and YOLOv8-segmentation (YOLOv8-seg) that balance trade-offs between speed, efficiency, and accuracy. Narayana & Ramana (2023) have used publicly available datasets called the early crop weed dataset and 4weed dataset and obtained 99.8% precision in real-time detection of weeds using YOLO7. The development of a weed detection system based on YOLOv5 architecture using machine learning and neural network algorithms was done by Urmashev, et al., (2021) and achieved an accuracy of 83.3 %, 87.5 %, and 80 % from K-Nearest Neighbors, Random Forest, and Decision Tree classifiers, respectively.

Literature has reported several studies on weed detection in paddy fields. Weed detection in paddy fields using an improved WeedDet based on the RetinaNet network was proposed by Peng, Li, Zhou, & Shao (2022). The modelling of an automatic weed predictor model to aid farmers in handling the weed coverage and scattering of weeds in the paddy field was proposed Vaidhehi & Malathy (2022). They proposed an automatic weed predictor to support paddy weed coverage and scattering of seeds. Jin et al. (2021) proposed a CenterNet model and achieved a precision of 95.6% and a recall of 95.0%. A color index-based segmentation was applied to extract weeds from the background utilizing image processing techniques. Literature related to the autonomous detection and identification of weeds in Sri Lankan paddy fields is very limited. Edirisinghe *et al.* (2022) have used image processing and statistical methods to identify the weed types using multispectral UAV images. However, this study used UAV-based image processing rather than field-deployed robotic systems.

Several studies such as Guzmán *et al.* (Guzmán, Acevedo, & Guevara, 2019) and Sethiya *et al.* (Sethia, Guragol, & Sandhya, 2020), propose detection and localization systems, but few have demonstrated field-tested integration with mechanical or robotic weed removal tools. There is still a gap in translating detection into effective removal. Many studies rely on limited or public datasets. This raises concerns about model

generalization to diverse weed species, particularly under imbalanced data conditions. Earlier approaches (e.g., simple classifiers, color-index segmentation, or ResNet-based WeedDet) struggle with fine-grained weed–crop differentiation. Therefore, there is a requirement for an advance feature extraction and segmentation method. While mobile robots and CNN-based detectors have been proposed, many remain at the prototype or lab-testing stage. Therefore, there is a requirement to introducing scalable, low-cost, and farmer-adoptable solutions suitable for large-scale deployment in paddy cultivation. The key contribution of this study lies in the development of an autonomous weed removal robot that integrates hardware-based mechanical extraction with deep learning–driven weed detection. A novel dataset of five major paddy weed species will be collected, and advanced computer vision models, including YOLOv8 variants and multiple pretrained feature extractors, will be employed to achieve detection and classification.

MATERIALS AND METHODS

The methodology of this research primarily focuses on the development of a weed removal robot, referred to as *Weedex,* as shown in Figure 2. The development process is structured into two major components: (i) hardware design and implementation for weed removal, and (ii) software model development for accurate weed detection.

Design and development of the weed removal prototype called 'Weedex'

Leveraging SolidWorks software, a three-dimensional model of the *Weedex* was created, which allowed for visualization to refine the design before development. Using the simulation capabilities of SolidWorks, the interaction of the *Weedex* components with each other was pre-experimented.

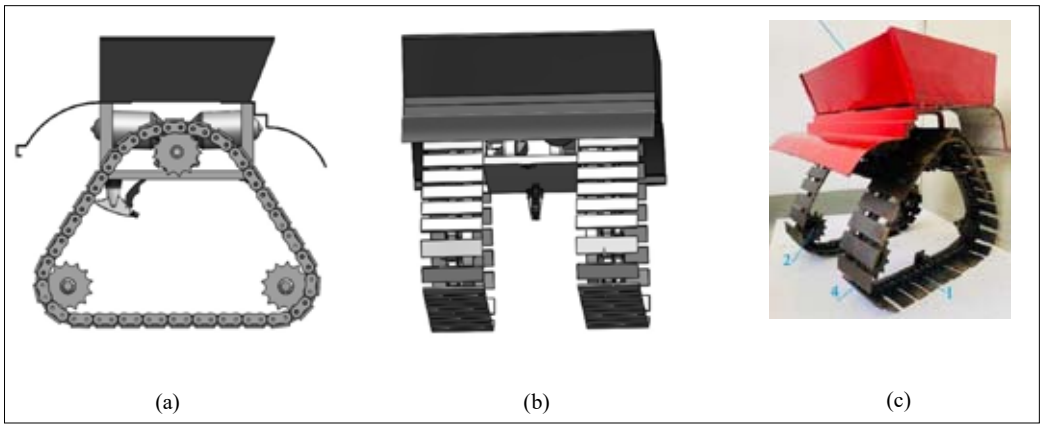


Figure 2: Weedex, (a) side view of the design (b) front view of the design (c) final Prototype

Overall, dimensions of the *Weedex* were taken according to field conditions because paddy is cultivated by the transplant method. The distance between two transplanted rows is 12 inches (30.48 cm). Considering that the width of the *Weedex* is taken as 15 inches (38.1cm) the *Weedex* can travel between the rows by observing the paddy row. The height of the prototype is taken as 10 inches (25.4cm spacing is required) since, the paddy has competition from the weeds for 4 to 6 weeks, and the maximum height of the paddy that can be grown is not more than 8 inches within that time period.

The *Weedex* assembly incorporates fundamental hardware components, including two chains, two DC motors, chassis, and body, all constructed from durable metals. The body is crafted from lightweight metal to ensure optimal functionality. The *Weedex* is made from iron components and has two sides that use tracks to move. The utilization of tracks as the primary mechanism for travel is a deliberate choice driven by the uneven ground settings typically encountered in agricultural fields. The strategic width of 15 inches between the two tracks is considered a choice, aligning with the standard length between two rows of transplanted paddy field. This dimension ensures that as the *Weedex* traverses between the transplanted rows, the track length adeptly accommodates this distance. Tracks welded with flat irons are used to ensure that the *Weedex* comes in contact with the ground. These tracks direct both sides of the prototype. The tracks are supported by a track guard

attached to the curved chassis. The chassis serves as the central structure where all hardware components are seamlessly integrated. This includes mounting the chains and sprocket wheels, crucial elements for the functionality of *Weedex*. The *Weedex* is powered by motors running on a 12V DC power supply as the primary power source.

Additionally, the control system integrates electronic components, featuring a Raspberry Pi 3B+ board as the core processing hub, a camera serves for the acquisition of images of paddy fields, a relay module for precise control, and two DC Pulse Width Modulation (PWM) motor controller drivers dedicated to the motors offering precise control over DC motors using PWM. This combination of mechanical components and electronic control systems forms the integral structure of the prototype.

Once the weeds are detected by the software model, *Weedex* integrates the detection outputs with its mechanical system to precisely localize each weed within the field. The robot then autonomously maneuvers its actuators to grasp and remove the identified weed plants, ensuring targeted weed management while minimizing damage to surrounding crops. Weedex has not been field tested in this study.

Image Acquisition

The images were collected from the paddy rows of the subplots in the Regional Rice Research and Development Center (RRRDC), Bombuwala. For dataset preparation, 246 images of *Limnocharis flava*,were collected from a Canon EOS 80 Dcamera equipped with an 18-135 mm lens. Other images of four distinct weed types such as 257 images of *Fimbristylis miliacea*, 234 images of *Panicum repens*, 213 images of *Cyperus rotundus*, and 267 images of *Commelina diffusa* were collected from the publiclya vailable dataset (Alietal., 2024). The five paddy weed species used in this research, were evenly and proportionally distributed across the training (70%), validation (15%), and test (15%) subsets such that, each of the five weed species maintained approximately the same class proportion in all three datasets, thereby preventing class imbalance and bias during model training and evaluation. This method guarantees that the model is exposed to a representative distribution of all weed species in every phase of the learning process. In this study, a one-time random split was employed rather than k-fold cross-validation due to computational constraints and dataset characteristics. Further, this approach minimizes the risk of over-fitting and provides a reliable estimate of the model's generalization capability.

Preprocessing of images

To enhance model generalization and address class imbalance, data augmentation techniques were applied, including random, horizontal, and vertical flips, as well as color and brightness adjustments. These transformations increased the diversity of the training set, thereby reducing the risk of overfitting. In the image preprocessing stage, all images were uniformly re-sized to 456 × 456 pixels to ensure consistency in input dimensions. Subsequently, pixel values were normalized using the standard ImageNet mean and standard deviation, facilitating stable convergence during training and enabling effective transfer learning from pretrained models.

Object detection

For object detection, the YOLOv8 model was implemented. This research has used YOLOv8-s, YOLOv8-x and YOLOv8-seg. The difference between YOLOv8-s and YOLOv8-x primarily lies in their model size, number of parameters, and the trade-offs between speed and accuracy. YOLOv8-s (small) is a lightweight model with fewer layers and parameters, optimized for speed and efficiency. YOLOv8-x (extra-large) is the most complex and parameter-heavy version of the YOLOv8 family. It provides higher detection accuracy, robustness, and better performance on complex datasets, but this comes at the cost of significantly greater computational requirements, longer training times, and slower inference speed. YOLOv8-seg represents a functional extension of the architecture designed for instance segmentation rather than just object detection. While standard YOLOv8 models output bounding boxes and class labels to identify objects within an image, YOLOv8-seg additionally provides pixel-level segmentation masks, enabling more precise delineation of object boundaries. The models generated bounding boxes around each detected object, enabling precise extraction of weed instances for further analysis and downstream processing.

Feature extraction and classification

For feature extraction, deep features were obtained from detected objects using multiple pretrained models, including EfficientNet-B0, EfficientNet-B5, ResNet50, MobileNetV2, and a custom CNN. A customized CNN model, inspired by the VGG16 architecture, was developed by incorporating five additional convolutional layers and one dense layer. The dense layer employed ReLU activation, followed by a final output layer with sigmoid activation for classification.

Following the feature extraction, a feature database was constructed using the training images from each weed class. The extracted embeddings were normalized and systematically stored, forming a structured reference database to support similarity-based classification during inference. For object classification, the detected regions were classified by comparing their extracted feature embeddings against the reference feature database using cosine similarity. Each region was then assigned to the class label corresponding to the highest similarity score, ensuring accurate categorization of the weed species.

Specifically, the YOLOv8s-seg model is employed for detecting and segmenting weed regions within input images. Following detection, EfficientNet-B5, pretrained on ImageNet, is utilized as a feature extractor rather than a classifier. Its classification layer is removed and replaced with a custom embedding network composed of dense, ReLU, dropout, linear, batch normalization, and normalization layers to generate discriminative feature vectors for each detected weed region. A feature database is then constructed by computing and storing these feature vectors from all training images, organized by class. During inference, a test image is processed through the pipeline: YOLOv8 detects weeds, each detected region is cropped and passed through EfficientNet-B5 to extract features, and these features are compared with the database using similarity metrics. The weed is subsequently assigned to the class label with the highest similarity score. Optionally, the system can also visualize results by drawing bounding boxes and confidence-based labels on the detected regions, providing an end-to-end solution for automated weed identification and classification.

Evaluation

The performance of the proposed weed detection and classification model was evaluated using metrics such as accuracy, precision, recall, and F1-score. Accuracy measures the overall correctness of the model by calculating the proportion of correctly classified instances among the total number of samples. Precision represents the ability of the model to correctly identify weed instances among all predicted weed samples, while recall indicates the capability of the model to detect actual weed instances present in the dataset. The F1-score, which is the harmonic mean of precision and recall, provides a balanced measure of the model's classification performance, particularly in cases of class imbalance. These evaluation metrics provide a comprehensive assessment of the effectiveness and reliability of the proposed approach for accurate paddy weed classification. To ensure a more reliable performance assessment, predictions labeled as "Unknown" were optionally filtered out. In addition, a comparative analysis of YOLOv8 variants along with the feature extraction models—EfficientNet-B0/B5, ResNet50, MobileNetV2, and the custom CNN—was conducted to determine the most effective architecture for accurate weed classification.

RESULTS AND DISCUSSION

The classification performance was evaluated across five weed classes: Class 0 – *Fimbristylis miliacea*, Class 1 – *Panicum repens*, Class 2 – *Cyperus rotundus*, Class 3 – *Commelina diffusa*, and Class 4 – *Limnocharis flava*.

As shown in Table 1, EfficientNet-B5 consistently achieved the highest precision and recall across nearly all classes, demonstrating its robustness even for smaller or less distinct categories such as Class 1 and Class 3. EfficientNet-B0 also performed well, though slightly lower than B5. In contrast, ResNet50 showed weaknesses in Class 1, while MobileNetV2 underperformed in Class 2, achieving only 0.61 precision. The Custom CNN recorded comparatively low scores across several classes, reflecting the limitations of its simpler architecture. Overall, EfficientNet-B5 proved to be the most reliable feature extractor, maintaining strong and consistent performance across all weed types.

As shown in Table 2, the EfficientNet-B5 model demonstrated robust performance across all weed classes, highlighting its ability to effectively balance both precision and recall. In contrast, the other models showed noticeable class imbalances, where high recall for certain classes was offset by poor precision in others, resulting in an increase in false positives or missed detections. This indicates that while alternative models may excel in capturing certain weed types, their inconsistency across classes makes them less reliable compared to the stability of EfficientNet-B5.

As shown in Figure 3, YOLOv8-x has improved detection when compared with YOLOv8-s for the same test image of *Commelina diffusa.*

Table 3 shows the results of YOLOv8-seg with the EfficientNet-B5 classifier.

Table 1: Results related to YOLOv8-s and classifiers

	EfficientNetNet-B0			EfficientNetNet-B5			ResNet50			MobileNetV2			CustomCNN		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Class 0	0.90	1.00	0.95	0.96	1.00	0.98	0.960.	0.90	0.92	0.72	0.78	0.75	0.79	0.56	0.65
Class 1	0.76	0.86	0.81	0.91	0.95	0.95	0.54	0.95	0.69	0.75	0.82	0.78	0.71	0.68	0.70
Class 2	0.96	0.88	0.92	1.00	0.88	0.94	0.81	0.88	0.85	0.61	0.88	0.72	0.44	0.72	0.55
Class 3	0.79	1.00	0.88	0.92	1.00	0.96	0.67	0.98	0.79	0.74	0.98	0.84	0.65	0.95	0.77
Class 4	0.99	0.88	0.93	1.00	0.98	0.99	1.00	0.74	0.85	1.00	0.80	0.89	0.95	0.76	0.84
Accuracy			0.91			0.98			0.83			0.84			0.77
Macro avg	0.88	0.92	0.90	0.97	0.96	0.96	0.80	0.89	0.82	0.76	0.85	0.80	0.71	0.73	0.70
Weighted avg	0.92	0.91	0.91	0.98	0.98	0.98	0.88	0.83	0.83	0.87	0.84	0.84	0.82	0.77	0.78

Table 2: Results related to YOLOv8-x and classifiers

	EfficientNet-B0			EfficientNet-B5			ResNet50			MobileNetV2		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Class 0	1.00	0.93	0.96	1.00	1.00	1.00	0.92	0.91	91	0.77	0.89	0.83
Class 1	0.89	0.77	0.83	0.95	0.95	0.95	0.48	0.96	0.62	0.86	0.82	0.84
Class 2	0.96	0.96	0.96	1.00	1.00	1.00	0.89	1.00	0.92	0.88	0.88	0.88
Class 3	0.73	1.00	0.85	0.89	1.00	0.94	1.00	0.75	0.85	0.70	1.00	0.82
Class 4	0.99	0.90	0.95	1.00	0.96	0.98	0.75	0.75	0.86	0.99	0.83	0.90
Accuracy			0.92			0.97			0.84			0.87
Macro avg	0.92	0.91	0.91	0.97	0.98	0.97	0.90	0.90	0.83	0.84	0.88	0.85
Weighted avg	0.93	0.92	0.92	0.98	0.97	0.97	0.84	0.84	0.85	0.90	0.87	0.87

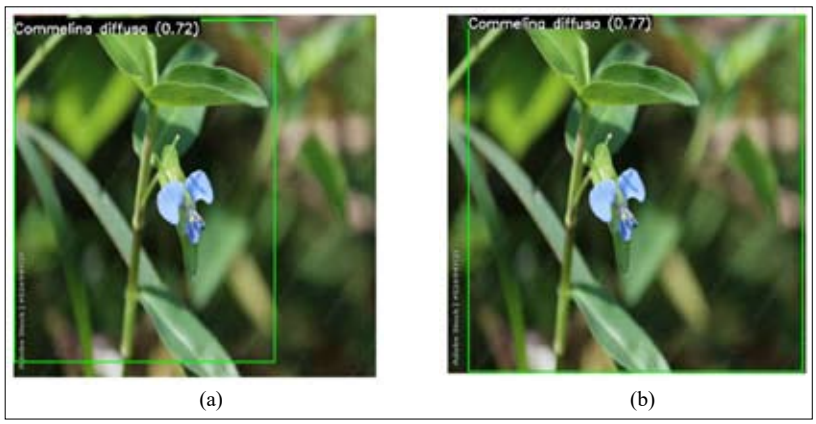


Figure 3: Comparison between (a) YOLOv8-s and (b) YOLOv8-x in weed detection

Table 3: Results related to YOLOv8-seg with EfficientNetNet-B5 classifier

	EfficientNetNet-B5		
	Precision	Recall	F1-score
Class 0	1.00	0.98	0.99
Class 1	1.00	0.98	0.99
Class 2	0.98	1.00	0.99
Class 3	0.98	1.00	0.99
Class 4	1.00	0.99	0.99
Accuracy			0.99
Macro avg	0.99	0.99	0.99
Weighted avg	1.00	0.99	1.00

YOLOv8-seg is used for weed detection because segmentation provides pixel-level masks, which improve localization and reduces misclassification for small and irregularly shaped weeds. Compared to standard YOLOv8 models (s and x), segmentation ensures that only the object area is analyzed during feature extraction, resulting in higher detection precision and recall.

Table 4 shows a summary of the performance of different YOLO variants (YOLOv8-S, YOLOv8-X, and YOLOv8-seg) integrated with various feature extractors for weed detection and classification. Within the YOLOv8-S model, EfficientNet-B0 served as a reliable baseline, while EfficientNet-B5 emerged as the strongest performer, outperforming other baseline extractors. ResNet50 underperformed due to sensitivity to class imbalance, whereas MobileNetV2 provided a

Table 4: Summary of the performance of YOLOv8-s vs YOLOv8-x vs YOLOv8-seg

YOLO Model	Feature extractor	Precision (avg)	Recall (avg)	F1-score (avg)	Accuracy
YOLOv8-s	EfficientNet-B0	0.88	0.92	0.90	91%
	EfficientNet-B5	0.97	0.96	0.96	98%
	ResNet50	0.80	0.89	0.82	83%
	MobileNetV2	0.76	0.85	0.80	84%
	Custom CNN	0.71	0.73	0.70	77%
YOLOv8-x	EfficientNet-B5	0.97	0.98	0.97	97%
	EfficientNet-B0	0.92	0.91	0.91	92%
	ResNet50	0.81	0.90	0.83	84%
	MobileNetV2	0.84	0.88	0.85	87%
YOLOv8-seg	EfficientNet-B5 + Custom embedding	0.99	0.99	0.99	99%

lightweight option with moderate success. The custom CNN achieved the lowest accuracy, highlighting the limitations of its simplified architecture.

For the YOLOv8-x model, overall performance showed incremental improvements compared to YOLOv8-s. In particular, EfficientNet-B5 remained the most effective and stable, followed by EfficientNet-B0, which showed slight gains compared to its baseline. ResNet50 and MobileNetV2 continued to reflect their earlier performance patterns, with modest improvements.

In summary, the YOLOv8-seg model combined EfficientNet-B5 with custom embedding in the proposed research achieved the best overall results indicating 99% for precision, recall, F1-score and accuracy. Similar models for weed detection and classification were reviewed and those results were compared with the outputs of the proposed research. Narayana et al. propose a YOLOv7-based object detection approach for creating a weed detection system. Experimental results revealed that the YOLOv7 model achieved the F-score, precision, and recall values for the bounding boxes as 97.6%, 99.8%, and 95.5% respectively on the early crop weed dataset and 78.53, 79.83, 86.34, and 74.24 on the 4weed dataset (Narayana & Ramana, 2023). Deng et al. propose a method for weed target detection based on YOLOX. In performance comparisons of different models, including YOLOv3, YOLOv4-tiny, YOLOv5-S, SSD and several models of the YOLOX series, namely, YOLOX-S, YOLOX-M, YOLOX-nano, and YOLOX-tiny, the results show that the YOLOX-tiny model performs best. The F1-score and recall values from the YOLOX-tiny model are 95%, and 98.3%, respectively (Deng et al. 2023). Wang *et al.* created a dataset of weeds called "Weed25". YOLOv3, YOLOv5, and Faster R-CNN were employed to train weed identification models using the Weed25 dataset and obtained average accuracies of 91.8%, 92.4%, and 92.15% respectively (Wang et al*.,* 2022).

Kumar et al. (2023) utilized five (05) convolutional neural networks (CNNs)-based architectures, namely MobileNets, MobileNetV2, and EfficientNets (B0–B2), with the DeepWeeds dataset. Results showed that EfficientNets outperformed MobileNets on all three datasets achieving accuracy value of 96.40%.

The experimental results demonstrate that the proposed approach outperforms the selected baseline classifiers in the literature. The improved performance of the proposed approach can be attributed to its ability to effectively extract and learn discriminative features from paddy field images.

CONCLUSION

This study presents a comprehensive approach for the automated detection and removal of paddy weeds using computer vision, deep learning, and robotics. A dataset containing five distinct weed species was collected and augmented to enhance model generalization and address class imbalance. The weed removal robot, *Weedex*, was developed with a dual focus on hardware implementation and software-based weed detection. Object detection was achieved using YOLOv8 variants (YOLOv8-s, YOLOv8-x, and YOLOv8-seg) in combination with feature extraction through multiple pretrained models, including EfficientNet-B5/B0, ResNet50, MobileNetV2, and a custom CNN, enabled accurate localization and classification of weed plants. Among the configurations tested, EfficientNet-B5 consistently outperformed other feature extractors, while the integration of YOLOv8-seg with EfficientNet-B5 embeddings achieved the best performance.

The outcomes of this study contribute to the field of agricultural image analysis by providing a practical and effective framework for improving weed detection under real-world conditions. Future research should focus on optimizing the weeder's design to reduce its overall weight, thereby enhancing operational efficiency in field applications. Additionally, the integration of path-tracking mechanisms aligned with transplanted paddy rows represents a crucial direction for development. Strengthening the training dataset with a larger collection of original field images would further enhance the robustness and accuracy of the detection model, ensuring more reliable weed identification in paddy cultivation.

Acknowledgement

This research was supported by the Regional Rice Research and Development Center, Bombuwala, Sri Lanka. Their guidance, expertise, and unwavering support were invaluable throughout the research.

REFERENCES

Ali, M. S., Rashid, M. R. A., Hossain, T., Kabir, M. A., Kamrul, M., Aumy, S. H. B., Mridha, M. H., Sajeeb, I. H., Islam, M. M., & Jabid, T. (2024). A comprehensive dataset of rice field weed detection from Bangladesh. *Data in Brief, 57*, 110981. https://doi.org/10.1016/j.dib.2024.110981

Choudhari, P. (2023). Weed detection using image processing. International Journal for Research in Applied Science and Engineering Technology, *11*(7), 522–525. https://doi.org/10.22214/ijraset.2023.54680

Deng, X., Qi, L., Liu, Z., Liang, S., Gong, K., & Qiu, G. (2023). Weed target detection at seedling stage in paddy fields based on YOLOX. *PLoS ONE*, *18*(12), e0294709. https://doi.org/10.1371/journal.pone.0294709

Edirisinghe, E., de Silva, N., Herath, H., Amarasingha, R. A., & Marambe, B. (2022). Identification of Major Weed Types in Broadcasted Paddy Fields Using Multispectral UAV Images. *International Research Conference of the SLTC Research University, Sri Lanka* .

Guzmán, L. E. S., Acevedo, M. L. R., & Guevara, A. R. (2019). Weed-removal system based on artificial vision and movement planning by A* and RRT techniques. Acta Scientiarum Agronomy, *41*(1), 42687. https://doi.org/10.4025/actasciagron.v41i1.42687

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. ArXiv. https://arxiv.org/abs/1512.03385

Jin, X., Che, J., & Chen, Y. (2021). Weed identification using deep learning and image processing in vegetable plantation. IEEE Access, *9*, 10940–10950. https://doi.org/10.1109/access.2021.3050296

Kumar, A., Jain, R., & Dwivedi, R. (2023). Weed detection in crops using lightweight EfficientNets. In Lecture notes in networks and systems (pp. 149–162). https://doi.org/10.1007/978-981-99-2100-3_13

Monteiro, A., & Santos, S. (2022). Sustainable approach to weed management: The role of precision weed management. Agronomy, *12*(1), 118. https://doi.org/10.3390/agronomy12010118

Narayana, C. L., & Ramana, K. V. (2023). An Efficient Real-Time Weed Detection Technique using YOLOv7. International Journal of Advanced Computer Science and Applications, *14*(2). https://doi.org/10.14569/ijacsa.2023.0140265

Peng, H., Li, Z., Zhou, Z., & Shao, Y. (2022). Weed detection in paddy field using an improved RetinaNet network. Computers and Electronics in Agriculture, 199, 107179. https://doi.org/10.1016/j.compag.2022.107179

Varghese, R., & Sambath, M. (2024). YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, 1-6. https://doi.org/10.1109/ADICS58448.2024.10533619

Rodenburg, J., Riches, C. R., & Kayeke, J. M. (2010). Addressing current and future problems of parasitic weeds in rice. Crop Protection, *29*(3), 210-221. https://doi.org/10.1016/j.cropro.2009.10.015

Sethia, G., Guragol, H.K., Sandhya, S.V., J, S., & N, R. (2020). Automated Computer Vision based Weed Removal Bot. 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), 1-6. https://doi.org/10.1109/CONECCT50063.2020.9198515

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv. https://arxiv.org/abs/1409.1556

Urmashev, B., Buribayev, Z., Amirgaliyeva, Z., Ataniyazova, A., Zhassuzak, M., & Turegali, A. (2021). Development of a weed detection system using machine learning and neural network algorithms. *Eastern-European Journal of Enterprise Technologies*, *6*(2 (114)). https://doi.org/10.15587/1729-4061.2021.246706

Vaidhehi, M., & Malathy, C. (2022). RETRACTED ARTICLE: An unique model for weed and paddy detection using regional convolutional neural networks. Acta Agriculturae Scandinavica Section B - Soil & Plant Science, *72*(1), 463–475. https://doi.org/10.1080/09064710.2021.2011395

Wang, P., Tang, Y., Luo, F., Wang, L., Li, C., Niu, Q., & Li, H. (2022). Weed25: A deep learning dataset for weed identification. *Frontiers in Plant Science*, *13*, 1053329. https://doi.org/10.3389/fpls.2022.1053329

Yu, F., Jin, Z., Guo, S., Guo, Z., Zhang, H., Xu, T., & Chen, C. (2022). Research on weed identification method in rice fields based on UAV remote sensing. Frontiers in Plant Science, 13. https://doi.org/10.3389/fpls.2022.1037760

J.Natn.Sci.Foundation Sri Lanka 2025 54 (1): 87 - 100
DOI: http://dx.doi.org/10.4038/jnsfsr.v54i1.12838

RESEARCH ARTICLE

Plant Pathology

Molecular identification of the plant parasitic nematode, *Meloidogyne* species in Sri Lanka

RVDUP Rajapakse, KGWW Bandara and CM Hettiarachchi*
Department of Chemistry, Faculty of Science , University of Colombo, PO Box 1490, Colombo 00300, Sri Lanka.

Submitted: 13 October 2025; Revised: 19 February 2026; Accepted: 27 February 2026

Abstract: Root-knot nematodes, a major type of plant pathogen of the genus *Meloidogyne*, cause substantial losses in many agricultural crops. Therefore, the correct identification of *Meloidogyne* species is essential for the application of appropriate nematode control measures for effective crop recovery. In this study, molecular identification of *Meloidogyne* species was conducted on 27 root samples with abnormal root-galls, collected from nine different types of crops. The samples were collected from 21 different agricultural fields in seven provinces of Sri Lanka, between 2014-2018. The presence of juvenile stages in 25 samples (93%) was confirmed using microscopic analysis, and DNA was isolated from purified males, eggs, and juvenile stages for molecular identification. Universal primers MF/MR, which amplified the 500 bp of ribosomal DNA, confirmed the presence of *Meloidogyne* in 23 samples (85%). Primers C2F3/1108 and 194/195 were used to confirm species-specific amplification of mitochondrial and ribosomal DNA. Species *Meloidogyne arenaria*, *Meloidogyne enterolobii,* and *Meloidogyne thailandica* were identified in 22%, 67%, and 4% of the samples collected, respectively, either as single species populations or as mixed populations. Collectively, these species were detected in several economically important crops, including tomato, cherry tomato, spinach, guava, chili, cabbage, and capsicum. *Meloidogyne enterolobii* was the most prevalent species and was detected on guava, capsicum, and tomato in the Northwestern, Central, and Western provinces. To the best of our knowledge, this is the first study on molecular identification of *Meloidogyne* species in Sri Lanka and the first report of *M. enterolobii* and *M. thailandica* in Sri Lanka. This expands the previously known worldwide distribution of these species. The presence of *Meloidogyne* species associated with economically important crops was observed in all seven provinces surveyed, indicating an urgent need for the implementation of effective nematode management strategies.

Keywords: Agricultural crops, *Meloidogyne* species, molecular identification, plant pathogens, root-knot nematode.

INTRODUCTION

Root-knot nematodes (RKNs) belonging to the genus *Meloidogyne* are economically important plant endoparasites distributed worldwide (Karssen, 2002; Perry et al., 2009; Jones et al., 2013). They are polyphagous obligate pathogens and are considered the most successful plant- parasitic nematodes in the world (Goverse et al., 2000), causing extensive economic losses (Dhandaydham et al., 2008). *Meloidogyne* spp. have a complex life cycle that includes both free-living and parasitic stages (Vanholme et al., 2004). Root-knot nematodes have a wide host range and infect almost all species of higher plants (Karssen, 2002; Perry et al., 2009).

The first discovery of *Meloidogyne* species was reported in 1855 by Reverend Miles Joseph Berkely in the roots of glasshouse-grown cucumber (*Cucumis* spp.) in England (Hunt & Handoo, 2009; Mitkowski & Abawi, 2003; Perry et al., 2009). Since then, over 100 species have been classified under the genus *Meloidogyne*. Among them, *Meloidogyne incognita*, *M. javanica, M. arenaria, M. hapla, M. chitwoodi*, and *M. graminicola* are the most widely distributed and economically important *Meloidogyne* species (Mitkowski & Abawi, 2003). The history of RKN discovery in Sri Lanka began with the identification of *M. brevicauda* from mature tea plants

* Corresponding author (chamarih@chem.cmb.ac.lk; https://orcid.org/0000-0002-9557-751X)

by Loose in 1953 (Hutchinson & Vythilingam, 1963), and since then, very few investigations have focused on RKN infestations. In Sri Lanka, morphological studies have reported *M. incognita*, *M. javanica*, *M. arenaria*, *M. hapla*, *M. brevicauda*, and *M. graminicola* on economically important crops, including vegetables, fruits, and rice, causing significant yield loss (Ekanayake et al., 2001; Ekanayake & Toida, 1997; Nishantha et al., 2025; Nugaliyedda et al., 2001). RKNs can be identified using morphological or molecular-based studies (Brito et al., 2004; Devran & Sogut, 2009; Ekanayake & Toida, 1997). However, the identification of RKNs is challenging due to their wide host range, different habitats in their life cycle, unclear species boundaries, conservative morphology, polyploidy, and sexual dimorphism (Blok & Powers, 2009). In addition, morphological identification methods for RKNs are time-consuming and comparatively difficult. At the same time, classical taxonomic knowledge and skills are continuously decreasing in many taxa, such as nematodes (Oliveira et al., 2011; Wesemael et al., 2011). Therefore, molecular methods for identifying RKNs are considered more reliable because they provide consistent results (Powers et al., 2005). These methods also have advantages, such as the ability to analyze a large number of samples and as identification is possible even in the presence of a mixture of species (Oliveira et al., 2011). Molecular identification of RKN species based on PCR-based techniques uses various target regions, such as mitochondrial DNA (Blok et al., 1997, 2002; Tigano et al., 2005), satellite DNA (Randig et al., 2009), intergenic spacer region (IGS) (Adam et al., 2007) and internal transcribed spacer regions (ITS) from ribosomal DNA (rDNA), external transcribed spacer region (ETS) (Palomares- Rius et al., 2007), and sequence-characterized amplified region (SCAR) of genomic DNA (Tigano et al., 2010).

In Sri Lanka, RKN identification has been conducted based on their morphological characteristics, and there is no available literature on their DNA-based molecular identification (Ekanayake & Toida, 1997). Hence, this study aimed to develop a DNA-based molecular approach for the accurate identification of RKN that infect economically important crops in Sri Lanka, and thereby to identify the presence and distribution of *Meloidogyne* species in different locations in Sri Lanka. The nine tested crops, including tomato, guava, long bean, cherry tomato, chili, capsicum, okra, spinach, and cabbage, are widely grown in Sri Lanka and play an important role in food security, farmer income, and local vegetable and fruit supplies. Therefore, understanding RKN diversity in these crops is essential to mitigate yield losses and support sustainable agriculture.

MATERIALS AND METHODS

Sample collection

Root samples (n=27) infected by RKNs from nine different crop plants were collected from 21 agricultural farms in different localities in seven provinces of Sri Lanka from 2014 to 2018 (Table 1, Figure 1). The sample collection areas were decided according to information obtained from the Agriculture Departments of each province. Suspected crop plants showing above-ground symptoms such as wilting, stress, disease, poor growth, and chlorosis, in the selected geographical area were collected and examined for the presence of root galls in their root systems, which is the major below-ground symptom of RKN infection. Plants showing root galls in their root system were chosen as samples for the molecular identification of *Meloidogyne* species. The infected roots were transferred into clean polythene bags and labeled appropriately. Small amounts of water were added to the collected root samples to maintain moisture during transportation of collected samples from the agricultural farms to the Biotechnology Laboratory of the Department of Chemistry, University of Colombo. Samples were washed thoroughly with tap water and inspected for egg masses on their root surface using a hand lens. Root galls were opened with forceps and

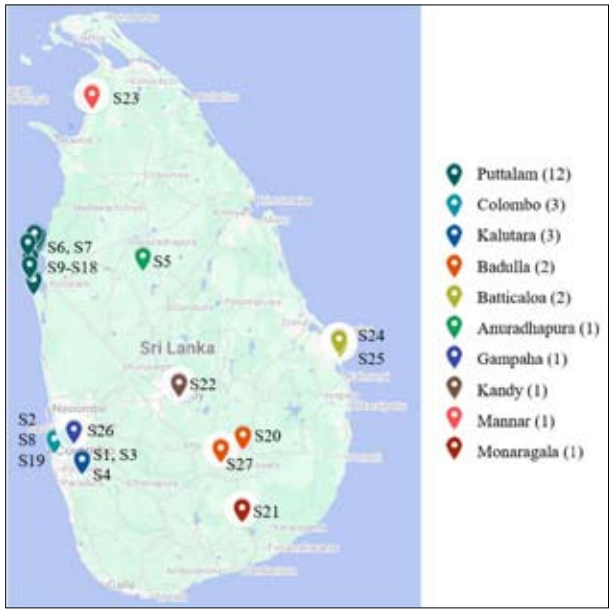


Figure 1: Sampling locations in Sri Lanka from which 27 root-knot nematode (RKN)–infected root samples were collected during 2014–2018

inspected for the presence of swollen adult females and egg masses. The females of RKNs were confirmed under a light microscope at 10×10 magnification. Roots were visually assessed to determine the gall index (GI) by counting the number of galls in each root system. The GI was scored on a scale of 0 to 5 scale, where 0 = no galls; 1 = 1-2 galls; 2 = 3-10 galls; 3 = 11-30 galls; 4 = 31-100 galls, and 5 = more than 100 galls per root system (Taylor & Sasser, 1978).

Table 1: Location and host crop of collected *Meloidogyne* infected samples.

Code	Location	District	Province	Field	Crop	Coordinates (latitude- longitude)
S1	Kananwila	Kaluthara	WP	F1	Tomato	6°45'44.9"N 80°03'43.5"E
S2	Narahenpita	Colombo	WP	F2	Tomato	6°53'60.0"N 79°52'39.5"E
S3	Kananwila	Kaluthara	WP	F1	Spinach	6°45'44.9"N 80°03'43.5"E
S4	Kananwila	Kaluthara	WP	F1	Tomato	6°45'44.9"N 80°03'43.5"E
S5	Mahailluppallama	Anuradhapura	NC	F3	Guava	8°06'42.4"N 80°28'01.1"E
S6	Kalpitiya	Puttalam	NW	F4	Tomato	8°13'41.1"N 79°45'31.0"E
S7	Kalpitiya	Puttalam	NW	F5	Guava	8°13'42.6"N 79°46'05.2"E
S8	Kaduwela	Colombo	WP	F6	Tomato	6°55'55.1"N 79°58'24.3"E
S9	Nirmalpura	Puttalam	NW	F7	Guava	7°57'15.2"N 79°44'28.7"E
S10	Kallepalliya	Puttalam	NW	F8	Guava	8°15'17.5"N 79°45'56.7"E
S11	Norochchola	Puttalam	NW	F9	Guava	8°15'53.9"N 79°44'48.4"E
S12	Eththale	Puttalam	NW	F10	Guava	8°06'55.2"N 79°43'05.4"E
S13	Nirmalpura	Puttalam	NW	F7	Long bean	7°57'15.2"N 79°44'28.7"E
S14	Nirmalpura	Puttalam	NW	F7	Tomato	7°57'15.2"N 79°44'28.7"E
S15	Palagama	Puttalam	NW	F11	Guava	8°03'42.9"N 79°42'33.5"E
S16	Illanthandiya	Puttalam	NW	F12	Guava	8°02'45.7"N 79°42'55.6"E
S17	Kandekuliya	Puttalam	NW	F13	Guava	8°12'18.4"N 79°41'58.4"E
S18	Palagama	Puttalam	NW	F11	Long bean	8°03'42.9"N 79°42'33.5"E
S19	Narahenpita	Colombo	WP	F2	Cherry tomato	6°53'60.0"N 79°52'39.5"E
S20	Passara	Badulla	UP	F14	Tomato	6°55'42.3"N 81°08'14.5"E
S21	Thanamalwila	Monaragala	UP	F15	Chili	6°26'28.6"N 81°08'03.5"E
S22	Pallekelle	Kandy	CP	F16	Capsicum	7°16'27.1"N 80°42'43.9"E
S23	Vellankulam	Mannar	NP	F17	Chili	9°11'13.2"N 80°07'52.8"E
S24	Kaluwanjikudy	Batticaloa	EP	F18	Okra	7°32'24.6"N 81°47'10.2"E
S25	Kaluwanjikudy	Batticaloa	EP	F19	Chili	7°33'56.2"N 81°46'59.6"E
S26	Biyagama	Gampaha	WP	F20	Spinach	6°57'55.9"N 80°00'19.5"E
S27	Bandarawela	Badulla	UP	F21	Cabbage	6°50'22.9"N 80°59'32.7"E

WP: Western Province, NC: North central Province, NW: Northwestern Province, CP: Central Province, UP: Uva Province, NP: Northern Province, EP: Eastern Province

Isolation of RKNs, DNA and PCR amplification

Purified males, eggs, and juvenile stages of *Meloidogyne* spp. were obtained using the incubation method (Young, 1954), which includes sieving and centrifugal flotation, to extract DNA using a mini-prep-column-based Nemato-Spin D™-Nematode DNA isolation kit (Ceygen Biotech, Sri Lanka) following the manufacturer's instructions. DNA from the isolated RKNs was used as the template for all PCR assays, which were carried out using five different primer pairs provided by New England Biolabs (UK) Ltd. Aliquots of 10 µL from the amplified PCR

products from each reaction were electrophoresed on a 1.0% (w/v) agarose gel stained with ethidium bromide at 80 V for 1 h.

DNA Sequencing and phylogenetic analyses

Randomly selected PCR-amplified products were sent for custom sequencing (Macrogen Inc, Korea). The raw sequences obtained were carefully observed, aligned using multAlin version 5.4.1 (Corpet, 1988), and manually edited using BioEdit version 7.0.9 (Hall, 1999) by correcting mismatches. The consensus sequences obtained were used to search and compare the sequences available in the GenBank, National Center of Biotechnology Information (NCBI) database, based on homology using the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990), to determine the identity of RKN isolates. In the phylogenetic analysis, the consensus sequences obtained for regions between COII and the large subunit of the rRNA gene (lrRNA) of the identified *Meloidogyne* species and homologous sequences retrieved from the GenBank database were aligned in ClustalX 2.1 (gap opening – 15; gap extension – 6.66). A total of 17 sequences, including five from the current study, were aligned for phylogenetic analysis. Some sequences of Sri Lankan *M. enterolobii* identified in this study were compared with each other and with those available in the GenBank to evaluate intra- and interspecies variations. The phylogenetic tree was constructed using COII sequences based on the Maximum Parsimony (MP) analysis (Tigano et al., 2005) in PAUP* version 4 software (Swofford, 2002) with a heuristic search performed by adding 100 random replicates and Tree Bisection-Recombination branch swapping (TBR). Gaps were considered as missing data. Bootstrap analysis was conducted with 1000 replicates to evaluate the support for each phylogenetic branch on the tree (Landa et al., 2008). For each dataset, both unweighted and weighted MP analyses were performed using PAUP* 4.0b 10 software, and support for each cluster was assessed using MP analysis with 1000 replicates. Consensus sequences obtained for the IGS/rDNA regions of the identified *Meloidogyne* species and homologous sequences retrieved from the GenBank database were aligned using ClustalW/MUSCLE MEGA version 5.0 (Tamura et al., 2011). A total of 15 sequences, including five from the current study, were aligned for phylogenetic analysis. Phylogenetic analysis was conducted using MEGA5 software with the Tamura-Nei model as the base substitution method and the Maximum Likelihood (ML) model as the statistical method with 1000 bootstrap replicates.

RESULTS AND DISCUSSION

Evaluation of RKN infestation in root samples collected from different crop fields

Galls were an integral part of the infected roots (Figure 2A). Hyaline to brown colored egg masses were detected on the surface of the galls, and when the galls were dissected, one to several whitish masses corresponding to adult females of RKNs embedded in the inner firmer tissues were detected either with the naked eye or using a hand lens. Direct microscopic examination of the collected root samples revealed the presence of females and egg masses of RKNs in 25 samples (93 %) of the 27

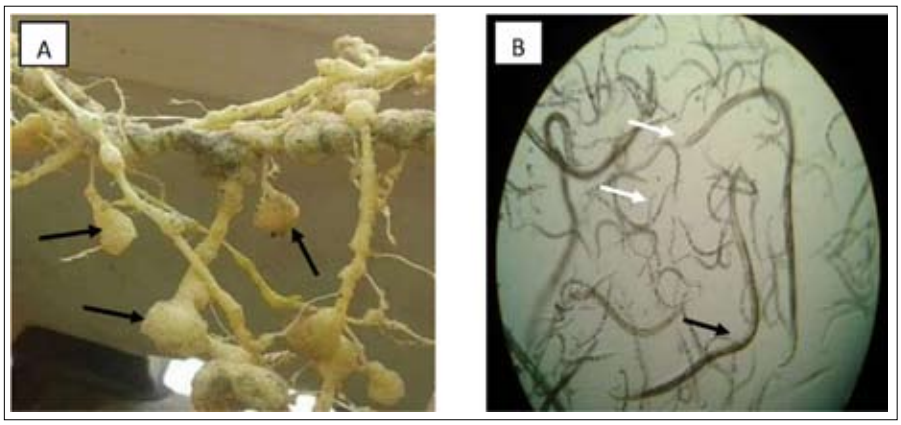


Figure 2: A. The gall morphology on the collected tomato plant roots. Galls indicated with arrows appear as integral parts of the tomato roots. B. Light micrograph of infected roots showing male (10 x 10) (arrowed in black) and juvenile stages (arrowed in white) of RKN.

samples collected. Microscopic analysis of RKN isolates also revealed the presence of juvenile stages, males, and eggs of RKNs in these samples (Figure 2B). Most of the collected root samples had a higher GI and clear large galls (Table 2). Crop-wise variations in the same field and area-based variations in the same crop were observed. The highest GI (5) was observed on guava, capsicum, and tomato. The microscopic observation of these samples showed that RKN infestations are scattered in all provinces tested.

Table 2: Gall indices of the *Meloidogyne*-infected root samples according to the method described by Taylor and Sasser (1978)

Gall index (GI)	RKN isolates
1	S14, S25
3	S2, S4, S5, S19, S20, S24, S27
4	S1, S3, S6, S7, S11, S16, S17, S21, S23, S26
5	S8, S9, S10, S12, S15, S22

GI: Gall index (0 = 0; 1 = 1-2; 2 = 3-10; 3 = 11-30; 4 = 31-100 and 5 = more than 100 galls per root system.

Molecular identification of the *Meloidogyne* genus

In Sri Lanka, several studies on RKN identification have been conducted (Ekanayake et al., 2001; Ekanayake & Toida, 1997; Nishantha et al., 2025; Nugaliyedda et al., 2001). However, all previous identifications were based on morphological characteristics. This may lead to the misidentification of *M. enterolobii* because of the overlap of its morphological characteristics with those of the dominant tropical species, since the perineal patterns of *M. enterolobii* are similar to *M. incognita* (Brito et al., 2004; Brito et al., 2015). Therefore, molecular methods have been utilized in most countries to identify *M. enterolobii* (Blok et al., 2002; Fargette & Braaksma, 1990). In this study, a simple, sensitive, and accurate diagnostic assay based on PCR was developed to identify *Meloidogyne* species collected from different crops in the region.

Amplification of the D2-D3 expansion domain of RKNs 28S ribosomal DNA with MF/MR universal primers resulted in a 500 bp fragment specific to the genus *Meloidogyne.* Out of 27 samples examined, 23 gave positive results for the amplification of the 500 bp fragment. The 500 bp amplified products for isolates S1 to S17 are shown in Figure 3.

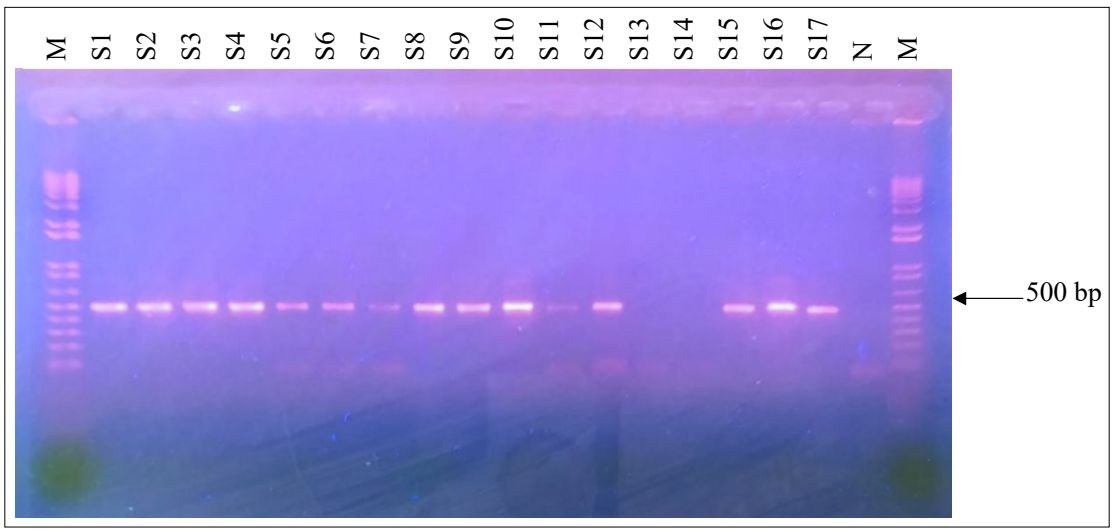


Figure 3: Gel image showing PCR amplification of nematode DNA samples with MF/MR primers. M, 1kb DNA ladder marker; S1-S17, amplified products of RKN DNA samples; N, Negative control.

However, isolates S13, S14 (Figure 3), S18, and S24 (data not shown) did not show any PCR amplification with *Meloidogyne* genus-specific primers. DNA sequencing of a 500 bp fragment of S1 (GenBank accession number MT254734.1) showed 100% similarity to the most common *Meloidogyne* species available in the GenBank database (accession numbers MN847618.1, MN661334.1, MN072431.1, etc.). Thus, samples

collected from all the districts surveyed were infected with RKNs, demonstrating the widespread occurrence of these RKN pathogens in Sri Lanka. These results further confirmed that, except for long bean and okra, all the other infected crops collected, including tomato, cherry tomato, spinach, guava, chili, cabbage, and capsicum were infected with RKNs, and need good management control of the pathogen.

Molecular identification of species of *Meloidogyne* based on mitochondrial DNA

Mitochondrial DNA is a potential source of genetic markers for species-level diagnostics of RKNs (Thomas & Wilson, 1991). *M. enterolobii* can be initially detected by producing a unique size of ~700 bp using the primers C2F3/1108 (Blok et al., 2002; Brito et al., 2004; Brito et al., 2015 Jeyaprakash et al., 2006; Powers, 2004; Tigano et al., 2005). According to previous studies, the 1100 bp amplicon for the COII/lrRNA region of the mitochondrial DNA is specific for *M. arenaria* (Powers & Harris, 1993).

The RKN DNA of 23 *Meloidogyne* genus positive isolates was amplified using primers C2F3/1108 for the region between COII/16S genes in mitochondrial DNA. Amplification of mitochondrial DNA with C2F3/1108 primers successfully discriminated the important species of RKNs in most of the infected samples. This resulted in three different amplified products in the 1100 bp, 700 bp and 520 bp for identification of *M. arenaria, M. enterolobii* and *M. daklakensis*, respectively (Figure 4A & B). *M. arenaria* was identified in isolates S1 & S2 with ~1100 bp DNA fragment (Figure 4A). *M. enterolobii* was identified in 18 isolates (S1, S5-S12, S15-S17, S19, S21-S23, S25, and S27) using the ~700 bp DNA fragment (amplification of S8 to S19 are shown in Figure 4B). Although the samples S3, S4, S20, and S26 were confirmed for the presence of *Meloidogyne* genus, no PCR amplification was obtained for species confirmation with C2F3/1108 primers. The RKN DNA of S6 and S19 isolates obtained from tomato and cherry tomato plants, respectively, showed ~520 bp band amplification, in addition to 700 bp amplified products (amplification of S19 is shown in Figure 4B).

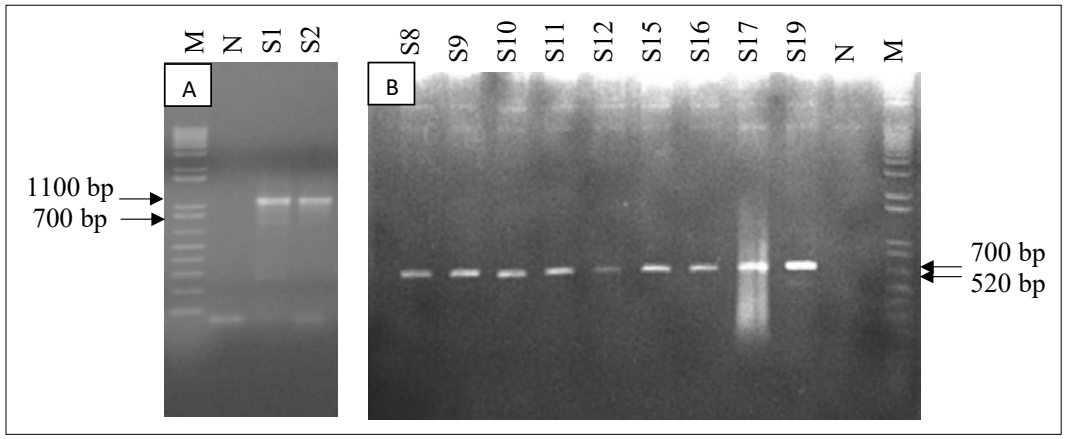


Figure 4: Gel image sAhowing amplified products of RKN DNA samples with C2F3/1108 primers. A. PCR amplification of nematode DNA samples S1 and S2. B. PCR amplification of nematode DNA samples S8-S12, S15-S17, S19. M, 1kb DNA ladder marker; N, Negative control.

The results obtained were further confirmed by DNA sequencing, and the sequencing result of the ~520 bp amplification product of the isolate S19 showed more than 70% similarity with the COII sequence of *M. daklakensis* (GenBank accession number. KU243337.1) (Table 3). The sequencing results of the 700 bp amplification products of isolates S21, S22, and S23 showed 100% similarity to *M. enterolobii* (Table 3). Moreover, the sequence of a 1100 bp amplicon of the S1 isolate showed more than 95% similarity to *M. arenaria* (Table 3).

Table 3: Identification of *Meloidogyne* species based on COII/16S rRNA BLAST results.

Isolate	Amplicon size (bp)	GenBank accession no. of the closest match	Closest match *Meloidogyne* spp.	Max score	e-value	Max identity (%)	GenBank accession no. of the isolate
S1	1100	KX983450.1	*M. arenaria*	729	0	95.65	MT741800
S19	520	KU243337.1	*M. daklakensis*	226	1e-54	71.22	MT741801
S21	700	MN269947.1	*M. enterolobii*	1081	0	100	MT741794
S22	700	MN269947.1	*M. enterolobii*	1157	0	100	MT741795
S23	700	MN269947.1	*M. enterolobii*	1139	0	100	MT741796

Globally, two genotypes of *M. arenaria* have been discovered, as reflected by amplicons of two different sizes from the analysis of the COII/lrRNA region. For example, Han et al. (2004) reported 1100 bp amplicons for RKN collected from the USA, 1700 bp amplicons for RKN collected from Japan and Korea, and both sizes of amplicons for samples collected from China, indicating the presence of both *M. arenaria* genotypes. However, in the present study, only one PCR product of 1100 bp was obtained with C2F3/1108 primers for the DNA of two isolates (S1 and S2) (Figure 4A), reflecting the presence of only one genotype of *M. arenaria* in Sri Lanka. This result agrees with previous studies that reported a 1100 bp amplicon for the COII/lrRNA region of *M. arenaria* (Powers & Harris, 1993).

Although the primer pair C2F3/1108 was stated as good enough to identify various RKN species in previous studies (Powers et al., 2005; Powers & Harris, 1993), no PCR amplification was obtained for samples S3, S4, S20, and S26, which were previously confirmed for the presence of the *Meloidogyne* genus. Several studies have previously reported the absence of PCR products with C2F3/1108 for species *M. arenaria*, *M. javanica*, and *M. incognita* (Devran & Sogut, 2009; Joseph et al., 2016; Mwesige et al. 2016). Therefore, in previous surveys, to distinguish between these sibling RKN species and to minimize false identifications, multiple molecular characteristics have been examined using different primer sets. In the present study, pale ~520 bp bands were obtained with the primer set C2F3/1108 for the S6 and S19 RKN isolates (Figure 4B). Sequencing results of these amplification products were not satisfactory despite repeated trials, which prompted the use of the primer set 194/195 for the exact identification of RKN present in these samples.

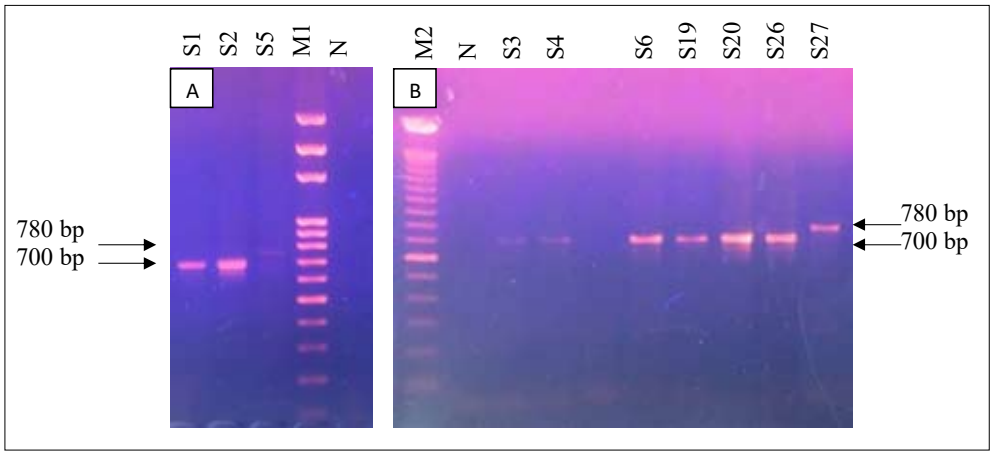


Figure 5: Gel image showing amplified products of RKN DNA samples with 194/195 primers. A. PCR amplification of nematode DNA samples S1, S2 and S5. B. PCR amplification of nematode DNA samples S3, S4 and S6, S19, S20, S26 and S27. M, 1kb DNA ladder marker; N, Negative control; M2, 100 bp ladder marker.

Molecular identification of the *Meloidogyne* species based on ribosomal DNA

The primer set 194/195 utilizes the sequence divergence exhibited by RKNs between the 5S-18S genes of the IGS-2 region of rDNA for species discrimination. Adam et al. (2007) reported an amplification product of 650-700 bp for *M. hapla*, 780 bp for *M. enterolobii* and 720 bp for the closely related tropical species *M. arenaria*, *M. incognita*, and *M. hapla.* Amplification of the IGS-2 region of ribosomal DNA with 194/195 primers yielded ~700 bp amplicons specific for *M. arenaria* and *M. hapla* for S1 and S2 isolates (Figure 5A), which produced 1100 bp size amplicons with C2F3/1108 primers. Samples S3, S4, S20 and S26, which showed no PCR amplifications with C2F3/1108 primers, showed 700 bp amplicons with 194/195 primers (Figure 5B). S6 and S19 isolates which produced multiple PCR products with C2F3/1108 primers also produced ~700 bp amplicons with 194/195 primers (Figure 5B). The 780 bp amplicon specific for *M. enterolobii* was obtained for the DNA of isolate S5 (Figure 5A) and S27 (Figure 5B). However, the 780 bp amplicons expected for the S1, S6, and S19 isolates were not discernible from the 700 bp amplicons. Selected PCR amplified products with 194/195 primers were sequenced and sequencing results revealed that the species responsible for samples S2, S3, S4 and S6 with ~700 bp products is *M. arenaria*. However, the S20 sample with ~700 bp product showed 100% identity for *M. thailandica* and ~780 bp product obtained for S5 showed 100% identity to *M. enterolobii* (Table 4).

Table 4: Identification of *Meloidogyne* species based on based on rDNA-IGS2 BLAST results.

Isolate	Amplicon size (bp)	GenBank accession no. of the closest match	Closest match *Meloidogyne* spp.	Max score	e-value	Maximum identity (%)	GenBank accession no. of the isolate
S2	700	U423442.1	*M. arenaria*	1175	0	100	MT741790
S3	700	GQ395518.1	*M. arenaria*	1130	0	100	MT741791
S4	700	U423442.1	*M. arenaria*	1047	0	100	MT741792
S6	700	GQ395518.1	*M. arenaria*	541	2e-154	97.82	MT741793
S20	700	HF568829.1	*M. thailandica*	1119	0	100	MT741789
S5	780	KP411228.1	*M. enterolobii*	1304	0	100	MT741788

Most of the RKN species relevant to the 520 bp size amplicon for the COII/lrRNA region displayed a length polymorphism between 5S-18S genes with the primer set 194/195 enabling their differentiation. Adam et al. (2007) reported a 700 bp amplicon for *M. hapla* and 1600-1700 bp products for *M. fallax* and *M. chitwoodi* with the primer set 194/195. Therefore, in the present study, the ~700 bp amplicon obtained for the DNA of isolate S6 with the primer set 194/195 was expected to be due to *M. hapla*. However, Adam et al. (2007) stated that the primer set 194/195 produces very similar sizes of PCR products for tropical species (*M. arenaria*, *M. javanica*, and *M. incognita*) and *M. hapla* (720 bp and 700bp, respectively) often leading to misidentification. In accordance with this statement, sequencing analysis of the ~700 bp amplicon obtained for the DNA of isolate S6 with the primer set 194/195 showed that the species was *M. arenaria.* These results indicated the importance of utilizing multiple gene targets to accurately identify *Meloidogyne* species.

Furthermore, this study clearly showed that amplicon sizes alone are incapable of differentiating closely related tropical RKN species *M. incognita*, *M. javanica*, and *M. arenaria*, which produce ~720 bp size amplicons, and *M. hapla*, which produces 650-700 bp amplicons. Therefore, the incorporation of additional genetic markers is important to differentiate sibling RKN species.

Confirmation of *M. arenaria* species with Far/Rar species specific SCAR primers

The isolates S1, S2, S3, S4, S6, S19, S20, and S26, which showed *M. arenaria* specific amplification with 194/195 primers, were further analyzed with *M. arenaria* specific Far/Rar SCAR primers. Except for S20, all isolates checked with Far/Rar confirmed the presence of *M. arenaria* producing a 420 bp amplification. PCR amplification of the 420 bp specific for the S26 *M. arenaria* isolate is shown in Figure 6 and for further confirmation, sequencing was carried out. The sequencing analysis

also showed 100% identity to *M. arenaria* isolates from Indonesia (Accession number. KP234264) in the GenBank database.

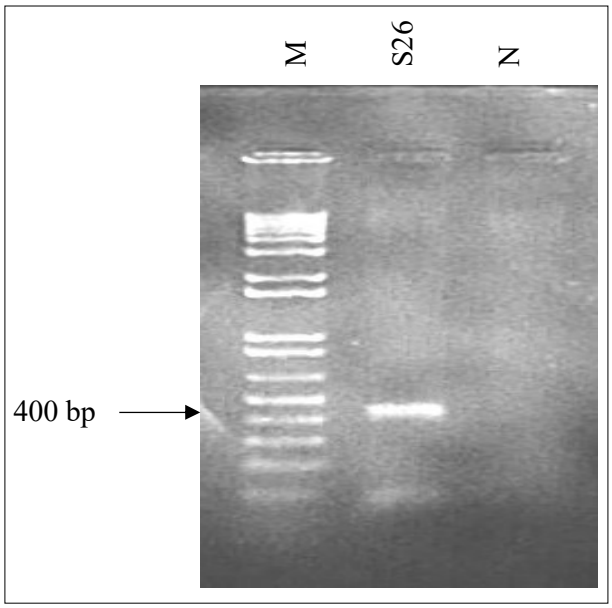


Figure 6: Gel image showing amplified product of RKN DNA sample S26 with *M. arenaria* specific Far/Rar primers. M, 1kb DNA ladder marker; N, Negative control.

No PCR amplification was observed for the S20 isolate with SCAR primers. However, isolate S20 produced an ~700 bp amplification product with 194/195 primers (Figure 5). Although isolates S1, S2, S3, S4, S6, S19, and S26 produced an ~700 bp fragment with 194/195 primers confirming the presence of *M. arenaria* with the sequencing results, isolate S20 showed the presence of *M. thailandica* (Table 4). Altogether, our results emphasize the importance of considering both PCR amplification and sequencing results to confirm the identification of different species in the *Meloidogyne* genus**.**

Confirmation of the *M. enterolobii* species with MeF/MeR species specific primers

The 18 RKN DNA samples (S1, S5-S12, S15-S17, S19, S21-S23, S25, and S27) that produced ~700 bp amplification products with C2F3/1108 primers, were further analyzed with *M. enterolobii* specific MeF/MeR primers that amplify a portion of the rDNA-IGS2 region. The presence of *M. enterolobii*, was confirmed by ~240 bp PCR products for the DNA of all RKN isolates tested. Isolates S7, S8 and S9 were sequenced and the sequencing analysis of the ~240 bp PCR products (Figure 7) obtained for S7 (GenBank accession number. MT741798), S8 (GenBank accession number. KY551570.1), and S9 (GenBank accession number. MT741797) showed 100% homology for the intragenic spacer region 2 (IGS2) of *M. enterolobii* isolates from Taiwan (Accession number. KP411228.1), Oman (Accession number. KM008548.1), China (Accession number. JN005846.1), and Switzerland (Accession number. GQ395524.1) in the GenBank database.

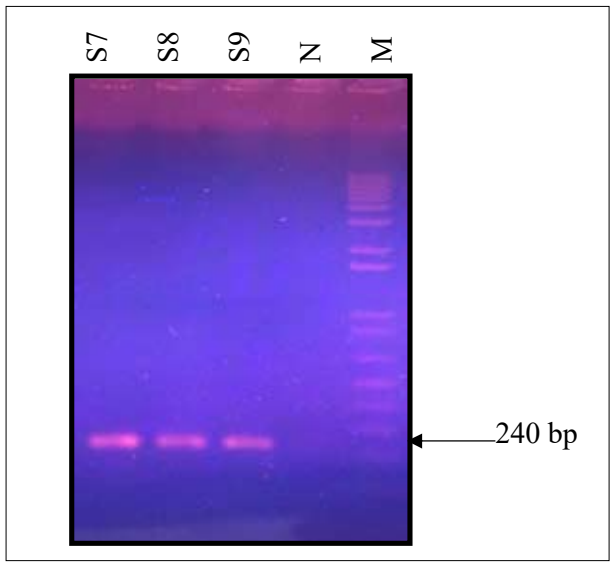


Figure 7: Gel image showing amplified products of RKN DNA samples S7, S8 and S9 with *M. enterolobii* specific MeF/MeR primers. N, Negative control; M, 1kb DNA ladder marker.

M. enterolobii was the most prevalent species, detected in 18 of 23 isolates (78.3%) positive for *Meloidogyne* genus-specific primers, occurring alone in 15 isolates and in mixed populations in 3 isolates. This is the first report of detecting *M. enterolobii* within Sri Lanka and our results further confirmed the presence of this very aggressive species in all the studied provinces in Sri Lanka. However, previous surveys conducted by Lamberti et al. (1987) and Ekanayake & Toida (1997), based on morphological characters, identified them as *M. incognita* and *M. arenaria*, which they reported as the dominant species infesting several vegetable crops in Sri Lanka. Morphological characters can always lead to misidentification of *M. enterolobii* due to the overlap of

its morphological characters with those of the dominant tropical species, i.e., perineal patterns of *M. enterolobii* are similar to *M. incognita* (Brito et al., 2004; Brito et al., 2015). *M. enterolobii* is considered the most aggressive emerging species (Brito et al., 2004) primarily due to its virulence against several root knot nematode resistance genes, such as the *Mi-1* gene in resistant tomato (Fargette, 1987). Due to this resistance breaking ability, this nematode displays a higher pathogenicity and reproductive potential than any other tropical species (Brito et al., 2004).

Phylogenetic analysis of *Meloidogyne* populations based on COII/16S rRNA sequences

The phylogenetic trees were constructed to study the diversity among RKNs in Sri Lanka. The phylogram obtained through maximum parsimony analysis contained several clades and sub-clades, which were separated based on Bootstrap Support (BS) values such as (i) *M. incognita*, *M. arenaria*, *M. javanica*, *M. thailandica*, and *M. floridensis* populations (BS=100%), (ii) *M. chitwoodi* and *M. fallax* populations (BS = 100%), (iii) *M. hapla* and *M. haplanaria* populations (BS = 100%), (iv) *M. enterolobii* populations (BS = 100%), and (v) *M. daklakensis* population (BS=82%) (Figure 8A). The outgroup sequence (GU177865.1) from the closely related *Bursaphelenchus* genus formed a separate clade. Putatively identified nematode populations from the current study grouped correctly with those identified from several parts of the world. The isolate S1 previously identified as *M. arenaria* was grouped with the species of clade (i). As shown in the phylogram, GenBank and Sri Lankan *M. enterolobii* populations (S21, S22, and S23) formed a monophyletic clade with a high bootstrap value (100%). Moreover, the sequence from the S19 isolate clustered with *M. daklakensis* in clade (v), but with a low bootstrap value of 57%.

The phylogenetic tree based on the COII-16S rRNA region showed no significant differences between the COII sequences of the Sri Lankan population and those retrieved from the GenBank, NCBI database for *M. enterolobii*, proving the lack of diversity among *M. enterolobii* populations from various geographic regions (Figure 8A). The homogenous nature of *M. enterolobii* populations identified through phylogenetic analyses suggested the applicability of similar control methods in farms infected with *M. enterolobii.* As evidenced by the bootstrap values, the clade containing *M. enterolobii* (100%) was closer to the tropical species clade comprising *M. javanica*, *M. incognita*, *M. arenaria*, *M. thailandica*, and *M. floridensis* (100%) than to the temperate species clade including *M. hapla*, and *M. haplanaria* (84%). This result is consistent with McClure et al., (2012) who reported that *M. enterolobii* is closely related to tropical species than to temperate ones based on topology of the phylogenetic tree constructed using mitochondrial DNA sequences. This close relationship can be explained by the mode of reproduction as both *M. enterolobii* and the majority of tropical species are mitotically parthenogenetic *Meloidogyne* species (Tigano et al., 2005).

Phylogenetic analysis of *Meloidogyne* populations based on IGS/rDNA sequences

Since ML has been previously proven in several studies as the ideal model that gives the best phylogenetic tree for IGS analysis, it was chosen to construct phylogenetic relationships based on IGS-rDNA region (Onkendi & Moleleki, 2013). Putatively identified nematode populations from the current study grouped correctly with those identified from several parts of the world (Figure 8B).

The phylogram obtained through maximum likelihood analysis contained several clades and sub-clades which were separated based on BS values: (i) *M. incognita*, *M. arenaria*, *M. javanica*, *M. thailandica*, and *M. floridensis* populations (BS=78%), (ii) *M. chitwoodi* and *M. fallax* populations (BS = 100%), (iii) *M. hapla* population (BS = 100%), (iv) *M. enterolobii* populations (BS = 100%), and (v) *M. haplanaria* population (BS=82%).

The isolates S2, S3, and S4 previously identified as *M. arenaria* and S20 previously identified as *M. thailandica* were grouped with the species of clade (i), and the isolate S5, which was identified as *M. enterolobii*, was in clade (iv).

Phylogenetic analysis based on the IGS-rDNA region resulted in grouping most of the tropical species into one clade that is distinct from the clade that grouped the temperate species. The clade with tropical species contained closely related apomictic species such as *M. arenaria*, *M. javanica*, and *M. incognita*, along with *M. thailandica* and *M. floridensis*, which are common in warmer climates. Clades with temperate species contained the automictic species *M. chitwood*i and *M. fallax* while *M enterolobii* formed an independent clade that is closer to the clade with tropical species than to the temperate ones. Phylogenetic analysis clearly separated the facultative parthenogenetic species *M. hapla* by forming an independent group between the automictic and apomictic species, but closer to the clade with automictic

ones. This agrees with previous studies that suggested *M. hapla* is closely related to automictic species than to apomictic species based on the percentage nucleotide base substitution model using total genomic DNA (Castagnone-Sereno et al., 1999; Onkendi & Moleleki, 2013). According to the phylogram constructed based on the IGS-rDNA, all the *Meloidogyne* species identified and used in the analysis were grouped closely into the clades with their respective *Meloidogyne* species, which were identified from several parts of the world. Altogether, the results presented in this study validate a basis to formulate alternative methods for RKN control.

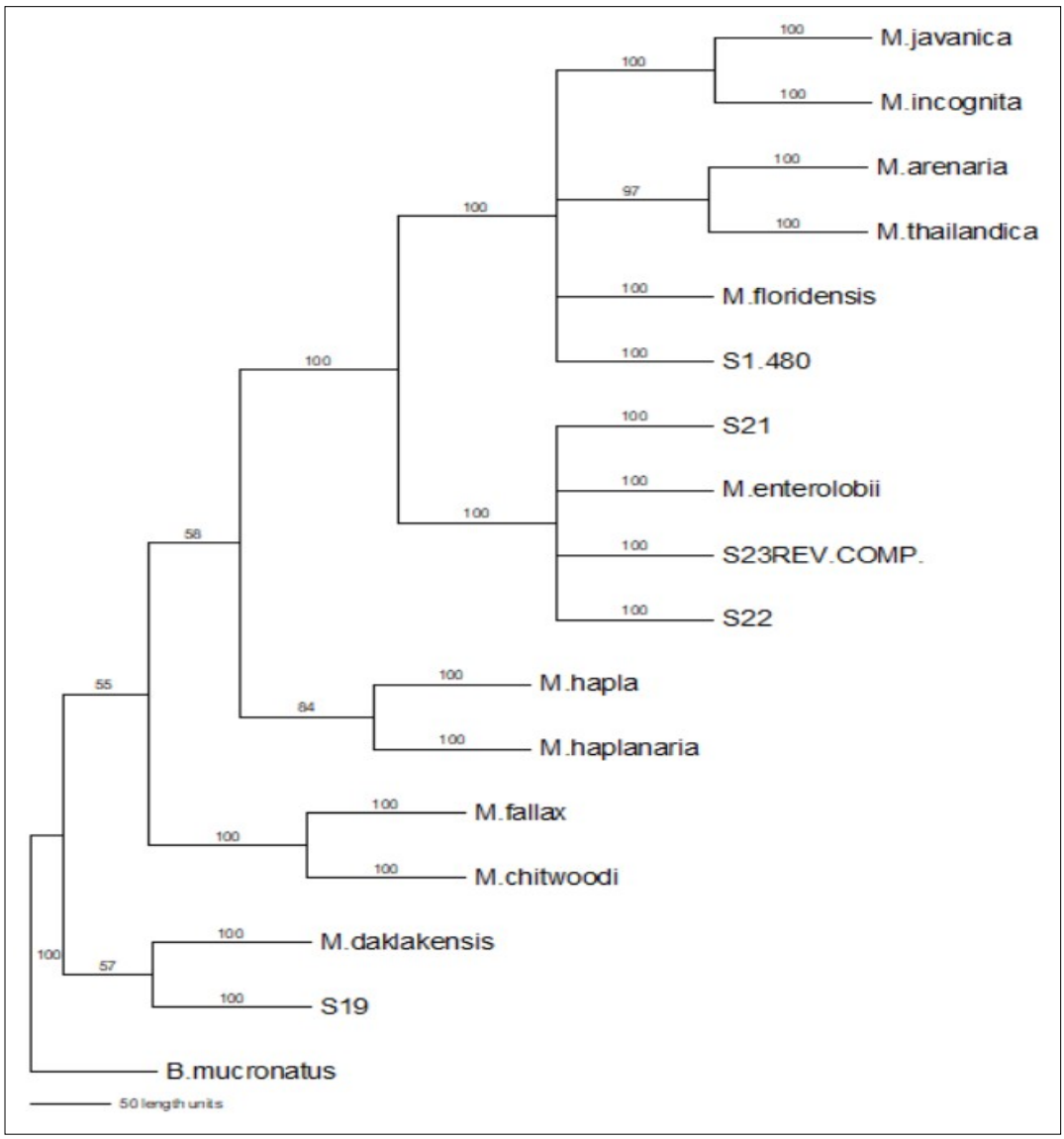


Figure 8A: Maximum parsimony tree based on mitochondrial COII-16S rRNA sequences of *Meloidogyne* species, preliminarily identified in the present study, and reference sequences obtained from the GenBank, NCBI database. S1, S19, S21, S22 and S23 represent the consensus sequences obtained in this study; the remaining sequences were retrieved from the GenBank/NCBI database as references. The analysis included 1,000 bootstrap replicates. The bootstrap support value of each clade is indicated on the nodes. *Bursaphelenchus mucronatus* (GU177865.1) was used as the outgroup in the alignment and construction of the phylogram.

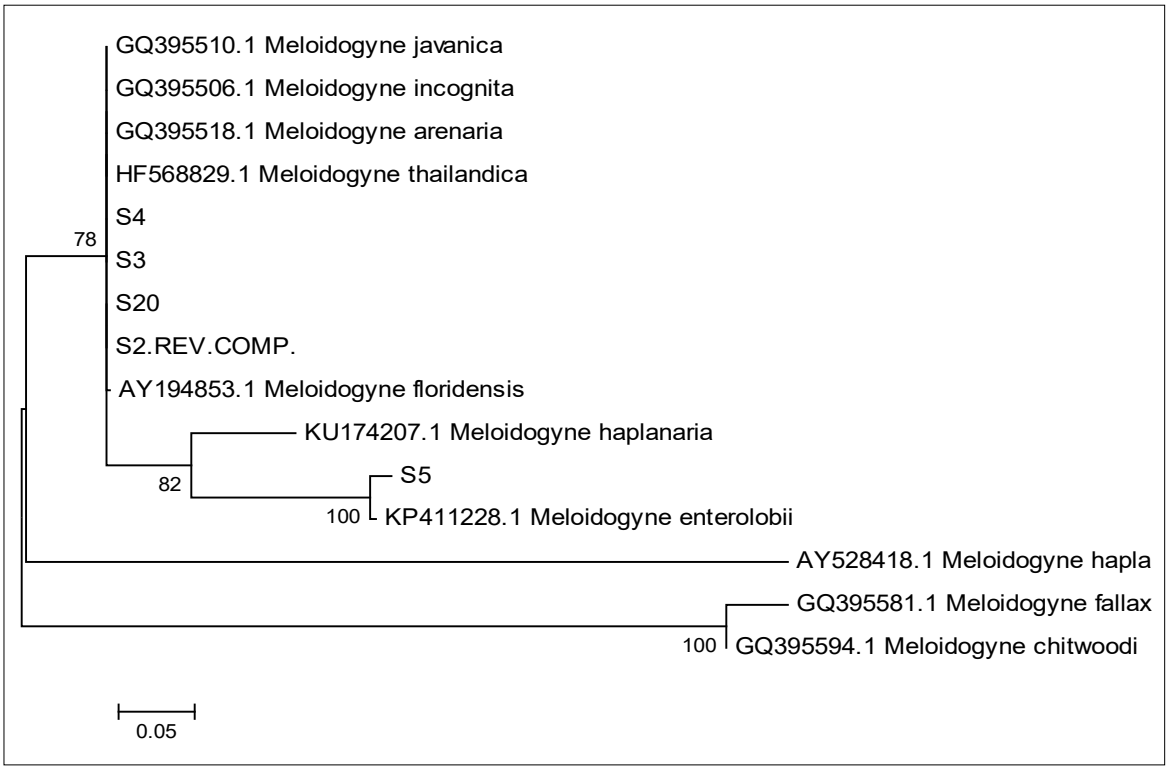


Figure 8B: Maximum likelihood tree constructed using the Tamura-Nei model based on the alignment of IGS/rDNA sequences of *Meloidogyne* species identified in this study, along with reference sequences from the GenBank, NCBI database. The sequences S4, S3, S20, S2 and S5 were obtained in this study. The analysis involved 1,000 bootstrap replicates and the bootstrap support value for each clade is indicated on the nodes. Since the 194/195 primers used in the IGS/rDNA analysis are specific to *Meloidogyne* species, no outgroup was included in the phylogram.

CONCLUSION

To the best of our knowledge, this is the first study of molecular identification of *Meloidogyne* species in Sri Lanka. In this study, *Meloidogyne* species *M. arenaria*, *M. enterolobii*, and *M. thailandica* were identified across multiple economically important crops including, tomato, cherry tomato, spinach, guava, chili, cabbage, and capsicum. Furthermore, *M. enterolobii* and *M. thailandica* were recorded for the first time in Sri Lanka through this study. The distribution of root-knot nematodes across all the provinces included in this investigation shows their threat to agriculture. These findings indicate the importance of accurate molecular identification of plant-pathogenic *Meloidogyne* species to support the implementation of effective root-knot nematode management strategies in Sri Lanka. Internal quarantine, varietal screening and prophylactic treatments are important in safeguarding the economically important crops susceptible to RKN species through in-depth studies.

Acknowledgments

The authors wish to acknowledge Mrs. S. P. Aluthge, Principal Scientist (Entomology, Fruit Research & Development Institute), Department of Agriculture, Sri Lanka and Dr. K. M. Mohotti, Nematologist and former Director, Tea Research Institute, Talawakelle, Sri Lanka, for providing required samples. This research was supported by the University of Colombo research grant (AP/3/2012/CG/25).

REFERENCES

Adam, M.A.M., Phillips, M.S. & Blok, V.C. (2007). Molecular diagnostic key for identification of single juveniles of seven common and economically important species of root-knot nematode (*Meloidogyne* spp.). *Plant Pathology, 56*(1), 190–197. https://doi.org/10.1111/j.1365-3059.2006.01455.x

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*, 403-410. https://doi: 10.1016/S0022-2836(05)80360-2.

Blok, V.C., Phillips, M.S. & Fragette, M. (1997). Comparison of sequences from the ribosomal DNA intergenic region of *Meloidogyne mayaguensis* and other major tropical root-knot nematodes. *Journal of Nematology, 29*(1), 16-22.

Blok, V.C. & Powers, T.O. (2009). Biochemical and molecular identification. In: Perry, R.N., Moens, M. & Starr, J.L. (Eds). *Root–knot nematodes*, 98-118.Oxfordshire, CABI Publishing.

Blok, V.C., Wishart, J., Fargette, M., Berthier, K. & Phillips, M.S. (2002). Mitochondrial DNA differences distinguishing *Meloidogyne mayaguensis* from the major species of tropical root-knot nematodes. *Nematology, 4*(7), 773-781. https://doi:10.1163/156854102760402559

Brito, J.A., Dickson, D.W., Kaur, R., Vau, S. & Stanley, J.D. (2015). The peach root-knot nematode: *Meloidogyne floridensis*, and its potential impact for the peach industry in Florida. *Nematology Circular, 224*, 7.

Brito, J.A., Powers, T.O., Mullin, P.G., Inserra, R.N., & Dickson, D.W. (2004). Morphological and molecular characterization of *Meloidogyne mayaguensis* isolates from Florida. *Journal of Nematology, 36*(3), 232-240.

Castagnone-Sereno, P., Leroy, F., Bongiovanni, M., Zijlstra, C. & Abad, P. (1999). Specific diagnosis of two root-knot nematodes, *Meloidogyne chitwoodi* and *M. fallax*, with satellite DNA probes. *Phytopathology, 89*(5), 380-400. https://doi.org/10.1094/PHYTO.1999.89.5.380

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research, 16*(22), 10881-10890. https://doi: 10.1093/nar/16.22.10881

Devran, Z., & Sogut, M.A. (2009). Distribution and identification of root-knot nematodes from Turkey. *Journal of Nematology, 41*(2), 128-133.

Dhandaydham, M., Charles, L., Zhu, H., Starr, J.L., Huguet, T., Cook, D.R., Prosperi, J.M., & Opperman, C. (2008). Characterization of root-knot nematode resistance in *Medicago truncatula*. *Journal of Nematology, 40*(1), 46-54.

Ekanayake, H.M.R.K. (2001). Histopathological changes caused by *Meloidogyne graminicola* in rice roots. *Annals of the Sri Lanka Department of Agriculture, 3*, 43-46.

Ekanayake, H.M.R.K. & Toida, Y. (1997). Nematode parasites of agricultural crops and their distribution in Sri Lanka. *JIRCAS Journal, 4*, 23-39.

Fargette, M. (1987). Use of esterase phenotype in the taxonomy of the genus *Meloidogyne.* 2. Esterase phenotypes observed in West African populations and their characterization. *Revue de Nematologie, 10*(1), 45-56.

Fargette, M. & Braaksma, R. (1990). Use of the esterase phenotype in taxonomy of the genus *Meloidogyne*. 3. A study of some 'B' race lines and their taxonomic position. *Revue de Nematologie, 13*(4), 375-386.

Goverse, A., de Engler, J.A., Verhees, J., van der Krol, S., Helder, J.H. & Gheysen, G. (2000). Cell cycle activation by plant parasitic nematodes. *Plant Molecular Biology, 43*(5-6), 747-761. https://doi:10.1023/a:1006367126077

Hall, T.A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series, 41*, 95-98.

Han, H., Cho, M.R., Jeon, H.Y., Lim, C.K. & Jang, H.I. (2004). PCR-RFLP identification of three major *Meloidogyne* species in Korea. *Journal of Asia-Pacific Entomology, 7*(2), 171-175. https://doi.org/10.1016/S1226-8615(08)60212-5

Hunt, D. & Handoo, Z. (2009). Taxonomy, identification and principal species. In Perry, R.N., Moens, M. & Starr, J.L. (Eds). *Root-knot nematodes*. London, CAP International, pp. 55-88.

Hutchinson, M.T. & Vythilingam, M.K. (1963). Distribution of plant parasitic nematodes in the soils of tea estates in Ceylon. *Tea Quarterly, 34*(3), 119-126.

Jeyaprakash, A., Tigano, M.S., Brito, J., Carneiro, R.M.D.G. & Dickson, D.W. (2006). Differentiation of *Meloidogyne floridensis* from *M. arenaria* using high-fidelity PCR amplified mitochondrial AT-rich sequences. *Nematropica, 36*, 1-12.

Jones, J.T., Haegemen, A., Danchin, E.G.J., Gaur, H.S., Helder, J., Jones, M.G.K., Kikuchi, T., Manzanilla-Lopez, R., Palomares-Rius, J.E., Wesemael, W.M.L. & Perry, R.N. (2013). Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology, 14*(9), 946-961. https://doi: 10.1111/mpp.12057

Joseph, S., Mekete, T., Danquah, W.B. & Noling, J. (2016). First report of *Meloidogyne haplanaria* infecting Mi-resistant tomato plants in Florida and its molecular diagnosis based on mitochondrial haplotype. *Plant Disease, 100* (7), 1438-1445. https://doi.org/10.1094/PDIS-09-15-1113-RE

Karssen, G. (2002). *The plant parasitic nematode genus Meloidogyne goldi, 1892 (Tylenchida) in Europe.* Leiden, Brill Academic Publishers.

Lamberti, F., Di Vito, M., Scrano, L., Chinappen, M. & Bucha, J. (1987). Root-knot nematodes, *Meloidogyne* species, in Mauritius and Rodrigues, *Plant Protection Bulletin, 35*, 43–50.

Landa, B.B., Rius, J.E.P., Vovlas, N., Carneiro, R.M.D.G., Maleita, C.M.N., Abrantes, I.M.D.O. & Castillo, P. (2008). Molecular characterization of *Meloidogyne hispanica* (Nematoda, Meloidogynidae) by phylogenetic analysis of genes within the rDNA in *Meloidogyne* spp. *Plant Disease, 92*(7), 1104-1110. https://doi: 10.1094/PDIS-92-7-1104.

McClure, M.A., Nischwitz, C., Skantar, A.M., Schmitt, M.E. & Subbotin, S.A. (2012). Root-knot nematodes in golf course greens of the Western United States. *Plant Disease*, *96*(5), 635-647. https://doi.org/10.1094/PDIS-09-11-0808

Mwesige, R., Seid, A. & Wesemael, W. (2016). Root-knot nematodes on tomatoes in Kyenjojo and Masaka districts in Uganda. *African Journal of Agricultural Research, 11*(38), 3598-3606. https://doi:10.5897/AJAR2016.11311

Mitkowski, N.A. & Abawi, G.S. (2003). Root-knot nematode. *The Plant Health Instructor*. *3*. https://doi:10.1094/PHI-I-2003-0917-01

Nishantha, K. M. D. W. P., Nugaliyadde, M. M., Costa, D. M. De, Nugaliyadde, L., Kumari, S. W. G. A. N., Dissanayake, D. M. I. C. B., & Malwenna, M. H. N. Y. (2025). Morphometric Taxonomy of *Meloidogyne* spp. Infesting Vegetable Crops in Sri Lanka via Female Perineal Pattern Analysis. *International Journal of Environment, Agriculture and Biotechnology*, *10*(2), 001–006. https://doi.org/10.22161/ijeab

Nugaliyadde, L., Dissanayake, D.M.N., Herath H.M.D.N., Dharmasena, C.M.D. and Jayasundara, D.M. (2001). Outbreak of rice root knot nematode, *Meloidogyne graminicola* (Golden & Birchfield) in Nikewaratiya, Kurunegala in Maha 2000/2001. (Short Communication). *Annals of the Sri Lanka Department of Agriculture, 3*, 373-374.

Oliveira, C.M.G., Monteiro, R. A. & Blok, V. C. (2011). Morphological and molecular diagnostics for plant-parasitic nematodes: Working together to get the identification done. *Tropical Plant Pathology, 36*(2), 65-73. https://doi.org/10.1590/S1982-56762011000200001

Onkendi, E.M. & Moleleki, L.N. (2013). Detection of *Meloidogyne enterolobii* in potatoes in South Africa and phylogenetic analysis based on intergenic region and the mitochondrial DNA sequences. *European Journal of Plant Pathology, 136*(1), 1-5. https://doi:10.1007/s10658-012-0142-y

Palomares- Rius J.E., Vovlas, N., Troccoli, G., Landa, B.B. & Castillo, P. (2007). A new root-knot nematode parasitizing sea rocket from Spanish Mediterranean coastal dunes: *Meloidogyne dunensis* n. sp. (Nematoda: Meloidogynidae). *Journal of Nematology, 39*(2), 190-202.

Perry, R.N., Moens, M. & Starr, J.L. (2009). *Meloidogyne* species - a diverse group of novel and important plant parasites. In: Perry, R.N., Moens, M. & Starr, J. L. (Eds). *Root knot nematodes.* Oxfordshire, CAB International, pp. 1-17

Powers, T. O. (2004). Nematode molecular diagnostics from bands to Barcodes. *Annual Review of Phytopathology, 42*, 367-383. https://doi: 10.1146/annurev.phyto.42.040803.140348.

Powers, T. O. & Harris, T. S. (1993). A Polymerase chain reaction method for identification five major *Meloidogyne* species. *Journal of Nematology, 25*(1), 1-6.

Powers, T. O., Mullin, P. G., Harris, T. S., Sutton, L. A. & Higgins, R. S. (2005). Incorporating molecular identification of *Meloidogyne* spp. into a large-scale regional nematode survey. *Journal of Nematology, 37*(2), 226-235.

Randig, O., Deau, F., Santos, M. F. A., Tiga no, M. S., Carneiro, R. M. D. G. & Catagnone-Sereno, P. (2009). A novel species-specific satellite DNA family in the I nvasive root-knot nematode *Meloidogyne mayaguensis* and its potential use for diagnostics. *European Journal of Plant Pathology, 125*, 485-495. https://doi.org/10.1007/s10658-009-9497-0

Swofford, D. L. (2002). PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods), 4.0b4a. Massachusetts: Sinauer Associates.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and evolution, 28*(10), 2731-2739. https://doi.org/10.1093/molbev/msr121

Taylor, A. L. & Sasser, J. N. (1978). Biology, identification and control of root-knot nematodes (*Meloidogyne* species). North Carolina: Cooperative publication of Department of Plant Pathology, North Carolina State University and US Agency for International Development, Raleigh, North Carolina.

Thomas, W. K. & A. C. Wilson. (1991). Mode and tempo of molecular evolution in the nematode Caenorhabditis: Cytochrome oxidase II and calmodulin sequences. *Genetics, 128*, 269-279. https://doi.org/ 10.1093/genetics/128.2.269

Tigano, M. S., Carneiro, R. M. D. G., Jeyaprakash, A., Dickson, D. W. & Adams, B. J. (2005). Phylogeny of *Meloidogyne* spp. based on 18S rDNA and the intergenic region of mitochondrial DNA sequences. *Nematology, 7*(6), 851-862. https://doi.org/ 10.1163/156854105776186325

Tigano, M. S., de Siqueira, K., Castagnone-Sereno, P., Mulet, K., Queiroz, P. & Dos Santos, M. (2010). Genetic diversity of the root-knot nematode *Meloidogyne enterolobii* and development of a SCAR marker for this guava-damaging species. *Plant Pathology, 59*(6), 1054-1061. https://doi.org/10.1111/j.1365-3059.2010.02350.x

Vanholme, B., de Meutter, J., Tytgat, T., Van Montagu, M., Coomans, A. & Gheysen, G. (2004). Secretions of plant-parasitic nematodes: A molecular update. *Gene, 332*(1), 13-27. https://doi:10.1016/j.gene.2004.02.024

Wesemael, W., Viaene, N. & Moens, M. (2011). Root-knot nematodes (*Meloidogyne* spp.) in Europe. *Nematology, 13*(1), 3-16. https://doi.org/10.1163/138855410X526831

Young, T. W. (1954). An incubation method for collecting migratory endo parasitic nematodes. *The Plant Disease Reporter, 38*(11), 794-795.

J.Natn.Sci.Foundation Sri Lanka 2026 54 (1): 101 - 111
DOI: http://dx.doi.org/10.4038/jnsfsr.v54i1.12547

RESEARCH ARTICLE

Environmental Science

Process optimization and yield enhancement of bioethanol production from *Musa paradisiaca* peel waste

S Yogendrarajah* **and K Ranganathan**
Department of Botany,Faculty of Science, University of Jaffna, Thirunelvely, Sri Lanka.

Submitted: 26 January 2025; Revised: 18 February 2026; Accepted: 05 March 2026

Abstract: The growing global demand for energy and the depletion of fossil fuels create serious challenges for energy security and increasing environmental problems such as pollution, global warming, rising sea levels, and CO_2 emissions. Switching to low-carbon fuels is necessary to fight climate change. Bioethanol, a renewable fuel made from plants, offers a sustainable way to meet energy needs while reducing environmental impact. This study investigates the use of *Musa paradisiaca* peels, an underutilized agro-waste, for bioethanol production. Different parts of the plant, including pseudostem, leaves, roots, peels, and bunch stalk, were tested. Among these, peels produced the highest levels of reducing sugars and bio-alcohol, making them the most effective carbon substrate for further optimization. Fermentation conditions were optimized step by step by altering one factor at a time while maintaining all other parameters constant. The optimization process involved determining the ideal fermentation time, hydrolysis agent, sulfuric acid concentration, yeast inoculum size, and substrate amount. Under the optimized conditions, a 40-hour fermentation time, the use of 0.75M sulfuric acid as the hydrolysis agent, 100 g/L yeast inoculum, and 75 g/100 mL substrate concentration resulted in a significant 8.15-fold increase in bio-alcohol yield compared to the non-optimized conditions. Analysis by GC revealed that the amount of ethanol in the crude alcohol sample was 89.6%. This study shows an effective way to convert banana peels into high purity bioethanol, supporting sustainable waste management and circular bioeconomy strategies.

Keywords: Bioethanol production, fermentation, *Musa paradisiaca*, *Saccharomyces cerevisiae.*

INTRODUCTION

With rising energy demand and the depletion of fossil fuel reserves, transitioning to low-carbon energy sources such as bioethanol has become essential to mitigate climate change and reduce CO_2 emissions (Broda et al., 2022). Bioethanol is produced from renewable, plant-based materials and contributes to energy security by utilizing agricultural residues. Lignocellulosic feedstocks such as rice straw and sugarcane bagasse are commonly used for second generation (2G) bioethanol production. Compared to first generation bioethanol, 2G bioethanol offers greater reductions in greenhouse gas emissions (Danmaliki et al., 2016; Sarkar et al., 2012). More advanced bioethanol technologies, including third-generation algae-based systems and fourth-generation processes involving genetically engineered microorganisms, have been developed to further improve production efficiency and sustainability (Edeh et al., 2021).

Bioethanol production involves pretreatment, hydrolysis, and fermentation. Pretreatment is used to disrupt the lignin structure, thereby improving accessibility to cellulose and hemicellulose. This can be achieved through physical, chemical, biological, or physicochemical methods (Meenakshisundaram et al., 2022; Dimos et al., 2019). Hydrolysis converts

* Corresponding author (sangavee23@gmail.com; https://orcid.org/0009-0007-7215-2045)

complex carbohydrates into fermentable sugars using acid, alkaline, or enzymatic methods (Devi et al., 2016; Binod et al., 2011). Fermentation, mainly carried out by *Saccharomyces cerevisiae*, converts fermentable sugars into ethanol (Hackl et al., 2011). Bioethanol has a high-octane rating, making it suitable for blending with gasoline in spark-ignition engines. However, its low vapor pressure presents challenges that can be addressed by blending it with more volatile fuels (Abdulsalam et al., 2021; Mushimiyimana et al., 2016).

In addition to its use as transportation fuel, bioethanol has wide industrial applications. It is commonly employed in chemical synthesis as a solvent, and also serves as a starting material to produce other biofuels (Yaverino et al., 2024). It acts as an important feedstock for producing bio-based chemicals through polymerization and oxidation, resulting in the formation of industrial plastics and other valuable compounds (Assaf et al., 2024). Bioethanol can be used in power generation, contributing to energy security and diversification of energy sources (Yaverino et al., 2024). Bioethanol is widely utilized in household products, including cleaning agents and personal care items due to its effective solvent properties (Mushimiyimana et al., 2016).

Bioethanol is a renewable energy source produced from agro-waste and other non-edible lignocellulosic biomass, offering an effective approach to waste reduction while minimizing environmental pollution (Mushimiyimana et al., 2016; Yaverino et al., 2024). Its high-octane number enhances engine performance and reduces greenhouse gas emissions, as CO_2 released during combustion is reabsorbed through photosynthesis (Gouthami et al., 2024). Because bioethanol can be produced from agricultural residues, forest residues, and solid waste, it supports sustainability and stimulates local economies (Abdulsalam et al., 2021; Yaverino et al., 2024). However, several challenges remain, including competition with food resources, land use concerns, and environmental impacts associated with crop cultivation and processing (Assaf et al., 2024; Wandscher Busanello et al., 2023).

Bioethanol production from lignocellulosic biomass, such as crop residues and wood, involves various technological challenges. Effective pretreatment methods are required for breaking down complex biomass into fermentable sugars. However, these processes can be costly and technically challenging. The existing technologies for producing second-generation (2G) bioethanol face difficulties in the pretreatment and hydrolysis stages, which can restrict their feasibility for large scale industrial implementation (Yaverino et al., 2024).

Banana cultivation, particularly the Kathali variety (*Musa paradisiaca*), is significant in Sri Lanka's Jaffna peninsula, contributing to the economy and food security. Although bananas are produced in large quantities, much of the plant, including peels, pseudostems, leaves, bunch stalks, and roots, are underutilized or discarded as waste. These parts contain sugars such as glucose, sucrose, and fructose, essential for fermenting bioethanol. Utilizing banana waste for biofuel production can improve waste management, reduce environmental pollution, and promote renewable energy, thereby contributing to sustainability and resource efficiency goals (Pazmiño et al., 2019). Current practices often leave banana residues to decompose in fields, which can lead to disease spread and water pollution Converting this biomass into bioethanol offers a practical approach to mitigating these environmental impacts while supporting cleaner energy production (Santa-Maria et al., 2013).

According to Dhas et al. (2025), red banana waste (*Musa acuminata*) was used as a substrate for bioethanol production using *Saccharomyces cerevisiae*. The bioethanol yield obtained was 35.9 ± 0.51 g/L, which was 11% higher than the yield under non-optimized conditions.

Santis-Espinosa et al., (2025) found that pretreatment of banana residues significantly increased bioethanol production, with acid hydrolysis yielding 55.18 g/L, sterilization 64.26 g/L, and pasteurization 34.11 g/L, showing that pretreatment improves sugar availability for fermentation). Uljanah et al., (2024) reported that banana pseudostem waste (*Musa balbisiana*) produced 41.5% (v/v) ethanol using 12% yeast over 13 days of fermentation, which was higher than yields from lower yeast concentrations and shorter fermentation times.

Alonso-Gómez et al., (2019) used whole unripe plantain (*Musa paradisiaca* L.) as raw material for bioethanol production. After 32 h of fermentation, the whole plantain produced 5.86 % of bioethanol, while the plantain pulp yielded 4.55 % of bioethanol.

Considerable research has been conducted on bioethanol production from different parts of the banana plant. However, there is a noticeable lack of publications specifically focused on bioethanol production from the peels of *Musa paradisiaca* using *Saccharomyces cerevisiae*. This presents a significant

opportunity to contribute novel findings in this area by utilizing underexplored biomass to meet energy demands. The objectives of the study were to identify the most underutilized parts of *Musa paradisiaca* that yield a higher quantity of bioalcohol and to optimize the culture conditions and media compositions to maximize bioalcohol output.

MATERIALS AND METHODS

Chemicals and culture media

All the chemicals used in this study were obtained from standard sources and were of analytical grade. Hydrolysis was carried out using 1 M H_2SO_4. The basal medium contained 4 g/L yeast extract, 8 g/L KH_2PO_4, 4 g/L $(NH_4)_2SO_4$, 2 g/L peptone, and 4 g/L $MgSO_4.7H_2O$. After autoclaving, 5 g of *Saccharomyces cerevisiae* (commercial yeast, 50 g/L) was added to the conical flask containing the media.

Collection of samples

Musa paradisiaca leaves, bunch stalks, roots, peels, and pseudostems were collected from several sites in the Jaffna district. The samples were packed in polythene and brought to the laboratory.

Inoculum preparation

Saccharomyces cerevisiae (Baker's yeast) was purchased from Cargill's Food City, Jaffna, Sri Lanka. *Saccharomyces cerevisiae* was cultured in 100 ml of sterile glucose (50 g/L) and sucrose (50 g/L) solution by inoculating yeast strains (50 g/L). The culture was shaken at 100 rpm at room temperature for 18 h.

Substrate preparation

Leaves from *Musa paradisiaca* were collected, and the midribs removed. After cleaning under running water, the leaves were air dried. Thereafter, they were oven-dried at 60 ˚C until reaching a constant weight. Finally, the dried leaves were ground into a fine powder and stored in plastic bottles. Bunch stalks, roots, peels, and pseudostems of *Musa paradisiaca* were collected separately, thoroughly cleaned with clean running tap water, sliced into small pieces using a sharp knife, and air dried. All the samples were oven dried at 60 ˚C to a constant weight. Finally they were ground thoroughly with a grinder and stored in plastic bottles.

Production of bioalcohol and alcohol measurement

Physical pre-treatment

After cleaning with tap water, the materials (leaves, bunch stalks, roots, peels, and pseudostems) were chopped into small pieces, air dried, and then oven dried at 60 ˚C until their weight remained constant. The materials were then ground thoroughly in a grinder and stored in bottles.

Pre-treatment via autoclaving

Each sample, weighing thirty grams, was placed in separate 500 mL conical flasks with 100 mL of distilled water. Flasks were sealed air-tight with cotton wool. The samples were then autoclaved for 15 minutes at 121˚C and 1 atm of pressure and left to cool.

Acid hydrolysis

Samples were acid hydrolysed with 100 mL of 1 M H_2SO_4 and autoclaved at 121˚C, 1 atm pressure for 15 minutes. The mixture was cooled to room temperature, filtered through muslin cloth, and the filtrates were centrifuged for 15 minutes at 8000 rpm. The supernatants were collected and neutralized with 4 M NaOH.

Fermentation process

The fermentation medium (75 mL) containing 4 g/L yeast extract, 8 g/L KH_2PO_4, 4 g/L $(NH_4)_2SO_4$, 2 g/L peptone, and 4 g/L $MgSO_4.7H_2O$ was added with 300 mL of neutralized solution in separate conical flasks. The mixture was then autoclaved at 121˚ C and 1 atm pressure for 15 minutes and the solutions were left to cool. The solutions were inoculated aseptically with 10% of an 18 h old culture of *Saccharomyces cerevisiae* inoculum. The mixture was allowed to ferment for 24 h at room temperature while being incubated at 100 rpm in an orbital shaker. For fermentation, the conical flasks were firmly sealed with cotton plugs to create an anaerobic environment.

Measuring alcohol percentage

From the fermented samples, 50 mL were transferred to a falcon tube and centrifuged at full speed (3000 rpm). The supernatant was carefully collected, and the pellet discarded. The ebulliometer method was then used to determine the alcohol percentage of the supernatant (Gnanasegaram et al., 2024). The five samples taken

from *Musa paradisiaca* underwent the same process as before.

Determination of reducing sugar

The reducing sugar content was measured using the 3,5-dinitrosalicylic acid (DNS) method. Standard glucose solutions were prepared by diluting different volumes (0.2–1.0 mL) of a stock glucose solution (1.0 g/L) into labeled test tubes, with the total volume adjusted to 1.0 mL using distilled water. Each tube was then treated with 1.0 mL of DNS reagent and heated in a boiling water bath for 5 minutes. After cooling, 10.0 mL of distilled water was added to each tube, and the absorbance was measured spectrophotometrically at 550 nm with a reagent blank serving as the reference. The reagent blank was prepared by using 1.0 mL of distilled water instead of the standard glucose solution. The glucose concentrations of unknown samples were measured by comparing them to the standard glucose curve after they were treated similarly (Christy et al., 2023).

Distillation and gas chromatography

Absolute ethanol (≥99.9%), methanol (≥99.9%), 3-methyl-1-butanol (≥99%), and 2-methyl-1-propanol (≥99%) standards were obtained from Sigma-Aldrich (USA). High purity helium (99.999%) was used as the carrier gas throughout the chromatographic analysis.

After fermentation was completed, the broth was initially filtered through muslin cloth followed by Whatman No. 1 filter paper to remove any residual solids. The clear filtrate was then distilled fractionally using a conventional laboratory apparatus consisting of a 500 mL round bottom flask fitted with a fractionating column, thermometer, Liebig condenser, and a receiving flask. Distillation was conducted using a heating mantle, and the fraction corresponding to 78–85 °C was collected. The collected distillate was stored in airtight amber bottles at 4 °C for further analysis.

The chemical composition of the distilled samples was determined using a Gas Chromatography–Mass Spectrometry (GC–MS) system (Agilent 7890B GC coupled with 5977A MSD, USA) equipped with an HP-5MS capillary column (30 m × 0.25 mm i.d., 0.25 µm film thickness). Helium served as the carrier gas, maintained at a constant flow rate of 1.0 mL min^{-1}. A 1 µL sample was injected in split mode (10:1) with the injector temperature adjusted to 250 °C. The oven temperature was initially set at 40 °C for 3 min, then increased at 10 °C min^{-1} to 200 °C, and held for 5 min. Mass spectrometric detection was performed using electron impact ionization at 70 eV, with an ion source temperature of 230 °C and a scan range of m/z 30–300. Compounds were identified by comparing their mass spectra with the NIST library database and confirmed by retention times of authentic standards. The relative percentage composition was calculated based on peak area normalization.

Optimization of conditions for bioalcohol production

Selection of suitable substrate

Leaves, bunch stalks, roots, peels, and pseudostems of *Musa paradisiaca* were used as substrates which underwent pre-treatment, acid hydrolysis and fermentation. The alcohol percentage produced from each substrate was determined and the substrate with the highest percentage of alcohol was selected for further optimization studies.

Optimization of time

Thirty grams of banana peel sample, after pre-treatment and acid hydrolysis, was fermented for 50 h on an orbital shaker at 100 rpm, room temperature. The alcohol percentage was measured every 10 h, and the time with the highest alcohol content was selected for further analysis.

Optimization of different hydrolysis agents

The banana peel sample was hydrolyzed using different acids (1 M H_2SO_4, HCl, and HNO_3) and alkalis (1 M NaOH and KOH), and then autoclaved. The mixture was fermented with *Saccharomyces cerevisiae* for 40 h and the alcohol percentage was determined. The hydrolysis agent, that produced the most alcohol was chosen for further studies.

Optimization of sulfuric acid concentration in acid hydrolysis

The pre-treated banana peel samples were hydrolysed with different concentrations of sulfuric acid (0.50, 0.75, 1.00, 1.25, and 1.50 M). After adding the resultant mixture to fermentation media containing *Saccharomyces cerevisiae*, it was incubated for 40 h at room temperature at 100 rpm and the alcohol percentage determined. The concentration of sulfuric acid, which produced the highest alcohol yield was chosen for further studies.

Optimization of the amount of yeast inoculum

Thirty grams of banana peel samples were pre-treated with distilled water and then acid hydrolyzed using 0.75 M sulfuric acid. Different amounts of yeast inoculum (25, 50, 75, 100, 125, and 150 g/L) were added to conical flasks containing fermentation media. After inoculating aseptically at room temperature, it was left to ferment for 40 h at room temperature in an orbital shaker at 100 rpm and the alcohol percentage was determined. The yeast inoculum concentration that produced the highest alcohol percentage was selected for further studies.

Optimization of the amount of substrate

Different quantities of banana peel samples (15, 30, 45, 60, 75, 90 g /100 mL) were used as substrate to produce alcohol by fermentation after pre-treatment with 0.75 M sulfuric acid for hydrolysis. The fermentation process was carried out using a yeast inoculum concentration of 100 g/L for 40 h and the alcohol percentage was determined.

Statistical analysis

Each experiment was conducted in triplicate, and the average results were displayed graphically. Minitab 17.0 was used for statistical analysis. Tukey's multiple comparison test was used to identify significant differences, at a significance level of $p < 0.05$, after analyzing the data using one-way ANOVA.

RESULTS AND DISCUSSION

Effect of different substrates on bioalcohol production

Various parts of the *Musa paradisiaca* plant such as leaves, bunch stalk, roots, peels, and pseudostem were used as substrates for bioalcohol production. The alcohol percentage was measured under non-optimized conditions using a 24 h fermentation period, 1 M sulfuric acid hydrolysis, 50 g/L yeast inoculum, and a substrate concentration of 30 g/100 mL. *Musa paradisiaca* peels yielded a significantly higher quantity of alcohol (0.2%) compared to other underutilized parts of *Musa paradisiaca* (root 0.03%, pseudostem 0.06%, leaf 0.1%, bunch stalk 0.06%) after 24 h of fermentation. The reducing sugar content also followed the same trend. Banana peels contain considerable amounts of carbohydrates, primarily in the form of fermentable sugars, which are crucial for bioalcohol production. Alcohol production was notably lower in the roots, leaves, bunch stalk, and pseudo-stem due to the absence of sufficient fermentable substances (Figure 1). Therefore, the peels of *Musa paradisiaca* were selected as a carbon substrate for further optimization.

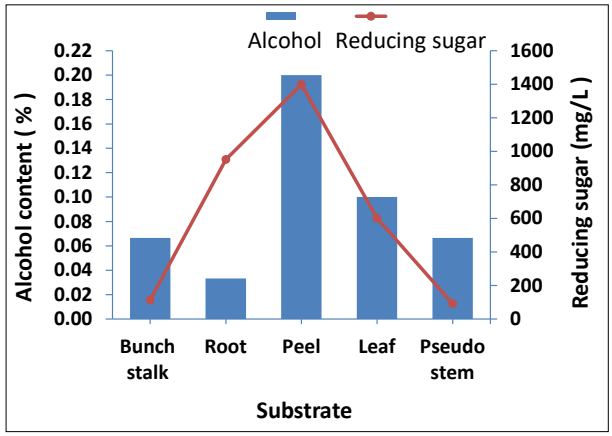


Figure 1: Effect of different plant parts of *Musa paradisiaca* as substrates on bioalcohol production using *Saccharomyces cerevisiae*.

Ilangarathna et al., (2022) used various coconut plant parts, such as young coconut fiber, coconut husk fiber, dried coconut leaves, green coconut leaves, and roots of coconut as carbon substrates for bioethanol production. Among these substrates, green coconut leaves produced the highest quantity of bioethanol compared to others.

Zainab et al., (2022) reported that among the three substrates utilized for bioethanol production - sugarcane bagasse, rice husk, and corn cob - the highest quantity of bioethanol was produced by sugarcane bagasse. Lignocellulosic biomass presents a viable alternative for producing bioethanol, being widely available, cost-effective, and not competing with food or feed crops (Bušić et al., 2018). The low lignin content in banana peels enables easier break down during the hydrolysis and pre-treatment stages. Less lignin indicates that the sugars can be released more efficiently because it may hinder cellulose's accessibility (Hamzah et al., 2019).

Effect of fermentation time

With an increasing fermentation time from 10 to 50 h, the maximum alcohol yield was observed at 40 h (Figure 2), with the yield increasing from 0.2% to 0.53% (v/v). Therefore, 40 h was determined to be the optimum fermentation time and chosen for the subsequent experiments.

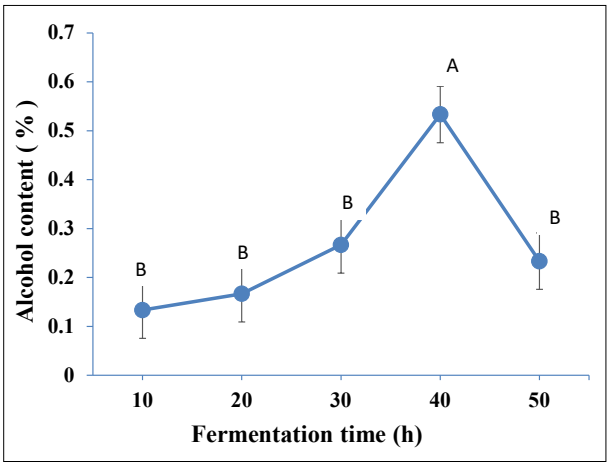


Figure 2: Effect of fermentation time on bioalcohol production from peels of *Musa paradisiaca* using *Saccharomyces cerevisiae.* Significant differences between mean values are indicated by different letters (A, B).

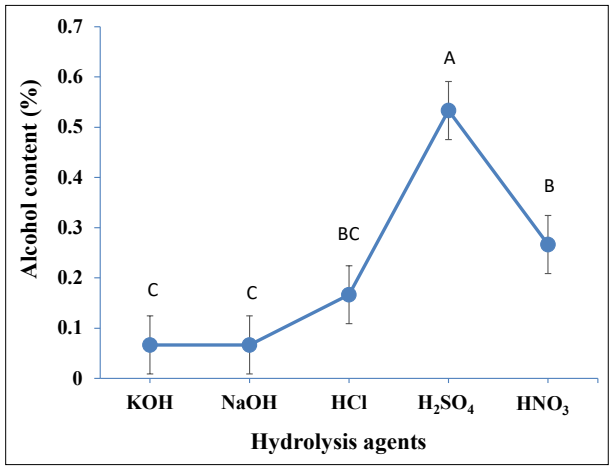


Figure 3: Effect of different hydrolysis agents on bioalcohol production from peels of *Musa paradisiaca* using *Saccharomyces cerevisiae*. Significant differences between mean values are indicated by different letters (A, B, C).

After 30 h of fermentation, a significantly higher bioethanol yield of 10.8% was obtained, which is 1.56 times greater than the yield from non-optimized conditions, using *Saccharomyces cerevisiae* on paddy husk of *Dahanala Red Naadu* (Christy et al., 2024). On the 4th day of fermentation, significantly higher quantities of bioethanol were produced from the whole plant of *Parthenium hysterophorus* using *Saccharomyces cerevisiae* (Gnanasegaram et al., 2024). The length of fermentation has an impact on yeast cell development. Furthermore, yeast cells expend energy during the first phase of fermentation to adjust to the growing environment. However, if the fermentation process is too long, the system's increased bioethanol concentrations could be toxic to the fermenting cells (Christy et al., 2023).

Effect of different hydrolysis agents

After 40 h of fermentation with *Saccharomyces cerevisiae*, hydrolysis of *Musa paradisiaca* peel using 1 M sulfuric acid produced an alcohol yield of 0.53% (v/v), which was significantly higher than the yields obtained with other acids tested (1 M nitric acid, and 1 M hydrochloric acid) and alkaline solutions (1 M sodium hydroxide and 1 M potassium hydroxide) (Figure 3). Therefore, acid hydrolysis using sulfuric acid was chosen for further studies. El-Tayeb et al., (2012) reported that the highest bioethanol was produced when sugar beet waste was treated with sulfuric acid (5% v/v) compared to hydrolyzing with hydrochloric acid, and phosphoric acid.

Production of bioethanol depends on the selection of the most efficient hydrolysis method for biomass. One important stage that greatly affects the process's overall efficiency is the hydrolysis of lignocellulosic material into fermentable sugars (Binod et al., 2011). The aim of the hydrolysis process is to break down the polysaccharide structures, making them more accessible for conversion into monomers by acids or alkalis (Christy et al., 2023).

The primary benefit of acid hydrolysis is its capacity to penetrate lignin and break down cellulose and hemicellulose polymers into individual sugar molecules without the need for any prior pretreatment of the biomass (Verardi et al., 2012). Since concentrated strong acids such as HCl and H_2SO_4 are effective cellulose hydrolyzers, they do not require additional enzymes thereafter, they are frequently used to treat lignocellulosic materials (Harmsen et al., 2010).

Effect of sulfuric acid concentration

When *Musa paradisiaca* peels were hydrolyzed with 0.75 M sulfuric acid, the alcohol yield increased from 0.53% to 0.63% (v/v), which was significantly higher than the yields obtained with other acid concentrations (Figure 4). Therefore, 0.75 M sulfuric acid was chosen for hydrolysis. The concentration of sulfuric acid is important to the hydrolysis process to produce bioalcohol from *Musa paradisiaca* peels. In this study, the production of bioalcohol increased significantly as the concentration of

sulfuric acid increased and reached a maximum value; thereafter, the production of bioalcohol was reduced by further increases in sulfuric acid concentration. The breakdown of sugars by the strong acidity may explain the decrease in bioethanol concentration (Kefale et al., 2012).

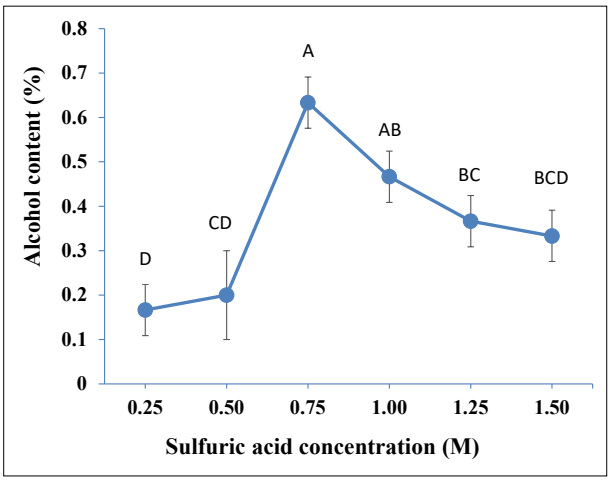


Figure 4: Effect of sulfuric acid concentration on bioalcohol production from peels of *Musa paradisiaca* using *Saccharomyces cerevisiae*. Significant differences between mean values are indicated by different letters (A, B, C, D).

Christy et al., (2023) reported that *C. globularis* produced significantly higher yields of bioethanol and reducing sugar when 0.75 M sulfuric acid was used. In agro-industrial wastes, acid concentrations of 1% sulfuric (v/v) produced a higher quantity of bioethanol than 5% (v/v) (El-Tayeb et al., 2012). High concentrations of H_2SO_4 can lead to browning or charring of the hydrolysate producing unwanted byproducts such as furfural and 5-hydroxymethylfurfural, which are harmful to *Saccharomyces cerevisiae*, hindering fermentation and reducing overall bioethanol production (Mitiku et al., 2020).

Effect of the amount of yeast inoculum

Musa paradisiaca peel produced a significantly higher alcohol yield when the *S. cerevisiae* inoculum concentration was 100 g/L, with the yield increasing from 0.63% to 0.93% (v/v) compared to other inoculum concentrations used (Figure 5). This concentration was selected as the optimum for further studies. Nikolić et al., (2009) tested three different inoculum concentrations (2%, 10%, and 20% (w/v)) using corn meal hydrolyzates; the highest ethanol yield was produced by an inoculum concentration of 2% (w/v). When the yeast inoculum size was increased to 0.5 g/100 mL, the ethanol yield increased significantly by 1.11 times (from 0.90% to 1.00%) in comparison to the non-optimized control inoculum size (Vivekanandaraja et al., 2021).

The inoculum concentration influences sugar utilization and ethanol production, but has little impact on the final ethanol concentration. Once the biocatalyst reaches a certain concentration, they saturate the system, which can reduce the efficiency of bioethanol production. High inoculum concentrations may saturate the system reducing bioethanol efficiency. Higher inoculum concentrations can lead to nutrient limitations or overcrowding, ultimately decreasing the overall ethanol yield (Laopaiboon *et al*., 2007). The inoculum size was increased by adding more yeast cells to the fermentation flask. Increasing the inoculum size led to the rapid increase in the yield of ethanol due to quick growth and multiplication of the yeast cells in the flask. Most of the substrate, which consisted of fermentable sugars, was quickly consumed and transformed into ethanol when additional yeast cells were added. However, yield of ethanol began to decline with further increases in inoculum size, which competed for the limited nutrients in the reactor reducing the multiplication of yeast cells and producing ethanol (Elizabeth et al., 2018).

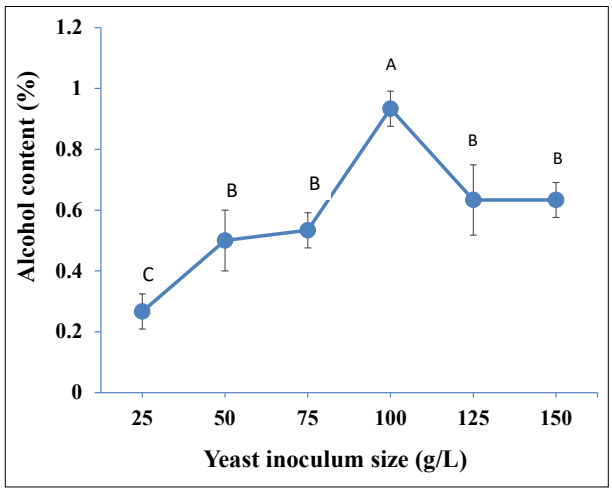


Figure 5: Effect of yeast inoculum size on bioalcohol production from peels of *Musa paradisiaca* using *Saccharomyces cerevisiae*. Significant differences between mean values are indicated by different letters (A, B, C).

Effect of amount of substrate

When evaluating different concentrations of *Musa paradisiaca* peel substrate, the highest alcohol yield was achieved at 75 g/100 mL after 40 h of fermentation, with the yield increasing from 0.93% to 1.63% (v/v). When the concentration of peel ranges from 15 g/100 mL to 75 g/100 mL, the yield of bioalcohol increases exponentially, but it takes longer for the yeast cells to adjust to the medium. However, at concentrations exceeding 75 g/100 mL, a slight decline in alcohol yield was observed (Figure 6). Therefore, 75 g/100 mL of peel substrate was selected as the optimized amount and used for further studies.

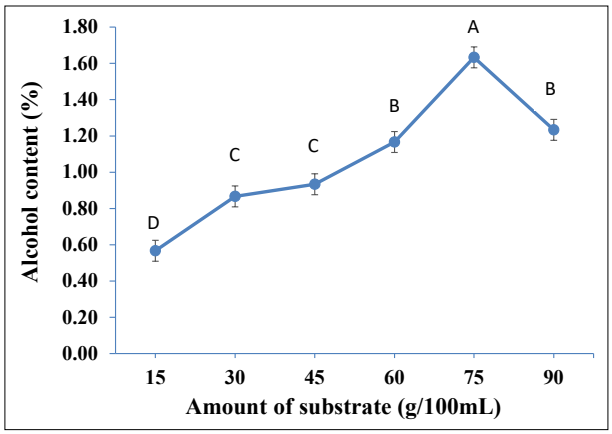


Figure 6: Effect of the amount of substrate on bioalcohol production from peels of *Musa paradisiaca* using *Saccharomyces cerevisiae*. Significant differences between mean values are indicated by different letters (A, B, C, D).

Using oil palm empty fruit bunches as the substrate, Mardawati et al., (2019) determined the production of bioethanol at different concentrations (3%, 6%, 9%, 12%, and 15%); after three h of fermentation, the maximum ethanol concentration was 0.25 g/L, with the highest ethanol yield at 9% concentration. Using three concentrations of cassava peel (10%, 15%, and 20% w/v) as a substrate Abdullah et al., (2024), found 15% concentration yielded the highest bioethanol, with a maximum concentration of 6.172 g/L. Since there is more raw material available for fermentation, increasing the amount of substrate can result in higher ethanol yields but may also cause substrate inhibition. High substrate concentrations may hinder yeast cell activity, a phenomenon known as substrate inhibition. This may result in less ethanol being obtained from fermentable sugars, which would lower the overall quantity of bioethanol produced during fermentation (Tenkolu et al., 2022).

Osmotic stress, which negatively affects yeast cells and reduces their growth and metabolic activity, is the reason for this decrease in ethanol yields (Chang et al., 2018). After the optimization process, the *Musa paradisiaca* peel substrate produced 8.15 times more alcohol than non-optimized conditions (increasing from 0.2 % to 1.63 %).

Distillation and GC analysis for ethanol purity assessment

The GC results for alcohol obtained from *Musa paradisiaca* peel substrate are shown in Figure 7. This showed that the most dominant compound in the sample was ethanol (89.6%) followed by 3-methyl-1-butanol (6.4%), and the least dominant compound was 2-methyl-1-propanol (3.8%). Methanol was found in trace amounts (0.2%). Similarly in the GC results for *C. globularis*, ethanol was the dominant compound (90.1%), followed by 3-methyl-1-butanol (7.3%), with 2-methyl-1-propanol (2.6%) the the least abundant (Christy et al., 2023).

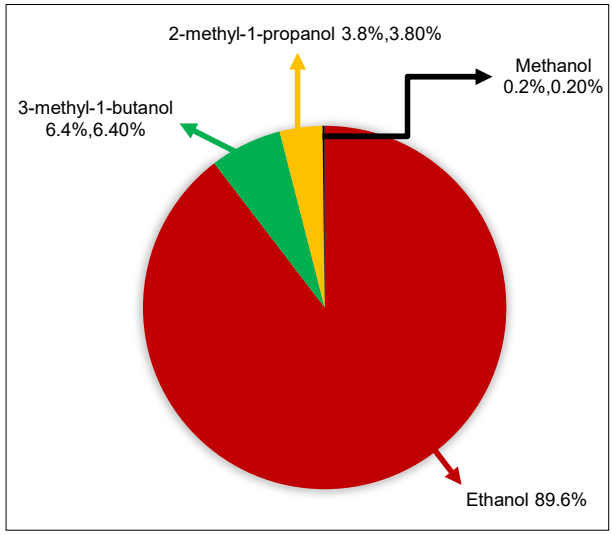


Figure 7: Gas chromatography result for alcohol obtained from *Musa paradisiaca* peel sample.

The high ethanol content and minimal methanol contents indicate efficient fermentation and distillation. The low concentration of higher alcohols and trace impurities suggest improved fuel quality, as excessive fuel alcohols

and methanol can adversely affect combustion efficiency and engine performance (Fini et al., 2021). Therefore, our results demonstrate that *Musa paradisiaca* peel substrate produces bioethanol with comparatively fewer undesirable by-products, supporting its potential as a sustainable and efficient biofuel feedstock.

Table 1: Comparison of optimized ethanol yield (% v/v) obtained in this study with reported yields from banana-based feedstocks.

Study	Substrate	Ethanol yield (% v/v)
This study	*Musa paradisiaca* peels	1.63
Zulnazri et al., 2023	Kepok banana peels (*Musa paradisiaca L.*)	5.04
Bilyartinus et al., 2021	Ambon Banana (*Musa paradisiaca var. sapientum Linn*) Peels	6
Benjamin et al., 2014	Banana peels	6.54
Ogunsuyi et al., 2021	Plantain stem biomass (*Musa paradisiaca L.)*	8.04
Kularathne et al., 2021	Overripe Banana (embul kesel variety)	13.39

Table 1 shows that the ethanol yield obtained in this study (1.63% v/v) is lower than several previously reported studies. The variation in yield may be attributed to differences in substrate type, pretreatment methods, fermentation conditions, and microbial strains used.

CONCLUSION

The peel of *Musa paradisiaca*, an underutilized resource, serves as an efficient substrate for alcohol production when fermented with *Saccharomyces cerevisiae* in liquid media. Optimization of fermentation parameters, including a 40 h fermentation period, sulfuric acid hydrolysis at a concentration of 0.75 M, a yeast inoculum of 100 g/L and 75 g/100 mL of *Musa paradisiaca* peel substrate, resulted in an 8.15-fold increase in alcohol yield, from 0.2% to 1.63% (v/v). Gas chromatography analysis confirmed that the crude alcohol sample contained 89.6% ethanol.

REFERENCES

Abdullah, N. A. E. P. (2024). Kinetics study of bioethanol production from cassava peels waste using *Saccharomyces diastaticus*. *International Journal of Chemical and Biochemical Sciences* (IJCBS), *25*(19), 884-892. https://doi.org/10.62877/106-IJCBS-24-25-19-106

Abdulsalam, A. A. A. S. M., Aliyu, B. B. L., Danillela, U. Y., Ahmad, Z. U., Arul Kumar, B., & Gbadamosi, L. A. (2024). Bioethanol: A sustainable liquid fuel as a substitute to gasoline. *International Journal for Science Technology and Engineering*, *12*(7), 360-371. https://doi.org/10.22214/ijraset.2024.63555

Alonso-Gómez, L. A., Heredia-Olea, E., Serna-Saldivar, S. O., & Bello-Pérez, L. A. (2019). Whole unripe plantain (Musa paradisiaca L.) as raw material for bioethanol production. *Journal of the Science of Food and Agriculture, 99*(13), 5784-5791. https://doi.org/10.1002/jsfa.9847

Assaf, J. C., Mortada, Z., Rezzoug, S. A., Maache-Rezzoug, Z., Debs, E., & Louka, N. (2024). Comparative review on the production and purification of bioethanol from biomass: A focus on corn. *Processes, 12*(5), 1001. https://doi.org/10.3390/pr12051001

Benjamin, C., Singh, P. K., Dipuraj, P. S., Singh, A., Rath, S., Kumar, Y., ... & Peter, J. (2014). Bio-ethanol production from banana peel by simultaneous saccharification and fermentation process using cocultures *Aspergillus niger* and *Saccharomyces cerevisiae*. *International Journal of Current Microbiology and Applied Sciences, 3*(5), 84-96.

Bilyartinus, G., & Siswanto, A. P. (2021). the effect of bacillus subtilis on bioethanol production from ambon banana (*Musa paradisiaca* var. sapientum Linn) peels by using fermentation process. *Journal of Vocational Studies on Applied Research, 3*(2), 26-30. http://dx.doi.org/10.14710/jvsar.v3i2.11081

Binod, P., Janu, K. U., Sindhu, R., & Pandey, A. (2011). Hydrolysis of Lignocellulosic Biomass for Bioethanol Production. In *Biofuels* (pp. 229–250). Elsevier. https://doi.org/10.1016/B978-0-12-385099-7.00010-3

Broda, M., Yelle, D. J., & Serwańska, K. (2022). Bioethanol production from lignocellulosic biomass - Challenges and solutions. *Molecules, 27*(24), 8717.

https://doi.org/10.3390/molecules27248717

Bušić, A., Marđetko, N., Kundas, S., Morzak, G., Belskaya, H., Ivančić Šantek, M., Komes, D., Novak, S., & Šantek, B. (2018). Bioethanol production from renewable raw materials and its separation and purification: A review. *Food Technology and Biotechnology, 56*(3), 289-311. https://doi.org/10.17113/ftb.56.03.18.5546

Chang, Y.-H., Chang, K.-S., Chen, C.-Y., Hsu, C.-L., Chang, T.-C., & Jang, H.-D. (2018). Enhancement of the efficiency of bioethanol production by Saccharomyces cerevisiae via gradually batch-wise and fed-batch increasing the glucose concentration. *Fermentation, 4*(1), 45. https://doi.org/10.3390/fermentation4020045

Christy, E. J. S. B. A., Branavan, K., & Kapilan, R. (2024). Bioethanol production from rice husk of Dahanala red naadu using Saccharomyces cerevisiae and yield enhancement by optimization of growing conditions. *Vingnanam Journal of Science, 19*(1), 37-43.

Christy, E. J. S. B. A., Kapilan, R., Wickramasinghe, I., & Wijesekara, I. (2023). bioethanol production from Chara globularis using yeast and yield improvement by optimization of conditions. *Ceylon Journal of Science, 52*(2), 219-229. https://doi.org/10.4038/cjs.v52i2.8042

Christy, E.J.S.B.A., Kapilan, R., Wickramasinghe, I. and Wijesekara, I. (2023) Optimization of pre-treatment and culture conditions for bioethanol yield enhancement from Azolla filiculoides substrate using Saccharomyces cerevisiae, *Journal of Dry Zone Agriculture, 9*(1), 1-25. https://doi.org/10.4038/jdza.v9i1.64

Danmaliki, G., Usman, B., Musa, A., & Ahmad, S. (2016). Bioethanol production from banana peels. *IOSR Journal of Environmental Science*, (JESTFT), 10, 56-62. https://doi.org/10.9790/2402-1006025662

Devi, S., Dhaka, A., & Singh, J. (2016). Acid and alkaline hydrolysis technologies for bioethanol production: An overview. *International Journal of Advanced Technology and Engineering Science, 4*(6), 94-106.

Dhas, D. S. D., Thiraviam, A. G. P., Muthuramamoorthy, M., Alodhayb, A., & Arokiyaraj, S. (2025). Bioethanol production from red banana waste via Saccharomyces cerevisiae under optimized conditions. *BioResources*, *20*(4), 10685–10705. https://doi.org/10.15376/biores.20.4.10685-10705

Dimos, K., Paschos, T., Louloudi, A., Kalogiannis, K. G., Lappas A. A, Papayannakos, N., Kekos. D., & Mamma, D. (2019). Effect of various pretreatment methods on bioethanol production from cotton stalks. *Fermentation, 5*(1), 5. https://doi.org/10.3390/fermentation5010005

Edeh, I. (2021). Bioethanol production: An overview. Bioethanol technologies, Chapter 1, 2-3. https://doi.org/10.5772/intechopen.94895

Elizabeth, O., Akwayo, I., Taiwo, O., Obanla, O., Ayoola, A., Ojewumi, E., & Oyeniyi, E. (2018). Bioconversion of sweet potato peel waste to bioethanol using Saccharomyces cerevisiae. *International Journal of Environmental Science and Technology, 8*(3), 46-54.

El-Tayeb, T. S., Abdelhafez, A. A., Ali, S. H., & Ramadan, E. M. (2012). Effect of acid hydrolysis and fungal biotreatment on agro-industrial wastes for obtainment of free sugars for bioethanol production. *Brazilian Journal of Microbiology, 43*(4), 1523-1535. https://doi.org/10.1590/S1517-83822012000400037

Fini, A.T and Fattahi, A. (2021). Bioethanol production from wastes: An experimental evaluating study for Iran. Journal *of Renewable Energy and Environment*, *8*(3), 86-93.

Gnanasegaram, A., & Kapilan, R. (2024). Bioethanol from Parthenium hysterophorus substrate by Saccharomyces cerevisiae. Journal of the University of Ruhuna, 12(1), 28-34. https://doi.org/10.4038/jur.v12i1.8022

Gouthami, Y., Praveen, Gidagiri., G, Bhuvaneshwari., G.S., Lekshmi., & Ms., Shameena. (2024). Bioethanol production from agricultural waste. *Futuristic Trends in Agriculture Engineering and Food Science, 3*, 536-554. https://doi.org/10.58532/V3BCAG15P4CH8

Hackl, M., Jakobi, T., Blom, J., Doppmeier, D., Brinkrolf, K., Szczepanowski, R., Bernhart, S. H., Höner zu Siederdissen, C., Hernandez Bort, J. A., Wieser, M., Kunert, R., Jeffs, S., Hofacker, I. L., Goesmann, A., Pühler, A., Borth, N., & Grillari, J. (2011). Next-generation sequencing of the Chinese hamster ovary microRNA transcriptome: Identification, annotation, and profiling of microRNAs as targets for cellular engineering. *Journal of Biotechnology, 153*(1-2), 62-75. https://doi.org/10.1016/j.jbiotec.2011.02.011

Hamzah, M. A., Alias, A. B., & Ahmad, N. E. (2019). Production of biofuel (bio-ethanol) from fruit waste: Banana peels. *International Journal of Engineering and Advanced Technology, 9*(1), 5897-5901. https://doi.org/10.35940/ijeat.A3024.109119

Harmsen, P., Huijgen, W., Bermudez, L., & Bakker, R. (2010). Literature review of physical and chemical pretreatment processes for lignocellulosic biomass. *Biomass and Bioenergy, 28*(5), 494-506.

Ilangarathna, D. A., & Kapilan, R. (2022). Bioethanol production from coconut fiber wastes using Saccharomyces cerevisiae. *Vingnanam Journal of Science, 17*(1), 9-19 https://doi.org/10.4038/vingnanam.v17i1.4205

Kefale, A., Redi, M., & Asfaw, A. (2012). Potential of bioethanol production and optimization test from agricultural waste: The case of wet coffee processing waste (pulp). I*nternational Journal of Renewable Energy Research, 2*(3), 446-450.

Kularathne, I. W., Gunathilaka, C. A., Ratnaweera, A. C., Kalpage, C. S., Rajapakse, S., & Gamage, P. (2021). Optimization of fermentation process parameters for bioethanol production from Sri Lankan overripe fruits. Engineer: *Journal of the Institution of Engineers, Sri Lanka, 54*(1). http://doi.org/10.4038/engineer.v54i1.7368

Laopaiboon, L., Thanonkeo, P., Jaisil, P., & Laopaiboon, P. (2007). Ethanol production from sweet sorghum juice in batch and fed-batch fermentations by Saccharomyces cerevisiae. *World Journal of Microbiology and Biotechnology, 23*(11):1497-1501. https://doi.org/10.1007/s11274-007-9383-x

Mardawati, E., Putri, A. V., Yuliana, T., Rahimah, S., Nurjanah, S., & Hanidah, I. (2019, February). Effects

of substrate concentration on bioethanol production from oil palm empty fruit bunches with simultaneous saccharification and fermentation (SSF). *IOP Conference Series: Earth and Environmental Science, 230*(1), 012079. https://doi.org/10.1088/1755-1315/230/1/012079

Meenakshisundaram, S., Fayeulle, A., Léonard, E., Ceballos, C., Liu, X., & Pauss, A. (2022). Combined biological and chemical/physicochemical pretreatment methods of lignocellulosic biomass for bioethanol and biomethane energy production: A review. *Applied Microbiology, 2*(4), 716-734. https://doi.org/10.3390/applmicrobiol2040055

Mitiku, A. A., & Hatsa, T. M. (2020). Bioethanol production from decaying fruit peels using Saccharomyces cerevisiae. *International Journal of Current Research and Academic Review, 8*(5), 50-59.

Mushimiyimana, I., & Tallapragada, P. (2016). Bioethanol production from agro wastes by acid hydrolysis and fermentation process. *Journal of Renewable Energy, 75*(6), 383-388.

Nikolić, S., Mojović, L., Rakin, M., Pejin, D., & Nedović, V. (2009). Effect of different fermentation parameters on bioethanol production from corn meal hydrolyzates by free and immobilized cells of Saccharomyces cerevisiae var. ellipsoideus. *Journal of Chemical Technology and Biotechnology*: *84*(4), 497-503. https://doi.org/10.1002/jctb.2068

Ogunsuyi, H. O., & Olawale, C. A. (2021). Evaluation of plantain biomass (Musa paradisiaca L.), as feedstock for bio-ethanol production. *Green and Sustainable Chemistry, 11*(2), 59-71. https://doi.org/10.4236/gsc.2021.112006

Pazmiño-Hernandez, M., Moreira, C. M., & Pullammanappallil, P. (2019). Feasibility assessment of waste banana peduncle as feedstock for biofuel production. *Biofuels, 10*(4), 473-484. https://doi.org/10.1080/17597269.2017.1323321

Santa-Maria, M., Ruiz-Colorado, A. A., Cruz, G., & Jeoh, T. (2013). Assessing the feasibility of biofuel production from lignocellulosic banana waste in rural agricultural communities in Peru and Colombia. *Bioenergy Research, 6*, 1000-1011. https://doi.org/10.1007/s12155-013-9333-4

Santis-Espinosa, L. F., López-Vidaña, E. C., Hernández-Altamirano, R., Grajeda, J. M. S., Aguilar-Aguilar, F. A., & Sebatian, P. J. (2025). Solar Drying and Bioethanol Production From Agro-Industrial Banana (musa paradisiaca l.) Residues. *Revista de Gestão Social e Ambiental*, *19*(1), e010709. https://doi.org/10.24857/rgsa.v19n1-023

Sarkar, N., Ghosh, S. K., Banerjee, S., & Aikat, K. (2012). Bioethanol production from agricultural wastes: An overview. *Renewable Energy, 37*(1), 19-27. https://doi.org/10.1016/j.renene.2011.06.045

Tenkolu, G. A., Kuffi, K. D., & Gindaba, G. T. (2024). Optimization of fermentation condition in bioethanol production from waste potato and product characterization. *Biomass Conversion and Biorefinery, 14*(4), 5205-5223. https://doi.org/10.1007/s13399-022-02974-4

Uljanah, D., Melwita, E., & Novia, N. (2024). Bioethanol Fermentation from Banana Pseudostems Waste (*Musa balbisiana*) Pretreated with Potassium Hydroxide Microwave. *Journal of Ecological Engineering*, *25*(9), 1–13. https://doi.org/10.12911/22998993/190691

Verardi, A., De Bari, I., Ricca, E., & Calabrò, V. (2012). Hydrolysis of lignocellulosic biomass: current status of processes and technologies and future perspectives. Bioethanol, Rijeka InTech, 95-122. https://doi.org/10.5772/23987

Vivekanandaraja, R., & Kapilan, R. (2021). Production and optimization of bioethanol from over ripen sour banana fruit wastes (Musa sapientum) using Saccharomyces cerevisiae. *Journal of dry zone agriculture, 7*(2), 53-76. https://doi.org/10.4038/jdza.v7i2.28

Wandscher Busanello, F., Sérgi Gomes, M. C., Paschoal, S. M., Rodrigues da Silva Baumgärtner, T., Barp, G., & Fiorentin-Ferrari, L. D. (2024). The influence of membrane separation technique in the biodiesel and bioethanol production process: a review. *Biofuels, 15*(7), 903-928. https://doi.org/10.1080/17597269.2023.2294593

Yaverino-Gutiérrez, M. A., Wong, A. Y. C. H., Ibarra-Muñoz, L. A., Chávez, A. C. F., Sosa-Martínez, J. D., Tagle-Pedroza, A. S., & Balagurusamy, N. (2024). Perspectives and progress in bioethanol processing and social economic impacts. *Sustainability, 16*(2), 608-638 https://doi.org/10.3390/su16020608

Zainab, I. S. G., Adiya, S. S., Adamu, M. A., Ibrahim, E. O., & Jamal, D. I. (2022). Comparative study of bioethanol produced from different agro-industrial biomass residues. Earthline Journal of *Chemical Sciences, 7*(2), 143-152. https://doi.org/10.34198/ejcs.7222.143152

Zulnazri, Z., Maulana, A., Muarif, A., Sylvia, N., & Dewi, R. (2023). Production of Bioethanol from Banana Peel (Musa paradisiaca l.) Using Sulfuric Acid Catalyst. *Jurnal Teknologi Kimia Unimal*, *12*(1), 97-111. https://doi.org/10.29103/jtku.v12i1.11654

J.Natn.Sci.Foundation Sri Lanka 2026 54 (1): 113 - 128
DOI: http://dx.doi.org/10.4038/jnsfsr.v54i1.12551

RESEARCH ARTICLE

Mathematical Modelling

Optimized feature selection approach for glaucoma detection using logistic-based chaotic whale optimization algorithm

AM Boopathi[1*], N Rayen[2], V Subha[3] and A Manonmani[4]
[1] *Department of Electrical and Electronics Engineering, Sethu Institute of Technology, Virudhunagar District, Tamilnadu, India.*
[2] *Department of Computer Science, St.Xavier's College, Tirunelveli, India.*
[3] *Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, India.*
[4] *Department of Electronics and Instrumentation Engineering, Saveetha Engineering College, Chennai, India.*

Submitted: 02 February 2025; Revised: 24 February 2026; Accepted: 06 March 2026

Abstract: Feature selection in image processing reduces data dimensionality by eliminating unnecessary features. Meta-heuristic algorithms inspired by natural phenomena help in accurate classification, mitigating the complexity found in medical diagnosis systems relating to features from a retinal dataset. This research aims to obtain a global optimal solution with a reduced feature set for glaucoma detection using the DRION-DB and DRISHTI-GS1 retinal image datasets and the YiWei Chen Retina Dataset with a decision tree (DT). This study employs a novel optimization approach termed logistic-based chaotic whale optimization algorithm (LCWOA) with an evolutionary search process for the exploration and exploitation phases of chaos theory. An efficient selection algorithm is proposed to improve the classification results of glaucoma images by selecting an optimal feature subset from extracted features using logistic map factor. The proposed method achieves better accuracy than the existing algorithms and state-of-the-art feature selector.

Keywords: Chaos theory, DT classification, feature extraction, LCWOA feature selection, whale optimization algorithm.

INTRODUCTION

A metaheuristics approach functions through nature-oriented (as in evolutionary and swarm optimization) and physics-inspired sources. Evolutionary and swarm-based source types are deployed in this research to classify a glaucoma image from a given retinal image feature set (Bhuvaneswari & Devi, 2022). The two sources reduce the problem behind optimization for different applications. Recently, the evolutionary approach performed well on a feature selection (FS) problem (Dhal et al., 2020). The classification of the selection method functions through two strategies, evaluation and search. The feature selection process has a range of common machine learning algorithms evaluating a subset of features (An et al., 2018).

The research attempted to improve classification accuracy by reducing redundant information using an enhanced feature selection approach. A subset of features may be selected by a wrapper approach that incorporates an evolutionary algorithm, which is a well-known mathematical approach that is applied to improve classification accuracy using chaos theory alongside the DT classifier. The proposed model employs a wrapper approach called logistic-based chaotic whale optimization algorithm (LCWOA), considering the four elements of population, selection, reproduction, and fitness function. The four elements are implemented differently by dividing the search process into two phases, exploration and exploitation (Mafarja & Mirjalili, 2017), and by setting parameters for chaos theory with the logistic map factor. This approach overcomes two major problems that are likely to surface, particularly the prevention of local optima entrapment and dimensionality reduction, by proceeding with a randomness requirement.

* Corresponding author (manivannaboopathi@sethu.ac.in; https://orcid.org/0000-0002-9996-6723)

Glaucoma (Abdullah et al., 2021) is a chronic disorder, a pathological condition that leads to vision loss. Glaucoma is diagnosed by extracting a feature from a set of 33 features using the gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRM), and gray-level difference matrix (GLDM) techniques (Salam et al., 2016). The proposed feature selection approach is implemented using the LCWOA and preceded, at the outset, by a comparative analysis with the genetic algorithm (GA), particle swarm optimization (PSO), and whale optimization algorithm (WOA) selection methods. The GA considers feature vectors taken individually to generate a population with random combinations. The key elements considered for this algorithm are the population, selection, reproduction, and fitness function. It generates a process flow and evaluates relevant features for successive generations to evaluate the optimum result (Snaselova & Zboril, 2015). The PSO, on the other hand, utilizes a multidimensional search space to create individual particles, complete with velocity and location. It updates the entire population by revising the position of each individual particle at successive iterations., an arrangement that is driven by a target function and an assembly of particles (Raja & Gangatharan, 2015). Another selection method, the WOA, follows the two phases of exploration and exploitation to produce a population at successive iterations. The mathematical model is formulated to optimize the search space using spiral bubble-net feeding behaviour. The study offers a global optimum solution, following a scrutiny of the shrinking, encircling, and spiral updating mechanisms (Sayed & Hassanien, 2018).

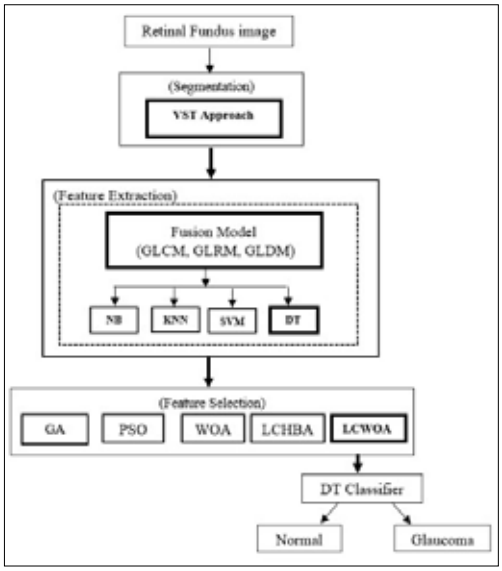


Figure 1: Framework for glaucoma detection

Feature selection works on the principle of supervised learning with a decision tree, which predicts normal or abnormal classes for the datasets. The results obtained are discussed in respect of an improved performance analysis. The methodology applied is presented in Figure 1, depicting the overall framework for glaucoma detection using the proposed feature selection approach, and incorporating the techniques of segmentation and feature extraction to ensure a reliable performance.

The main contributions of this work are as follows:

1. A logistic-based chaotic whale optimization algorithm (LCWOA) is proposed to obtain an optimal and compact feature subset for glaucoma detection.
2. The use of a logistic chaotic map improves the search behaviour and helps the model escape local optima.
3. The selected features reduce redundancy and enhance the performance of the Decision Tree classifier.
4. The effectiveness of the proposed method is validated on the DRION-DB, DRISHTI-GS1, and YiWei Chen retinal datasets and achieved better performance than other state-of-the-art techniques.

Literature review

Glaucoma detection has been analysed in regard to segmentation, feature extraction, feature selection, and classification. The following studies discuss the different techniques involved in each of these domains. Zou et al. developed a saliency-based segmentation technique to distinguish the optic disc (OD) region from a fundus image in the same location. The location of the OD is determined by the density of the vessel and matching templates. The OD region is exploited by boundary and connectivity measures for the segmentation, while the detection process is formulated as a saliency object (Zou et al., 2019). The texture of the segmented region of the fundus image is analysed by feature space techniques such as the GLCM, GLRM, and GLDM. Simonthomas et al. (2014) proposed an algorithm with Haralick features to detect glaucoma, applying a matrix with four directions and using the GLCM for the input image. The values of each direction are combined into a single matrix using a mathematical formulation. The resultant texture feature was computed from the Haralick-based technique to classify retinal images, using the Decision Tree (DT) classifier, into healthy or glaucomatous (Simonthomas et al., 2014).

(Zemmal et al., 2018) followed an automated process to detect glaucoma from the unlabelled data from the RIM-ONE retinal database. The proposed system used and improved upon a machine learning-based feature

selection algorithm constructed with a feature vector that comprises the GLCM as well as the Hu and central moments. The methodology designed produced an optimized feature set with high accuracy metrics using the wrapper approach, a transductive support vector machine (SVM), and a genetic algorithm. Padma and Sukanesh applied the wavelet transformation to extract dominant GLRM texture-based features from the input image. The proposed technique employed medical imaging and detected the affected region, using the SVM, to obtain 98% accuracy (Padma & Sukanesh, 2011). The feature space is used to improve classification accuracy by dividing the samples into two categories, normal and abnormal.

Kumar et al. surveyed the existing research to compare the accuracy of different machine learning algorithms, depending on the datasets and their attributes. The proposed method was evaluated using the Naive Bayes, decision tree, SVM, and KNN machine-learning techniques to predict disease based on the given dataset and it was observed that the SVM produced the best classification accuracy of all (Kumar et al., 2019). Much research has focused on producing an optimal solution for feature space dimensionality from the features extracted. It is possible, therefore, to select effective features from multidimensional feature spaces using a feature selection algorithm. Singh et al. followed a 2-system structure module with genetic algorithm procedures driven by feature selection and constructive induction. The proposed module's performance was explored with the ID3 to overcome the problems associated with texture classification. The constructive module improved performance to achieve better results than the existing system (Singh et al., 2022).

An et al. (2018) classified the glaucomatous OD into four types to quantify 91 parameters in terms of blood flow and ocular structure. The hybrid feature selection method is followed to find a possible feature for the neural network, Naïve Bayes, support vector machine, and gradient boosted decision tree classification models. It combines a feature selection method such as the genetic algorithm with the minimum redundancy. The performance was validated with 87.8% accuracy using nine parameters. Rajaji *et al.* implemented an artificial intelligence technique to classify the fundus image into two types using the fuzzy C-means centroid optimization method with features like the Cup-to-Disc Ratio (CDR) and the Neuroretinal Rim Ratio (NRR) in the inferior, superior, nasal, and temporal (ISNT) quadrants. This work compares the optimization techniques with the artificial fish swarm optimization (AFSO), Firefly optimization (FFO), and Particle Swarm Optimization (PSO), and the outcome was evaluated by metrics like accuracy, sensitivity, and specificity to obtain a rate of 84% (Rajaji & Prabakaran, 2019).

Bai presented an optimization technique based on a global heuristic method called particle swarm optimization (PSO). The research focused on optimizing the algorithm's performance with swarm intelligence. The rapid development of the algorithm is required for the implementation of a few particles that are widely used in optimization (Bai, 2010). Khalil *et al.* analysed glaucoma detection techniques with different datasets. The detection was conducted using Optical Coherence Tomography (OCT) on an automated prediction system that includes preprocessing, feature extraction, feature selection, and machine learning techniques. It was noted that the system achieved minimum accuracy (Khalil *et al.*, 2014). Somogyi evaluated the performance of machine learning models constructed by the SVM, KNN, and DT algorithms. The performance of each learning model hinges on measures that drive its overall performance. The proposed model evaluates performance in terms of the different metrics followed by the algorithms applied (Somogyi, 2021).

Li *et al.* introduced a weight factor for enhanced whale optimization performance using the chaos strategy with 13 benchmark functions. In this work, the exploration and exploitation phases are balanced with the corresponding population size and individual generations based on WOA variants. The weight factor is used to adjust the degree of influence between the phases to generate an optimal solution with greater accuracy and convergence levels. The implemented strategy is followed by a cosine function to manage population diversity (Li et al., 2019). Li et al. also produced a modified novel whale optimization-based algorithm using the two-tent chaos map and tournament selection strategies. This algorithm is an improvement over the standards of the existing whale algorithm, which is liable to diminish the optimal solution. It increases population diversity and the randomness of the tent map value during the population initialization. Thus, the proposed optimization algorithm offers an improved tournament selection strategy with high accuracy (Li et al., 2020).

The literature survey reveals that the existing frameworks for glaucoma detection may degrade the performance of classification schemes. The present research makes several key contributions. Firstly, texture features are extracted using the first and second order of statistical probability of the GLCM, GLRM, and GLDM

from segmented fundus images considered as salient objects and evaluated with a different classification algorithm. Secondly, the evolutionary performance is applied using a well-known mathematical approach with the logistic chaos theory. Thirdly, it focuses on arriving at an optimal solution for glaucoma detection in a retinal image dataset using a feature selection method with a DT classifier. Lastly, based on the wrapper model, the proposed feature selection approach is developed to maximize the accuracy rate by reducing redundant feature information from the dataset. The principal contribution of this study lies in its attempt to obtain an applicable accuracy rate by comparing the extraction results with those of machine learning algorithms such as the naive bayes (NB), K nearest neighbor (KNN), support vector machine (SVM), and decision tree (DT). Finally, a performance analysis concludes that the DT outperforms the other algorithms using the relevant metrics. A recent glaucoma detection approach combined DR2T-based feature extraction with Golden Jackal Optimization and LS-SVM classification, achieving high accuracy with fewer features on standard datasets (Sharma et al., 2025). A HyGKS-SAO–based hybrid feature selection approach improved classification performance on high-dimensional medical datasets by effectively balancing exploration and exploitation (Zaitoon et al., 2025).

The rest of the research paper is structured as follows: Section 3 deals with the proposed methodology, where segmentation, extraction, feature selection, and classification are discussed in detail. The results obtained, based on standard databases, are described in Section 4. Section 5 presents the concluding remarks.

MATERIALS AND METHODS

The methodology of the framework employed to detect glaucoma from fundus eye images is presented in the following stages.

Segmentation of ROI by visual saliency threshold (VST)

Segmentation is carried out by isolating the optic disc (OD) region with a specific threshold-based probability structure termed a saliency map (Gao *et al.*, 2019). The OD region is detected using the proposed approach VST along with the DT classifier. This approach is implemented to yield the finest result with a minimum computation time and attain hopeful results to identify an object region in OD from FEI. It decomposes the particular input images into condensed and varied elements in the following phases,

i. Brightening the region using histogram equalization
ii. Calculate distance of a pixel from the region using visual saliency descriptors
iii. Define the threshold value by means of the saliency method to isolate the region with the maximum pixel value using the saliency map, as follows:

$$Smap(x,y) = ||I\mu - IS(x,y)|| \quad ...(1)$$

where '$I\mu$' denotes the mean vector value of the preprocessed image, and '*IS*' matches Gaussian blurred version of the given image pixel vector value to the same input image pixel vector.

The resulting segmentation is refined using a saliency method that is suitable to detect a salient object with highlighted regions, as shown in Figure 2. An accurate segmentation of the object region is detected by the proposed novel approach VST in an efficient manner with maximum accuracy (S.Niraja P Rayen & Subha, 2022).

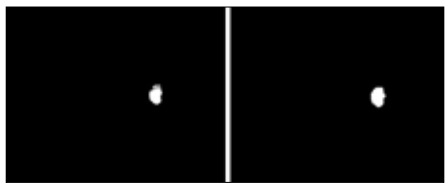


Figure 2: (a) Before saliency (b) After saliency

Feature extraction

Extraction techniques have been used to convert an image into a feature structure with texture and statistically oriented numerical values (Guo *et al.*, 2020). The proposed method explores feature extraction with a fusion model comprising the GLCM, GLRM, and GLDM feature extractors. The model extracts features of the segmented OD region from the glaucomatous image. In this research, the features extracted are used to enhance the efficacy of the feature set and help optimize it. The extraction process in this research requires a fusion model that is implemented with the following techniques:

i. GLCM: The GLCM extracts a statistical feature by supporting the co-occurrence value of a related pixel with a gray level. The GLCM probability is acquired by following the density function in a specified direction, as follows.

$$cpdf(i, j \mid dist, \theta) \quad ...(2)$$

where any pair of pixels (i,j) is located at a distance 'dist' with 'ɵ' gray level direction.

ii. GLRM: Given a specific direction, the GLRM extracts gray-level spatial dependencies in a row and column form. Throughout this process, higher-order statistical information is utilized to boost the intensity of a pixel with the two-dimensional form in a given direction. Therefore, the number of pixels is used to identify a process execution length in each direction.

iii. GLDM: This extracts an absolute gray level of any two pixels that have been separated by a precise probability of gray level using the density function. The Probability focuses on the feature's density value at a different angle, using the equation below to measure its new level for the given image.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} dpdf(i) \quad ...(3)$$

where $dpdf(i)$ is the total probability value of the pixel's gray level, and 'N' the number of image gray levels in the particular image pixel 'i'. Table 1 displays the feature attributes of each extraction technique.

Table 1: Feature attributes

Extraction techniques	Feature attributes
GLCM	Autocorrelation (AU), Contrast (Cont), Correlation (Corr), Cluster Prominence (CP), Cluster Shade (CS), Dissimilarity (Diss), Energy (E), Entropy (Entr), Homogeneity1 (H1), Homogeneity2 (H2), Maximum probability (MP), Sum of squares (Variance), Sum average (SA), Sum variance (SV), Sum entropy (SE), Difference variance (DV), Difference entropy (DE), Information measure of correlation1(IMC1), Information measure of correlation2 (IMC2), Inverse difference (ID), Inverse difference normalized (IDN), Inverse difference moment normalized (IDMN).
GLRM	Short Run Emphasis (SRE), Long Run Emphasis (LRE), Gray Level Non-uniformity (GLN), Run Percentage (RP), Run Length Non-uniformity (RLN), Low Gray Level Run Emphasis (LGRE), High Gray Level Run Emphasis (HGRE).
GLDM	Contrast (Con), Angular second moment (ASM), Entropy (E), and Mean (M) based Probability Density Function at a different angle

The significant features are extracted from the segmented objects using GLCM, GLRM, and GLDM techniques (Beura et al., 2015; Ahmadi & Akbarizadeh, 2020; Subha & Rayen, 2022). In all, 33 features are extracted separately for each normal and abnormal image, using the fusion model. Additionally, for the purpose of classifying the fundus images, the extracted features are trained by efficient classifiers and differentiate into two classes.

Feature selection

A selection technique is used to optimize the feature set from the features extracted. Feature selection plays a vital role in classification problems, helping achieve better accuracy and faster convergence. The number of features is reduced by eliminating irrelevant and redundant features for acceptable classification accuracy. There is a need, therefore, to convert the features selected to a more appropriate input form for the classifier to be able to detect glaucoma and non-glaucoma images. The proposed LCWOA optimizer is used in this research to select the best features from the dataset using the chaos-based whale optimization algorithm and to categorize the features in the retinal fundus images as well, applying DT classification. In this research, proposed optimizers called LCWOA were used to select the best features from the given dataset, using the chaotic-based whale optimization, and classify the features in retinal fundus images with DT classification (Abdel-Basset et al., 2021).

Whale optimization algorithm (WOA)

Sayed et al. (2018) presented the whale optimization algorithm (WOA) for optimizing numerical problems. The algorithm is inspired by the intelligent behaviour displayed by humpback whales during hunting, in particular the hunting strategy known as bubble-net feeding. It imitates the observed cooperative hunting activity of whales that commences with an encircling procedure. The encircling motion creates a spherical path that helps in the search for prey and, in due course,

shrinks appropriately to spiral upwards to the surface, where the feeding takes place. The shrinking encircling and spiral updating are key mechanisms that are integral to the working of the algorithm. The search position of the whales centres around a spiral at the surface of the water, made of bubbles in two directions, upwards and downwards, in a range encompassing around 12 meters. Figure 3(a) illustrates the exploration and exploitation phase, while Figure 3(b) shows the spiral updating mechanism.

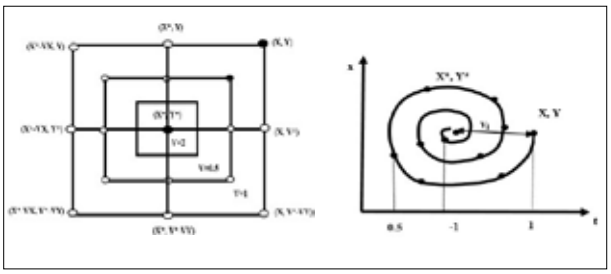


Figure 3: (a) Exploration and exploitation mechanism; (b) Spiral updating mechanism

The mechanism of the WOA algorithm includes the following three phases derived from the hunting behaviour of whales:

i. Encircling prey: In this phase, the position of the prey area of the whales in an unknown search space is determined, following which the best solution for the search agent is ascertained and updated.
ii. Bubble-net attacking: This phase defines the search process and minimizes it by means of random path selection using the two approaches of shrinking encircling and spiral updating.
iii. Searching for prey: The search position is limited by a specified random vector sized between 0 and 1. The global search is exposed with a random search agent, and the environment for each phase is set by a corresponding mathematical model.

Logistic-based chaotic whale optimization algorithm (LCWOA)

The challenges brought on optimization problems have been resolved satisfactorily by meta-heuristic algorithms. In the classification process, these algorithms overcome the problems associated with parameter optimization and feature selection based on nature-inspired methods. This research paper applies a chaos strategy to optimize classification accuracy using a decision tree (DT) for a parameter-related selection approach, based on the WOA for improved accuracy. The efficiency of the chaos is boosted by random initialization of population diversity and population distribution within the search space.

Chaos strategy: An optimization algorithm correlates a map function with the random behaviour of a parameter in a nature system, and mapping so produced is known as chaotic behaviour. Optimization applications have structures to deal with local optima problems by applying a nonlinear method termed chaos strategy. It requires an unbalanced system to arrive at a global optimum solution within a definite search range [0,1] using active features. Consequently, this structure comprises a map function with an 'n' dimensions dynamic model at a particular time theory equation as

$$cu_i^{(k+1)} = f\left(cu_i^{(k)}\right), i = 1,2,3,\dots,n \quad \dots(4)$$

The described function defines a state of chaos strategy in an ordered form $cu_i^{(k)}$, where k = 0, 1, 2. Each $cu_i^{(0)}$ is a system vector value at a particular interval with an ergodic property. This evolution system is developed to produce a search function to overcome the inefficiency of the trapping problems of local optima by the variants' chaotic map function, namely Chebyshev, Circle, Gauss/ mouse, Iterative, Logistic, Piecewise, Sine, Singer, Sinusoidal, and Tent.

The behaviour of each function is manipulated dynamically and incorporated as chaos-based optimization. Here, the logistic map improves the performance of the presented non-linear system for feature selection by adopting population initialization.

Integration of chaos with WOA

The WOA is utilized in the present study, given its effectiveness in solving optimization problems that lack structural reformation. When the chaos mechanism is combined with the WOA, the convergence speed gets enhanced and makes it possible to escape being trapped in a local optimal solution for large dimensional problems. The proposed method employs the chaotic map function above, along with the WOA, to investigate the optimization process by means of a chaos key factor called the logistic map. The results of the chaotic maps prove that a logistic map is needed to advance performance with greater accuracy.

In this research, the shrinking encircling and spiral updating WOA mechanisms are, in their respective phases, slated for a superior performance with the

application of the logistic chaos strategy, as follows:

Exploitation phase: This is meant for the encircling mechanism with maximum iterations and search agent position (s) along with best search agent direction (d). The mathematical formulation is described as:

If (s < 0.5 and mod(d) < 1) ...(5)

In the next iterations, the search agent position s(i+1) is given by the following equations as the candidate position for each iteration level

$$B = \left|U \cdot \overrightarrow{s^*}(i) - U.\vec{s}(i)\right| \quad ...(6)$$

$$\vec{s}(i+1) = \overrightarrow{s^*}(i) - V \cdot B \quad ...(7)$$

where s is the search agent position at each iteration with a constant value, s^* holds the optimal solution by selecting the best search agent and s(i+1) is the best search agent position at the current iteration in the current situation. The value for V and are U specified by the following equations

$$V = 2.\vec{x} \cdot y - \vec{x} \quad ...(8)$$

$$U = 2.y \quad ...(9)$$

where x ranges from 2 to 0 linearly and y holds chaotic map value with random selection. Then,

$$x = 2 - i \times \frac{2}{MX_{ilevel}} \quad ...(10)$$

Spiral phase: This phase maintains an updated position of whale and prey by the helix-based simulation during the exploration phase. The movement of a whale (new position) is expressed as following a spiral model,

$$\vec{s}(i+1) = \begin{cases} \overrightarrow{s^*}(i) - \vec{V} \cdot \vec{B} & \text{if } s < 0.5 \\ B \cdot e^{brl} \cdot \cos(2\pi rl) + \overrightarrow{s^*}(i) & \text{if } s \geq 0.5 \end{cases} \quad ...(11)$$

where B is the distance from the new generation's updated position to the new position, b holds a linear constant between 0 and 1 for the spiral move, and rl is the matrix form of an element with a random level in the interval [-1,1]. These combinations of the WOA strategy are applied to evaluate the feature selection process.

Exploration phase: Intended for the prey searching mechanism, the exploration phase works by selecting a random position for the search agent to update its position at each iteration. It avoids the local optima in the WOA, replacing it with the optimum solution as,

$$\begin{aligned} B &= U \cdot \overrightarrow{S}_{\text{rand}} - \overrightarrow{S} \\ \vec{S}_i(i+1) &= \overrightarrow{S}_{\text{rand}} - V \cdot B \end{aligned} \quad ...(12)$$

At each iteration i, the searching position is updated by the random selection vector (S^) ¯_"rand " ^ with two criteria, *i.e.*, V < 1 for position updating and V > 1 to select the best solution. This phase can be switched to one where a spiral position is updated or a phase terminated by applying the fitness function principle.

In this approach, each phase is evolved with a logistic map and corresponding parameters to evaluate the performance of optimization over GA, PSO, and WOA.

Logistic map: A logistic map can relate a polynomial mapping with 2 degrees of a nonlinear system to achieve chaotic behaviour by the expression,

$$x_n = \mu * x_{n-1} * (1 - x_{n-1}) \quad ...(13)$$

Here, μ used to measure the performance of the system between intervals (0, 4). If μ > 4, there is no time interval [0, 1] and if μ = 4, there is a possibility of chaotic occurrence (Li *et al.*, 2020).

The LCWOA implements wrapper mode feature selection with an optimal feature subset, which is initiated by a random solution set for a search agent to regulate the LCWOA algorithm in different size combinations. The search agent updates its position at each iteration based on a predefined fitness function. The detailed steps of the proposed LCWOA are presented in Algorithm 1, and the overall workflow is shown in Figure 4.

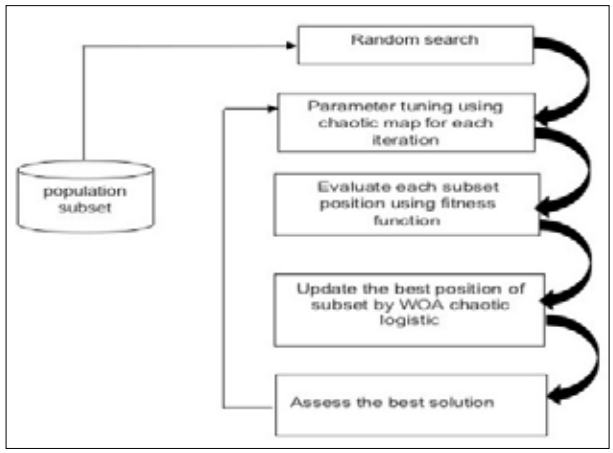


Figure 4: Flow of LCWOA

ALGORITHM 1: LOGISTIC-BASED CHAOTIC WHALE OPTIMIZATION ALGORITHM

1.Initialize a random population Xi (i = 1,2…n) with a generation count gt

2.Select the best search agent SG* by fitness evaluation of each search agent from generation

3.Assign chaotic map cmap0 values randomly

4.while (gt < max_iter)

5.Choose the respective chaotic number with logistic map

6.for each search agent in generation

7. Position of search agent is updated in a shrinking, encircling and spiral mechanism with random vector |A| using probability as follows

if (p< 0.5)

if(|A|< l)

Using Eq.7 search agent " is updated with current position

else if(|A|> 1)

i.Select a random search agent (SGrand) by Eq.12

ii. Choose best solution SG*

end if

else if(p>0.5)

Using Eq.11 search agent " is updated with current position

endif

8.end for

9. repeat step 7 until search agent is further selected from search space by

calculating the fitness of each search agent using Eq.14

10.return the best search agent by update SG* as solution

11.gt=gt+1

12.end while

13.return SG* as best solution

Classification

Typically, the performance of feature selection is influenced by the classification process to classify the glaucoma disease using the choice of the objective function. This function ensures the performance of the optimization problem expressed in terms of the classifier's error. It is conveyed with the selected feature and false positives to measure the behaviour of the classifier. In addition to that, the proposed optimization applies DT classification to provide an optimal solution by finding the best feature with the objective function

$$FF_t = \max\left(CAcc + w_f * \left(1 - \frac{L_{sf}}{L_{tf}}\right)\right) \quad ...(14)$$

where CAcc is the total accurate measure of classification calculated from a number of labelled samples with the total number of samples, w_f denotes labelled instance with [0, 1] weight factor value, L_{sf} is the selected feature-set length and L_{tf} is the total feature-set length.

Decision tree (DT)

The DT classifier generates a tree-structured classifier along with two nodes called 'decision nodes' and 'leaf nodes', in which internal nodes represent a feature set of data, a branch holds a decision rule and each leaf node represents the output (Berenguer-Vidal *et al.*, 2022). On the basis of the feature of a given dataset, the test is performed. Each decision node tested 'X' element of the input data and features of branches; then each branch holds a result of 'X'. Every leaf node signifies a group to making an effective decision.

A fundamental method of building a DT is a split procedure by dividing a feature space into a number of sub-classes (Holmes *et al.*, 2002). A training feature set 'T' can hold data of 'k' classes (c_1, c_2…c_k) or can only contain a single class. If 'T' restrains no data, it is assigned as a leaf node and its connected class is going to act as the main class of its root node. If 'T' holds the cases of different classes, a test holding some attribute parameter 'ai' of the training data is accepted and can be split into 'n' subsets (T_1, T_2,…..,T_n), where 'n' is the number of results of the test over attribute 'a_i'. A corresponding process of constructing 'DT' is recursively executed over every T_i, where T_i ranges from 1 to n, till every subset fits into one class.

The algorithm starts the whole process of classification at a root node of the tree. The root divides the feature space into more sub-classes of the feature set. The classes are assigned by weights. Based on features weights are computed during the learning process and these weights are used to classify testing (unknown) data.

RESULTS AND DISCUSSION

Experiment setup

The experiments that were carried out have attempted to analyze the feature selection method and classification algorithm applied using MATLAB. The experiments were evaluated to validate the performance of the proposed approach in terms of classifying the affected images.

Dataset

In this research, three glaucoma databases, namely, the DRION-DB (Medical Image Processing Group), DRISHTI-GS1 (Carmona *et al.*, 2008), and YiWei Chen Retina_dataset (Suzhou Institute of Biomedical Engineering and Technology) were used. These datasets contain 110, 101, and 101 fundus eye images (FEIs), respectively, including both normal and abnormal cases. Figure 5 (a), (b), and (c) shows examples of images from each of the three databases and their respective OD masks. An example of a database image and its OD marking is demonstrated in Figure 5. Thus, the proposed method is implemented to classify a glaucoma image using an optimized feature set selected from the datasets, and the experimental evaluation has taken 80% for training and 20% for testing.

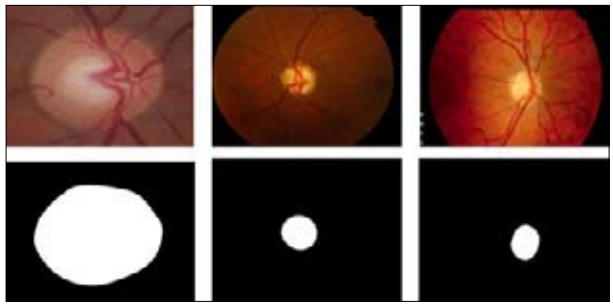


Figure 5: Sample images from respective database

Performance measures

The proposed method is evaluated using performance metrics namely recall rate, specificity, precision rate, F1-score, positive predictive value (PosPV), negative predictive value (NegPV), false positive rate (FPosR), prevalence (PVal), and accuracy.

The proposed framework has been validated through feature extraction techniques using the GLCM, GLRM, GLDM and fusion model. The fusion model has been modelled on a combination of the GLCM, GLRM, and GLDM. The constructed model produced accurate feature sets and the extracted features were classified using supervised algorithms such as the NB, KNN, SVM, and DT. From the classified results, the proposed feature selection algorithm was executed to produce an optimal feature with LCWOA. The experimental results of the study are discussed below.

Table 2: Overall Accuracy of Feature Extraction Techniques

Dataset	Feature Extraction Technique	NB	KNN	SVM	DT
DRISHTI-GS1	GLCM	74.5	67.7	77.7	80.5
	GLRM	74.5	70.9	77.7	78.9
	GLDM	78.0	80.2	69.5	89.3
	Proposed Fusion Model	80.6	83.0	81.4	94.0
DRION-DB	GLCM	72.7	68.3	75.4	79.8
	GLRM	73.3	70.2	75.7	76.3
	GLDM	76.9	79.6	70.1	78.4
	Proposed Fusion Model	80.0	76.0	82.0	88.0
YiWei Chen retinal dataset	GLCM	73.8	69.4	76.3	80.2
	GLRM	72.5	71.8	76.7	79.7
	GLDM	77.0	80.3	68.5	89.3
	Proposed Fusion Model	80.4	78.5	86.0	90.0

Analysis of proposed framework with feature extraction techniques

Experimental analysis #1:

The given dataset was used to analyze the performance of the classifier-based feature extraction model and the performance values are produced in Table 2.

It is observed from the analysis above that each extraction technique was executed with the corresponding feature attributes, presented in Table 1, on three different datasets. The extracted features were classified by the four-classification algorithm to categorize the results into normal and abnormal fundus images. Table 2 shows that the proposed fusion feature extraction model offers higher accuracy with 33 features under the DT classifier, producing better results than those of existing feature extraction techniques.

Experimental analysis #2

For the purpose of evaluation, 80% of the data were used for training and 20% for testing. In order to test and assess the proposed model, a total of 312 glaucoma images were taken for classification. The performance analysis above makes it plain that the proposed feature extraction fusion model yields better accuracy, comparatively speaking. Further, the performance of the proposed feature extraction model with the given datasets has been analyzed, based on the metrics mentioned earlier, and the results shown in Table 3.

Table 3: Performance comparison of classifiers using various evolution parameters

	Classifiers			
Metrics	NB	KNN	SVM	DT
Recall	60	70.1	0	87.7
Specificity	63.6	89.4	69.5	97.0
Precision	83.6	77	0	93.4
F1-Score	69.8	73.4	0	90.5
PosPV	83.6	77.0	0	93.4
NegPV	75.5	85.6	93.5	94.2
FPosR	0.086	0.105	0.305	0.029
Pval	0.425	0.335	0	0.325

The experiment has evaluated the contribution of the classifiers in regard to the performance metrics to examine the number of true positives, false positives, true negatives, and false negatives in glaucoma detection from the dataset. From Table III, the results presented show that the DT classifier produces better results in the proposed feature extraction model than the remaining classifiers. Thus, the experiments conducted have been evaluated, based on a combination of features, to classify normal and abnormal cases in a dataset.

Performance evaluation of LCWOA

This section compares the proposed LCWOA with other feature selection methods, namely GA, PSO, and WOA. The following table shows the experimental results of the selection algorithm from the given datasets. The classification process recognizes glaucoma by splitting the cases into two classes with the training and testing features.

Parameters initialization

Table 4 presents a parameter initialization that experiments with the feature selection method LCWOA. For a fair comparison, the same experimental conditions were maintained for GA, PSO, and WOA. By empirically evaluating the LCWOA, the parameters were chosen to make a trade-off between performance and computational cost. When the population size was 20 and the number of iterations was 60, the method was able to converge stably and the classification accuracy was very consistent. Hence, the same values were used in all the experiments.

Table 4: Initialization of parameter values

Parameter	Values for LCWOA
Population Size (Search Agents)	20
Iterations Level (Generation)	60
LB	1
UB	33
Dimension	33

Performance metrics for proposed Feature Selection approach

The following metrics are used to analyze the proposed LCWOA performance. The mathematical formulation of each metric is defined as,

$$Standard\ Deviation\ (SD) = \sqrt{\frac{\sum_{i=1}^{MI}(p_i - \mu)^2}{Max}}$$

$$\text{Meanfit (Mean fitness)} = \frac{1}{MI}\sum_{i=1}^{MI} P_i$$

$$\text{Bestfit (Best fitness)} = \max_{i=1}^{MI} P_i$$

$$\text{Worstfit (Worst fitness)} = \min_{i=1}^{MI} P_i$$

$$\text{ASS (Average feature selection size)} = \frac{1}{\text{MI}}\sum_{\text{i=1}}^{\text{MI}} \frac{\text{length } (\text{P}_\text{i})}{\text{N}}$$

where MI is the iteration's maximum level, P is the obtained best level at each iteration, and N is the total features of the dataset.

The extent of the random population does not ensure uniform distribution for the search space. An underlying condition of the WOA is that it culls the population in terms of size at each iteration. Owing to population diversity, WOA efficiency is too low to reduce the convergence speed in order to keep pace with the quality of the local optimal solution available. To overcome a population diversity problem, the proposed feature selection measures an optimal solution to detect glaucoma disease using the mentioned parameter values and metrics. Chaotic map values were applied to find a suitable map using those metrics. As in Table 5, the logistic-based map function provides the best fitness outcome for the given fundus image dataset.

Table 5: Finding a suitable chaotic map (cmap0) for population with Bestfit

Map function	SD	Meanfit	Bestfit	Worstfit	ASS
Chebyshev	0.05	1.45	1.62	1.21	0.16
Circle	0.07	1.42	1.67	1.24	0.18
Gauss/mouse	0.04	1.48	1.59	1.38	0.33
Iterative	0.06	1.40	1.63	1.26	0.21
Logistic	0.04	1.49	1.51	1.29	0.20
Piecewise	0.-08	1.43	1.62	1.23	0.17
Sine	0.11	1.39	1.60	1.32	0.20
Singer	0.06	1.32	1.58	1.39	0.18
Sinusoidal	0.09	1.35	1.60	1.32	0.17
Tent	0.03	1.49	1.54	1.41	0.17

The classification accuracy is maximized by the optimized features to assess constancy by utilizing best, worst, and fitness metrics of the chaotic map function. The extracted features from the fusion feature extraction model were minimized with a logistic-based WOA algorithm.

The aim of chaos theory is to produce an optimal solution from the random population of a subset of

features. The proposed LCWOA was implemented with a focus on the best optimal subset with reduced convergence speed. Figure 6 shows the convergence curves of the extracted feature set for all the chaotic maps. Furthermore, the statistical analysis reveals superior performance with the selection of a minimum feature level using the Bestfit of the logistic map.

Figure 6: Convergence curve for different chaotic map

Experimental analysis # 1

Table 6 has focused on the performance of several feature optimization techniques for the classification of glaucoma, calculated the accuracy of the proposed algorithm, and compared it with existing techniques such as the GA, PSO, WOA and LCHBA. For an enhanced performance, classification accuracy was fine-tuned by the DT classifier with ten chaotic maps. The benchmark datasets have been used to compare the performance of the LCWOA, particularly in regard to enhancing exploration and exploitation in terms of the classification performance and number of selected features. In addition, the performance of the proposed approach has been evaluated with the original WOA and the quality of the solutions enhanced. Table 6, show that the LCWOA produces higher accuracy with 9 features under the DT classifier, using the DRISHTI-GS1 dataset, and outperforms the existing feature selection algorithms.

Table 6: Performance comparison of proposed feature selection algorithm with existing methods under DT Classifier

Dataset	Feature selection method	Accuracy (%)	No. of selected features
DRISHTI-GS1	GA	96.0	12
	PSO	88.7	15
	WOA	93.3	13
	LCWOA	97.8	9
	LCHBA	95.5	11
DRION-DB	GA	94.7	16
	PSO	86.4	14
	WOA	92.1	12
	LCWOA	96.7	11
	LCHBA	94.2	13
YiWei Chen retinal dataset	GA	95.6	15
	PSO	88.9	13
	WOA	93.1	12
	LCWOA	97.0	10
	LCHBA	95.5	12

Experimental analysis #2

The proposed framework using the chaotic map-based whale optimization algorithm performed better than the rest in terms of sensitivity, precision and F1-score. The proposed framework, using feature extraction, has been tested with the DRISTI-GS1, DRION DB, and YiWei Chen Retina_ dataset using the DT. Table 8 shows a performance evaluation of the proposed feature selection framework in terms of normal and abnormal classification with the DT. The experiment has evaluated the contribution of the DT classifier in glaucoma detection from the dataset. The LCWOA is analysed, based on the performance metrics with the dataset, and the values are produced in Table 7.

Table 7: Performance evaluation of feature selection using various evolution parameters with DT classifier

Metrics	GA-DT	PSO-DT	WOA-DT	LCWOA-DT	LCHBA-DT
Recall (%)	88.8	64.9	66.2	95.3	94.8
Specificity (%)	90.9	88.6	98.1	98.5	97.3
Precision (%)	83.3	78.1	96.0	96.7	94.4
F1-Score (%)	85.9	70.8	78.0	94.0	90.6
PPV (%)	83.3	78.1	96.7	96.9	95.6
NPV (%)	93.7	80.1	78.4	97.8	96.0
FPR	0.107	0.113	0.018	0.014	0.109
p-value	0.340	0.385	0.445	0.325	0.363

Table 7 results show that the DT classifier produces better results in the proposed feature selection approach than the existing algorithms. The tables above have shown a performance evaluation of the existing selection algorithm with the proposed approach. It is observed, in conclusion, that the proposed feature selection scheme outperforms the existing ones with different measures.

Analysis of proposed glaucoma detection framework with existing methods

A comparative study has been analysed with the proposed framework and existing methods as shown in Table 8. The performance is analyzed with respect to the classifiers and features for both classes, normal and abnormal.

Table 8: Comparative analysis with existing methods

References	Feature selection techniques	Classifier	Dataset	Acc %
(Ali *et al.*, 2022)	ReliefF	Deep Belief Network (DBN) optimized by Elephant-Herding Optimization (EHO)	Public & Private Retinal Fundus Datasets (7280 images)	99.4%
(Singh *et al.*, 2023)	PSO, ABC, BCS, PSO-ABC (Soft Computing Feature Selection Algorithms)	RF, SVM, Ensemble of RF, SVM, and Logistic Regression	ORIGA, REFUGE, and their combination	94.1%
(Singh *et al.*, 2024a)	GSOA	SVM, DT, RF, k-NN, Naïve Bayes, XGBoost.	Private+ORIGA	95.36 %
(Singh *et al.*, 2024b)	GWO, WOA, and hybrid hGWWO	SVM, DT,RF, k-NN, Naïve Bayes, XGBoost.	ORIGA	96.8%
Proposed LCWOA	GLCM, GLRM, GLDM	DT	312	97.8%

The proposed framework manages to deliver some promising results, but the dataset used is quite limited in size, which may raise the issue of overfitting when dealing with a multi-stage optimization scenario. This issue is somewhat mitigated through the feature selection process that decreases redundancy and model complexity. Besides this, the method was tested on three separate retinal datasets, and the uniform performance of the method on these datasets testifies to its robustness. In future work, the model will be validated on larger and more diverse datasets, and cross-dataset testing and data augmentation will be explored to further assess its generalization capability.

Figure 7 provides a thorough examination of the solution space in terms of the chaotic parameter and execution time. This study has offered a comprehensive statistical analysis of the performance of various algorithms across a number of datasets. The integration of chaotic maps with the Whale Optimization Algorithm (WOA) is justified by their potential to enhance the exploration and exploitation capabilities of the optimization process.

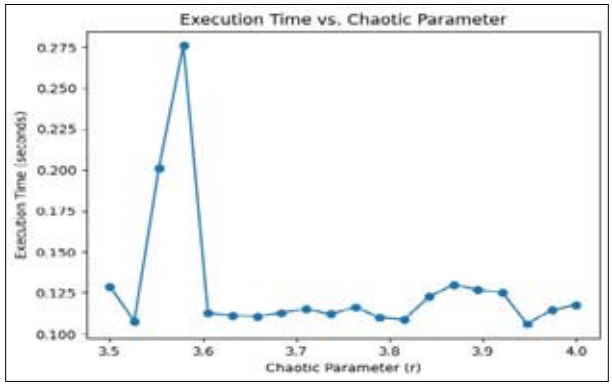


Figure 7: Computational Estimation

CONCLUSION

The present study has developed a computational framework for glaucoma disease classification using a fundus image database which is essential for its early detection. The proposed method presents an effective

approach to isolate the OD region with a specific threshold-based probability structure called a saliency map. Further, it ensures reliable texture-based feature extraction to detect glaucoma using both the statistical order and probability techniques like the GLCM, GLRM, and GLDM. It incorporates four classifiers to reduce feature dimensionality using a mathematical approach with chaos theory. The proposed LCWOA algorithm selects 9 optimized feature sets with an improved accuracy of 97.8% and yielded higher accuracy with the DT than the existing algorithm in both classes of the classification results. In future work, the proposed model will be validated on larger and more diverse retinal datasets to further examine its generalization capability and will be extended to other ophthalmic diseases. Hybrid feature selection strategies that combine filter and wrapper mechanisms will also be explored to further reduce false positive and false negative rates.

REFERENCES

Abdullah, F., Imtiaz, R., Madni, H. A., Khan, H. A., Khan, T. M., Khan, M. A., & Naqvi, S. S. (2021). A review on glaucoma disease detection using computerized techniques. *IEEE Access, 9*, 37311–37333. https://doi.org/10.1109/ACCESS.2021.3061451

Abdel-Basset, M., Chang, V., & Mohamed, R. (2021). A novel equilibrium optimization algorithm for multi-thresholding image segmentation problems. *Neural Computing and Applications, 33*(17), 10685–10718. https://doi.org/10.1007/s00521-020-04820-y

Ahmadi, N., & Akbarizadeh, G. (2020). Iris tissue recognition based on GLDM feature extraction and hybrid MLPNN-ICA classifier. *Neural Computing and Applications, 32*(7), 2267–2281. https://doi.org/10.1007/s00521-018-3754-0

Ali, M. A., Balasubramanian, K., Krishnamoorthy, G. D., Muthusamy, S., Pandiyan, S., Panchal, H., & Abd Elminaam, D. S. (2022). Classification of glaucoma based on elephant-herding optimization algorithm and deep belief network. *Electronics, 11*(11), 1763.

An, G., Omodaka, K., Tsuda, S., Shiga, Y., Takada, N., Kikawa, T., Nakazawa, T., Yokota, H., & Akiba, M. (2018). Comparison of machine-learning classification models for glaucoma management. *Journal of Healthcare Engineering, 2018*, 1–8. https://doi.org/10.1155/2018/6874765

Bai, Q. (2010). Analysis of particle swarm optimization algorithm. *Computer and Information Science, 3*(1), 180–184. https://doi.org/10.5539/cis.v3n1p180

Beura, S., Majhi, B., & Dash, R. (2015). Mammogram classification using two-dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer. *Neurocomputing, 154*, 1–14. https://doi.org/10.1016/j.neucom.2014.12.032

Berenguer-Vidal, R., Verdu-Monedero, R., Morales-Sanchez, J., Selles-Navarro, I., Kovalyk, O., & Sancho-Gomez, J. L. (2022). Decision trees for glaucoma screening based on the asymmetry of the retinal nerve fiber layer in optical coherence tomography. *Sensors, 22*, 4842. https://doi.org/10.3390/s22134842

Bhuvaneswari, A., & Devi, T. M. (2022). Contrast enhancement of retinal images using green plan masking and whale optimization algorithm. *Wireless Personal Communications, 125*, 1047–1073. https://doi.org/10.1007/s11277-022-09586-1

Carmona E.J. M, Rincon J, García-Feijoo and J. M. Martínez-de-laCasa. (2008). Identification of the optic nerve head with genetic algorithms. *Artificial Intelligence in Medicine*, 43, 243–259. https://doi.org/10.1016/j.artmed.2008.04.005

Dhal, K. G., Das, A., Ray, S., Galvez, J., & Das, S. (2020). Nature-inspired optimization algorithms and their application in multi-thresholding image segmentation. *Archives of Computational Methods in Engineering, 27*(3), 855–888. https://doi.org/10.1007/s11831-019-09334-y

Gao, Y., Yu, X., Wu, C., Zhou, W., Wang, X., & Zhuang, Y. (2019). Accurate optic disc and cup segmentation from retinal images using a multi-feature based approach for glaucoma assessment. *Symmetry, 11*, 1267. https://doi.org/10.3390/sym11101267

Guo, F., Li, W., Tang, J., Zou, B., & Fan, Z. (2020). Automated glaucoma screening method based on image segmentation and feature extraction. *Medical & Biological Engineering & Computing, 58*(10), 2567–2586. https://doi.org/10.1007/s11517-020-02237-2

Holmes, G., Pfahringer, B., Kirkby, R., Frank, E., & Hall, M. (2002). Multiclass alternating decision trees. In *European Conference on Machine Learning* (pp. 161–172). Springer. https://doi.org/10.1007/3-540-36755-1_14

Khalil,T.S. Khalid and A. M. Syed. (2014). Review of Machine Learning techniques for glaucoma detection and prediction. *Science and Information Conference,* 438-442. DOI: https://doi.org/10.1109/SAI.2014.6918224.

Kumar, A., Sushil, R., & Tiwari, A. K. (2019). Machine learning based approaches for cancer prediction: A survey. In *Proceedings of the 2nd International Conference on Advanced Computing and Software Engineering* (pp. 325–330).

Li, Y., Han, T., Han, B., Zhao, H., & Wei, Z. (2019). Whale optimization algorithm with chaos strategy and weight factor. *Journal of Physics: Conference Series, 1213*(3), 032004. https://doi.org/10.1088/1742-6596/1213/3/032004

Li, Y., Han, M., & Guo, Q. (2020). Modified whale optimization algorithm based on tent chaotic mapping and its application in structural optimization. *KSCE Journal of Civil Engineering, 24*, 3703–3713. https://doi.org/10.1007/s12205-020-0504-5

Ma, J., Yu, S., & Cheng, W. (2022). Composite fault diagnosis of rolling bearing based on chaotic honey badger algorithm optimizing VMD and ELM. *Machines, 10*(6), 469. https://doi.org/10.3390/machines10060469

Majdi M. Mafarja, SeyedaliMirjalili. (2017). Hybrid Whale Optimization Algorithm with simulated annealing for feature selection. *Neuro Computing*, *260*, 302-312. https://doi.org/10.1016/j.neucom.2017.04.053

Padma A and Sukanesh R. (2011). Automatic classification and segmentation of brain tumor in CT images using optimal dominant gray level run length texture features. *International Journal of Advanced Computer Science and Applications*, *2*(10), 53-59. https://doi.org/10.14569/IJACSA.2011.021009

Raja, Chandrasekaran, and Narayanan Gangatharan. (2015). A hybrid swarm algorithm for optimizing glaucoma diagnosis. *Computers in biology and medicine*, *63*, 196-207. https://doi.org/10.1016/j.compbiomed.2015.05.018

Rajaji, Shanthi and Prabakaran, S. (2019). Effective Utilization of Artificial Intelligence for Predicting Glaucoma in Fundus Image. *Caribbean Journal of Science*, *53*(1), 902-912

Rayen, S. N. P. & Subha V. (2022). A Novel Visual saliency Thresholding Method for Accurate ROI Segmentation in Optic Disc from FEIs for Glaucoma detection. *International Journal of Computing and Digital Systems*, 11, 477-489. https://dx.doi.org/10.12785/ijcds/110139

Salam, A. A., Khalil, T., & Akram, M. U. (2016). Automated detection of glaucoma using structural and non-structural features. *SpringerPlus, 5*, 1519. https://doi.org/10.1186/s40064-016-3175-4

Sayed, G. I., Darwish, A., & Hassanien, A. E. (2018). A new chaotic whale optimization algorithm for features selection. *Journal of Classification, 35*(2), 300–344. https://doi.org/10.1007/s00357-018-9261-2

Simonthomas. S, Thulasi. N, Asharaf. P. (2014). Automated diagnosis of glaucoma using Haralick texture features. *International Conference on Information Communication and Embedded Systems (ICICES* 2014) *IEEE*, 1-6. https://doi.org/10.1109/ICICES.2014.7033743

Singh, L. K., Khanna, M., Thawkar, S., & Singh, R. (2024a). A novel hybridized feature selection strategy for the effective prediction of glaucoma in retinal fundus images. *Multimedia Tools and Applications, 83*(15), 46087–46159.

Singh, L. K., Khanna, M., Garg, H., & Singh, R. (2024b). Efficient feature selection based novel clinical decision support system for glaucoma prediction from retinal fundus images. *Medical Engineering & Physics, 123*, 104077.

Singh, L. K., Khanna, M., Thawkar, S., & Singh, R. (2023). Nature-inspired computing and machine learning based classification approach for glaucoma in retinal fundus images. *Multimedia Tools and Applications, 82*(27), 42851–42899.

Singh, L.K., Pooja, Garg, H. (2022). An IoT based predictive modeling for Glaucoma detection in optical coherence tomography images using hybrid genetic algorithm. *Multimedia Tools and Applications*, *81*, 37203–37242. https://doi.org/10.1007/s11042-022-13540-5

Subha, V. & Rayen, S. N. P. (2022). Automated Glaucoma Detection from Fundus Eye Images Using Grey Level based Feature Extraction Methods and Supervised Learning Classification. *Turkish Online Journal of Qualitative Inquiry*, *12*, 2987-3008. https://www.tojqi.net/index.php/journal/article/view/6863/4892.

Somogyi Z. (2021). Performance Evaluation of Machine Learning Models. *In The Application of Artificial Intelligence, Springer*, 87-112. https://doi.org/10.1007/978-3-030-60032-7_3

Şüyun, S. B., Yurdakul, M., Taşdemir, Ş., & Biliş, S. (2025). Triple-stream deep feature selection with metaheuristic optimization and machine learning for multi-stage hypertensive retinopathy diagnosis. *Applied Sciences, 15*(12), 6485.

Sharma, S. K., Muduli, D., Rath, A., Dash, S., Panda, G., Shankar, A., & Dobhal, D. C. (2025). Discrete ripplet-II transform feature extraction and metaheuristic-optimized feature selection for enhanced glaucoma detection in fundus images using least square-support vector machine. *Multimedia Tools and Applications, 84*(23), 26447–26479.

Snaselova, P. and Zboril, F. (2015). Genetic algorithm using theory of chaos. *Procedia computer science*, 51, 316-325. https://doi.org/10.1016/j.procs.2015.05.248

Zaitoon, R., Begum, S. S. A., Mohanty, S. N., & Jose, D. (2025). Feature selection using hybridized Genghis Khan shark with snow ablation optimization technique for multi-disease prognosis. *Intelligence-Based Medicine, 11*, 100249.

Zemmal, Nawel, Nabiha Azizi, Mokhtar Sellami, Djamel Zenakhra, Soraya Cheriguene, Nilanjan Dey, and Amira S. Ashour. (2018). Robust feature selection algorithm based on transductive SVM wrapper and genetic algorithm: application on computer-aided glaucoma classification. *International Journal of Intelligent Systems Technologies and Applications*, *17*(3), 310-346. https://doi.org/10.1504/IJISTA.2018.094018

Zou B., Liu Q., Yue K., Chen Z., Chen J. and Zhao G. (2019). Saliency-based segmentation of optic disc in retinal images. *Chinese Journal of Electronics*, *28*(1), 71-75. https://doi.org/10.1049/cje.2017.12.007

J.Natn.Sci.Foundation Sri Lanka 2026 54 (1): 129- 142
DOI: http://dx.doi.org/10.4038/jnsfsr.v54i1.12284

RESEARCH ARTICLE

Imaging Technology

Detection of sugar adulteration in black tea using multispectral imaging

WAND Wickramasinghe[1], G. Thilakarathne[2*], EMSLB Ekanayake[3], AD Wijesinghe[1], KAST Senarath[1], HMVR Herath[1], GMRI Godaliyadda[1], MPB Ekanayake[1], T Madhujith[4] and KMM Mohotti[5]

[1] *Department of Electrical and Electronic Engineering, Faculty of Engineering, University of Peradeniya, Peradeniya, Sri Lanka.*
[2] *Department of Electrical and Information Engineering, Faculty of Engineering, University of Ruhuna, Hapugala, Galle, Sri Lanka.*
[3] *School of Engineering and Technology, Sri Lanka Technological Campus, Meepe, Padukka, Sri Lanka.*
[4] *Department of Food Science and Technology, University of Peradeniya, Sri Lanka.*
[5] *Tea Research Institute, Thalawakelle, Sri Lanka.*

Submitted: 11 July 2024; Revised: 01 October 2025; Accepted: 14 March 2026

Abstract: Black tea, valued globally for its flavor, aroma, and nutritional benefits, is a major export commodity for countries such as China, India, and Sri Lanka. Rising demand has led to sugar adulteration to enhance color, twist, and weight, compromising quality and posing health risks, highlighting the need for rapid and reliable verification methods. This study presents a multispectral imaging (MSI) based approach for detecting and quantifying sugar adulteration in black tea. A custom-built system with thirteen narrow band LEDs (365 nm to 940 nm) sequentially illuminated powdered and brewed samples, capturing 26 spectral images in reflectance and transmittance modes, respectively. Spectral features corresponding to sugar induced color changes were extracted independently of natural tea variability. Preprocessing steps, including dark current subtraction, cropping, histogram equalization, and dimensionality reduction via linear discriminant analysis (LDA), ensured high quality data for analysis. Classification was performed using linear discriminant analysis (LDA), K-nearest neighbors (K-NN), support vector machine (SVM), feed forward neural network (FFNN), and convolutional neural network (CNN), achieving accuracies above 93% across the tested models. These methods showed high sensitivity in detecting adulteration at levels as low as 5% (w/w) and strong specificity in distinguishing pure from adulterated brewed samples. Polynomial regression was applied to quantify sugar content, yielding R^2 values above 0.97 for polynomial orders from the first to the fifth. A third order polynomial was selected as it provided a slightly improved fit (R^2 = 0.9739) while maintaining low model complexity. These results demonstrate that multispectral imaging combined with machine learning enables reliable detection of sugar adulteration and continuous estimation of adulteration levels between 5% and 25%, supporting rapid and non-destructive monitoring of black tea authenticity.

Keywords: Black tea, food quality assessment, machine learning, multispectral imaging, sugar adulteration of tea.

INTRODUCTION

Tea, or "cha," is derived from the leaves of *Camellia sinensis* and related varieties. Originally discovered in southwestern China, its global dissemination was particularly driven by the British, alongside other Western powers, and facilitated via the Silk Road (Meng et al., 2019). As the world's second most consumed beverage after water (Cabrera et al., 2006), tea production encompasses post-fermented (black), fully fermented (oolong), and unfermented (green or pu-erh) categories (Fraser et al., 2014). This study focuses on black tea, particularly popular in Asia, the European Union, and the Middle East (XU et al., 2022).

With annual global production reaching approximately 6.5 billion tons (Markets and Trade Division – Economic and Social Development Stream, 2022), tea represents a major source of foreign exchange. Leading producers

* Corresponding author (gayathrit@eie.ruh.ac.lk; https://orcid.org/0009-0004-0306-2116)

China, India, and Kenya derive USD 2.08 billion, USD 750.63 million, and USD 109 million respectively (IBEF, 2022; Kamer, 2022; Ridder, 2023), while Sri Lankan tea exports contribute 2% to the GDP, generating USD 1.3 billion annually (Nadeera, 2022).

Tea is among the most widely consumed beverages worldwide, necessitating stringent quality standards established by national and international bodies (Chen et al., 2006). Quality assessment typically evaluates caffeine, water extract, total polyphenols, and free amino acids, ensuring consistency, safety, and market acceptance (Polat et al., 2022). Tea's bioactive compounds, including polyphenols, amino acids, minerals, caffeine, and catechins, shape its flavor, aroma, and astringency while also providing potential health benefits in antioxidative activity and diabetes prevention (Meng et al., 2019; Samanta, 2020).

Pure black tea contains up to 6.5% natural sugars (fructose, glucose, sucrose, m-inositol, maltose, and raffinose), while higher levels typically indicate external addition (Engelhardt, 2010, pp. 999–1032). Adulteration with sugar or glucose syrup becomes visually detectable above 20% concentration (Luqing et al., 2015), representing the realistic adulteration range targeted for selective detection.

Black tea production involves harvesting, withering, rolling, fermenting, drying, and grading, all requiring precise control (Aaqil et al., 2023). Quality declines often occur due to sugar addition in low grade tea to improve appearance (Wedagedara et al., 2019; Wang et al., 2021), which promotes bacterial growth, moisture absorption (Luqing et al., 2015), and health risks such as obesity, cardiovascular disease, and dental caries. Therefore, accurate discrimination between pure and adulterated tea is essential.

Figure 1: The key phases of black tea production

Traditional detection methods such as, Benedict's tests (Hernández-López et al., 2020), strip methods (el Moctar et al., 2021, pp. 59–86), and sensory evaluations (Jia et al., 2022) have proven insufficient for reliable, real-time monitoring. Classical analytical techniques, including GC, CE, HPLC, and SPE-LC-MS/MS, accurately measure sugars (Liu et al., 2008; Fraser et al., 2014; Alwis et al., 2020; Piyasena et al., 2022; Przybylska et al., 2021; Wang et al., 2021) but remain time-intensive, laborious, and unsuitable for high throughput screening (Ren et al., 2013). Moreover, single point optical detectors cannot capture spatial heterogeneity in tea samples.

While hyperspectral imaging (HSI) and NIR spectroscopy have been applied to tea analysis (Chen et al., 2006; Luqing et al., 2015; Ahmad et al., 2021; Wang, Liu, et al., 2021; Hu et al., 2023), most studies rely primarily on spectral information, overlooking textural characteristics. This reduces robustness in heterogeneous samples, highlighting the need to integrate spatial and spectral data for rapid, reliable evaluation.

This study presents a nondestructive, algorithm driven multispectral imaging (MSI) system integrating reflectance and transmittance modes, employing a monochrome FLIR Blackfly USB3 camera, thirteen narrow-band LEDs spanning 365 nm to 940 nm, an integrating hemisphere, and dedicated switching circuitry. Following median filtering and Fisher's Discriminant Analysis preprocessing, classification via LDA, K-NN, SVM, FFNN, and CNN (Ekanayake et al., 2022; Wickramasinghe et al., 2021) and polynomial regression for quantification (Ostertagová, 2012) achieved over 93% classification accuracy in brewed tea and R^2 greater than 0.97 for sugar content prediction, validating multispectral imaging as a rapid, noninvasive, cost-effective solution for black tea adulteration detection.

MATERIALS AND METHODS

Sample preparation

Black tea samples were provided by the Hantana Tea Research Institute (TRI) of Sri Lanka in sealed packages to ensure integrity. Samples originated from mid-country tea plantations in the Kandy and Matale districts (Tea Research Institute, 2025). The sample set comprised a pure tea variant with natural sugars and pre-adulterated variants at 5%, 10%, 15%, 25%, and 35% (w/w). Sugar was uniformly blended during the first rolling stage using a miniature environment-controlled manufacturing (ECM) system (National Science Foundation, 2025).

Black tea grades are determined by leaf shape and size (Kamal et al., 2008). Since the proposed method relies on spectral image analysis, samples were ground into powder to eliminate grade variations and standardized particle size. Grinding and sieving minimize leaf morphology influence and enhance light interaction consistency, improving the accuracy of multispectral analysis. The grinding process is conducted by professionals at the TRI, resulting in tea particles filtered through a 1 mm mesh (Tea Research Institute, Sri Lanka, 2003). Six adulteration levels (5% to 35%), including the pure sample, were prepared with five replicates each, acquiring a total of 30 multispectral images per sample type.

Brewed tea was prepared using 20 g of tea powder per 250 ml of boiling water (~100 °C), comparable tea-to-water ratio to ISO 3103 (ISO 3103, 2019) and brewed for 5 to 6 minutes. After cooling to room temperature (21 °C) to prevent vapor condensation on the camera. A 150 ml sample was obtained by filtering through a 0.5 mm mesh to reduce light scattering, following gentle stirring for 2 minutes to ensure homogeneous sugar distribution while minimizing bubble formation. Brewed samples were imaged following the same procedure as powdered samples.

Multispectral image acquisition

The MSI system was calibrated prior to each acquisition to ensure reproducible and high-quality spectral data. Imaging parameters were standardized, with zoom adjusted to minimize background interference, aperture optimized for brightness and depth of field and focus fine-tuned to capture surface details. LED intensity, zoom, aperture, and focus were fixed across all samples, ensuring consistent, high fidelity image collection.

Monochrome images were sequentially captured at thirteen narrow band wavelengths spanning NUV to NIR regions, recording spectral and textural variations critical for detecting sugar adulteration. The system comprised a FLIR BFS-U3-13Y3M CMOS monochrome camera (1.3 MP, USB3 Vision v1.0, 1280 × 1024 resolution, 10-bit ADC) with an onsemi PYTHON1300 sensor offering 365 nm to 940 nm spectral sensitivity aligned with the selected LEDs. Figure 2 illustrates the MSI setup, and Table 1 lists the spectral bands.

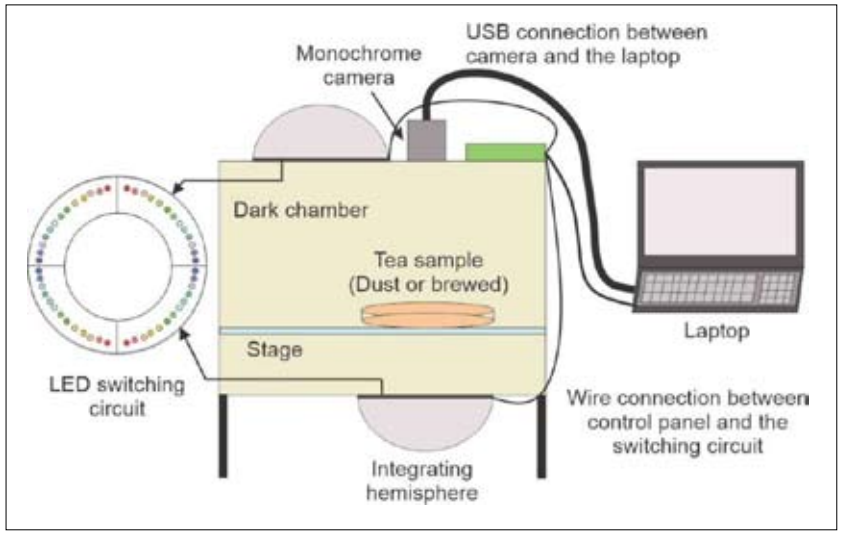


Figure 2: Schematic representation of the multispectral imaging setup

Table 1: An overview of narrow band LEDs used in the integrating hemisphere

Standard LED Part Number Dominant center	Wavelength (nm)
1080-ELUA2016OGB-P6070Q43040020-VA1MCT-ND	365
VLMU3100-GS08CT-ND	405
1830-1108-1-ND	473
1830-1109-1-ND	530
516-2827-1ND	575
516-HSMJ-A100-Q00J1CT-ND	621
754-1614-1-ND	660
1080-67-21S/NFR2CP2050A3B21522Z6/2TCT-ND	735
1125-MTSM 0077-843-IRCTC-ND	770
751-1214-ND	830
VSMY3850-GS08CT-ND	850
751-VSMY5890X01CT-ND	890
1830-IN-P32ZTHIRCT-ND	940

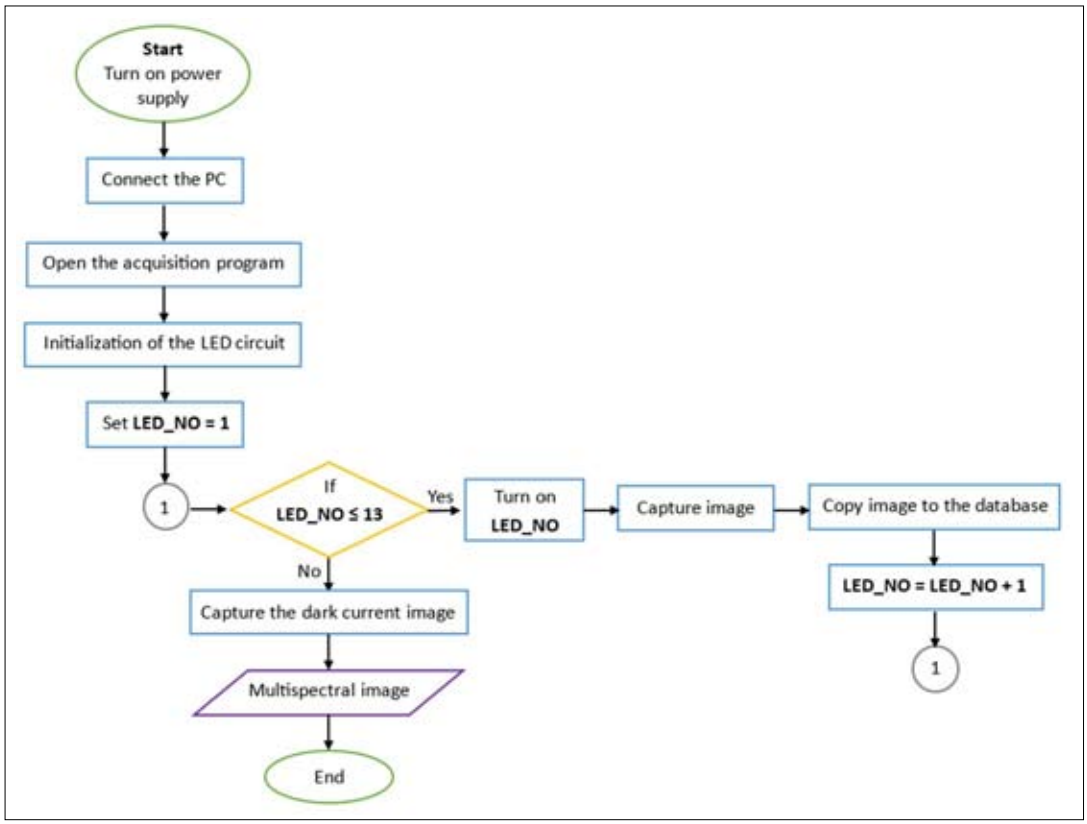


Figure 3: Flow chart of the multispectral imaging process

A custom Graphical User Interface (GUI) synchronized LED activation and image capture, eliminating operator variability. LED switching was controlled via an Arduino Due microcontroller, while the host computer managed acquisition and preprocessing. Powdered samples were imaged in reflectance mode on a wooden stage and cooled brewed samples in transmittance mode on a glass stage. Each sample was imaged across all thirteen spectral bands with dark current correction. The workflow is presented in Figure 3.

Spectral data preprocessing

Image preprocessing

The MSI device captures images of 1280 × 1024 pixels, but only a 100 × 100 pixel² window is used to focus on the brewed or powdered tea, excluding background elements such as the stage or petri dish. This window size balances sufficient data points with computational efficiency, as smaller windows reduce accuracy while larger windows offer minimal improvement.

Multispectral images obtained may contain erroneous pixel values owing to the diverse types of noise within the MSI system, predominantly originating from the camera during pixel acquisition. The photon absorption stimulates electrons within the photodetector (Skauli, 2011), which are converted into electrical signals representing the pixel values.

$$N[i,j] = \eta[i,j]N_{ph}[i,j] + I_d[i,j]t + \delta[N] \quad ...(1)$$

where $N[i,j]$ is the measured photoelectron count at the (i, j) position of the image, $\eta[i,j]$ is the quantum efficiency of the photodetector (the probability of generating a photoelectron from an incoming photon), $N_{ph}[i,j]$ is the incident photon count, $I_d[i,j]$ represents the dark current, t is the measurement time, and $\delta[N]$ denotes the readout noise associated with amplifying the photoelectric signal and converting it to digital values. The last two terms of Equation (1) are noise.

Charges can be generated in the photodetector due to thermal radiation over a unit of time, even without incident light, a phenomenon known as dark current noise (Shieh & Djordjevic, 2010, pp. 83–87). As the enclosure of the MSI device is not completely sealed during image acquisition, environmental radiation may also activate the photodiodes during image acquisition. These errors are mitigated by subtracting the dark current image from each spectral band (Bandara et al., 2020).

$$A[\lambda] = B[\lambda] - D \quad (2)$$

where $B[\lambda]$ is the raw pixel values at λ wavelength, D is the dark current image, and $A[\lambda]$ is the dark corrected pixel value.

A 2D median filter was applied to dark current corrected multispectral images to suppress high frequency electronic fluctuations and illumination nonuniformity. The filter was employed across both rows and columns simultaneously, with each pixel replaced by the median intensity of its 50 × 50 pixel² neighborhood, chosen through trial-and-error by evaluating classification accuracy across window sizes from 5 to 90 pixels. This configuration balances noise reduction and feature preservation, as larger windows may remove spectral details while smaller windows insufficiently improve the signal-to-noise ratio (SNR), maintaining spatial and spectral integrity.

Spectral signature

Each multispectral image consisted of 13 spectral bands, with five acquisitions (replicates) per sugar adulteration level, constrained by available resources at the TRI. Notably, the spatial arrangement (order) of these pixels within the cropped region was disregarded, as the analysis treated individual pixel intensities as independent data points rather than preserving their 2D locations. This pixel level approach substantially augments the effective dataset size for training and testing, minimizing the need for a large number of full images while enabling robust statistical modeling.

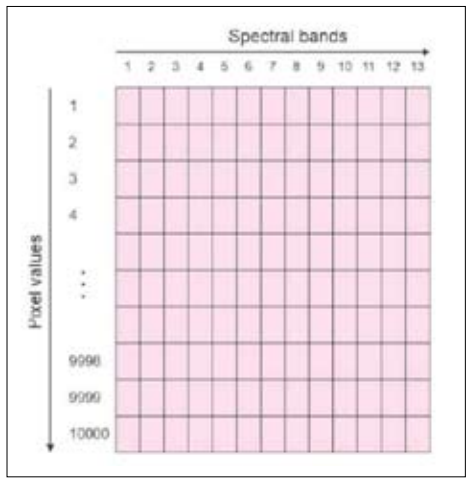


Figure 4: The generated data matrix for one sample

For data organization, a 100 × 100 pixel² region was cropped from each image, resulting in 10,000 pixels per band. These pixels from each of the 13 bands were flattened into column vectors and systematically concatenated in a controlled manner to form a 10,000 × 13 pixel² matrix per sample, where R represents the number of pixels (10,000) and B represents the number of spectral bands (13), defining the general matrix size as $R \times B$ as shown in Figure 4. This process was repeated for all 30 images (5 replicates × 6 sugar levels: 0%, 5%, 10%, 15%, 25%, 35%).

To generate spectral signatures, the mean pixel intensity was calculated across all pixels and replicates for each sugar level and spectral band. The resulting mean values were plotted against the dominant peak wavelength of each band to visualize the spectral response across adulteration levels, as shown in Figure 5.

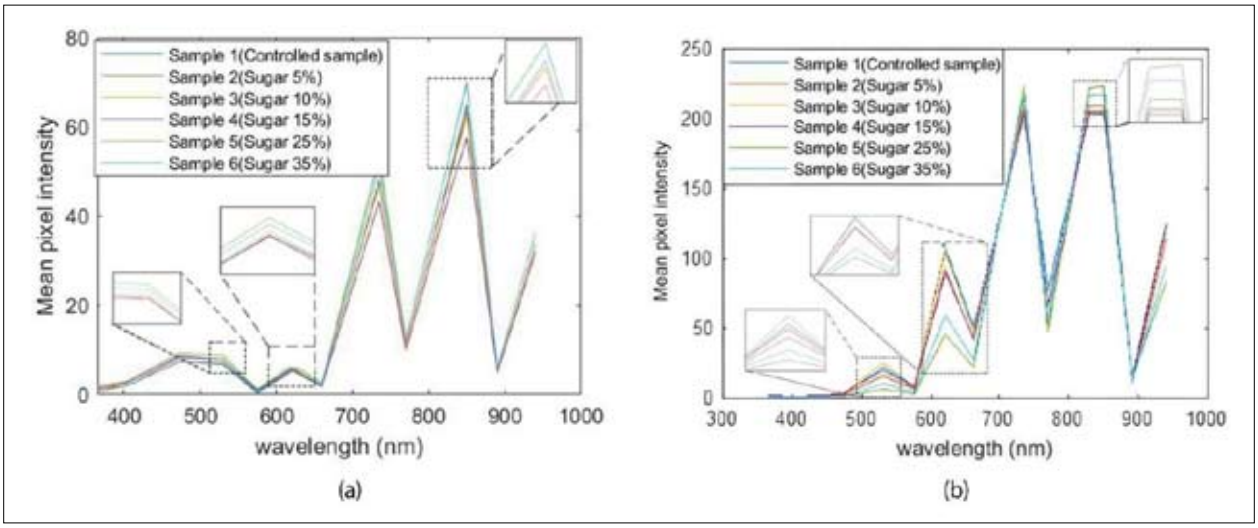


Figure 5: The spectral signature of the (a) powdered and (b) brewed black tea samples

The spectral signatures reveal differentiation among sugar levels at various wavelengths. Notably, brewed black tea shows clearer separations than powdered tea, as illustrated in Figure 5, particularly at peak wavelengths of 530 nm, 621 nm, 830 nm, and 850 nm. This suggests that brewed tea may yield higher classification accuracy. For dataset splitting, an 80:20 training-to-testing ratio was adopted, as recent studies indicate it optimizes performance in machine learning applications (Nguyen et al., 2021). Accordingly, four replicates per sugar level were allocated to training, and one replicate to testing.

The training matrix was formed by vertically stacking the 10,000 × 13 matrices from the training replicates across all classes, yielding a 240,000 × 13 matrix (4 replicates × 10,000 pixels × 6 classes). Similarly, the testing matrix comprised a 60,000 × 13 matrix (1 replicate × 10,000 pixels × 6 classes). Then, the training data matrix and testing data matrix are formed by vertically stacking corresponding $R \times B$ pixel² matrices one below the other. Ultimately, 2D matrices $R_{Tr} \times B$ pixel² (where $R_{Tr} = R \times$ training multispectral images per each class × number of classes = 240,000) and $R_{Te} \times B$ pixel² (where $R_{Te} = R \times$ testing multispectral images per each class × number of classes = 60,000) are created as the training data set and the testing data set, respectively.

Fisher's discriminant analysis

Fisher's Discriminant Analysis (FDA) was applied to enhance class separability before classification (Shashoa et al., 2016). FDA projects the data into a lower-dimensional space by maximizing between-class variation while minimizing within-class variation.

Let the training data set be represented as

$$X = [x_0, x_5, x_{10}, x_{15}, x_{25}, x_{35}]^T$$

where $x_k x_k$ (k = 0, 5, 10, 15, 25, and 35) denotes the dataset for each sugar adulteration level with a dimensionality of $R_{cl} \times B$ (number of rows per each sugar adulterant level, $R_{cl} = R \times$ number of training multispectral images = 40000) (Herath et al., 2023).

Step 01: Calculation of the mean vector of each sugar level.

$$\mu_k = \frac{1}{R_{cl}} \sum_{i=1}^{R_{cl}} x_k(i) \quad ...(3)$$

where $x_k(i)$ is the i^{th} row of x_k sugar adulteration level matrix.

Step 02: Calculation of the total mean vector.

$$\mu = \frac{1}{R_{Tr}} \sum_{i=1}^{R_{Tr}} X(i) \quad ...(4)$$

where $X(i)$ is the i^{th} row of the training data set (X).

Step 03: Calculation of the between class scatter matrix.

$$S_b = \sum_{i=1}^{L} (\mu_i - \mu)(\mu_i - \mu)^T \quad ...(5)$$

where L is the number of sugar adulterant levels, in this particular case L= 6 and μ_i is the mean vector of each sugar level.

Step 04: Calculation of the within class scatter matrix.

$$S_w = \sum_{i=1}^{L} C_i \quad ...(6)$$

$$C_k = \sum_{i=1}^{L} (x_k(i) - \mu_k)(x_k(i) - \mu_k)^T \quad ...(7)$$

where C_k is the covariance matrix of each sugar level (k = 0, 5, 10, 15, 25, and 35).

Step 05: Solve the following eigenvalue problem.

$$S_b w_p = \lambda_p S_w w_p \quad ...(8)$$

where w_p is the eigenvector corresponding to the λ_p eigenvalue (p = 1, 2, . . ., B) and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq \cdots \geq \lambda_{13}$

Step 06: Generation of the projection matrix.

$$W = [w_1, w_2, \ldots, w_k, \ldots, w_p] \quad ...(9)$$

where W is generated by extracting and sorting the eigenvectors corresponding to the eigenvalues in descending order. Here, the dimensionality of W is $B \times B$, and it is possible to reduce the dimensionality of W to $B \times m$ ($m \leq B$).

Step 07: Projection of the training data set into the new space.

$$Z(X) = XW \quad ...(10)$$

where Z is the projected training data set with the dimension of $R_{Tr} \times B$.

The FDA procedure involved computing class and overall means, estimating within and between class scatter, and solving the generalized eigenvalue problem to derive discriminant vectors. These vectors were used to project the datasets into a discriminant subspace, enhancing the visibility of class level differences. The $Z(X)$ illustration presents multiple rotated views to highlight the spatial distribution, as shown in Figures 6 and 7.

The 3D scatter plots project powdered and brewed black tea samples onto the first three FDA components, representing different sugar levels. Figure 6 shows partial separation among classes, with overlaps between intermediate levels (10% to 25%), indicating moderate discriminative capability. In contrast, Figure 7 reveals clearer separation in brewed tea, with distinct clusters along the principal axes, demonstrating FDA's effectiveness in maximizing between-class variance. Overall, brewed tea exhibits a more pronounced separation among sugar levels than powdered tea.

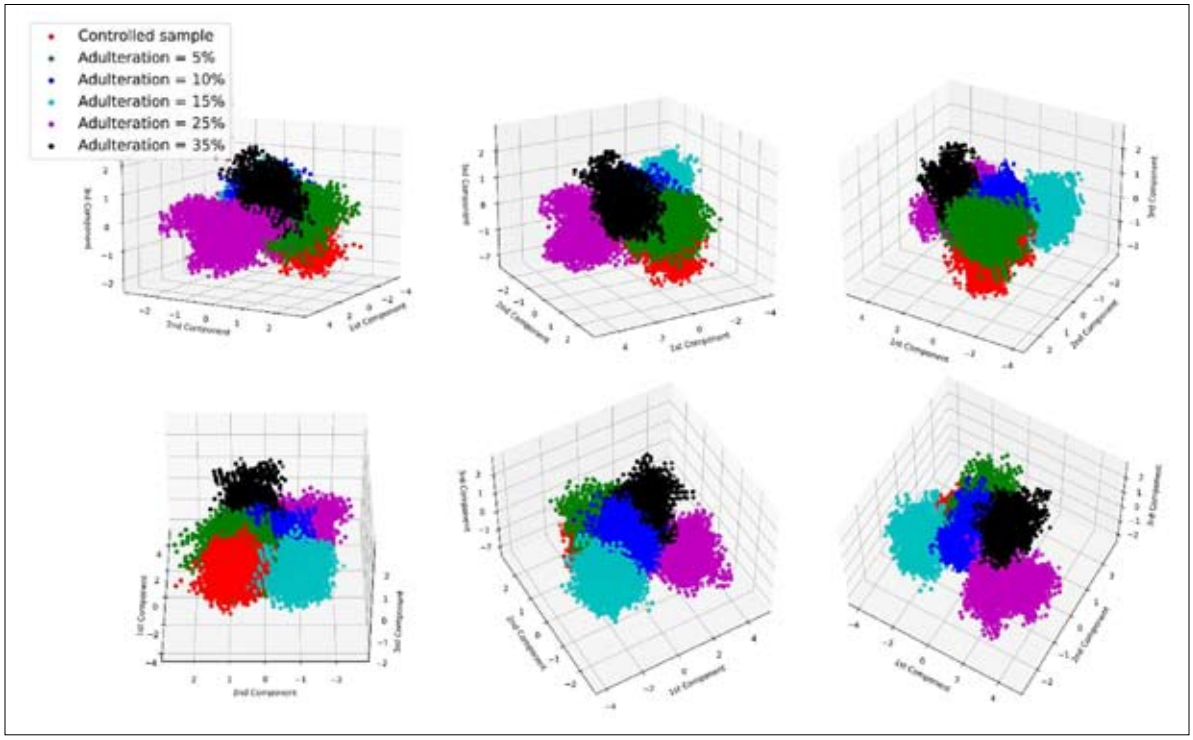


Figure 6: 3D visualization of FDA projected powdered black tea samples across the first three components with multiple rotated views

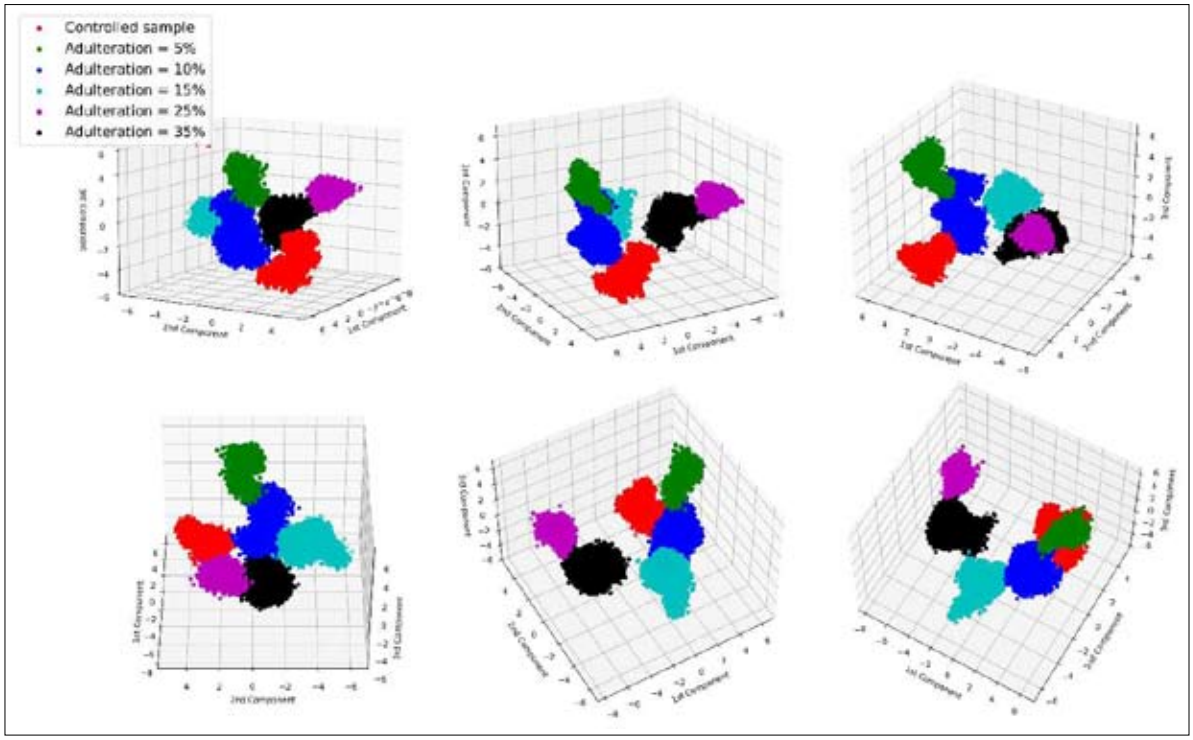


Figure 7: 3D visualization of FDA projected brewed black tea samples across the first three components with multiple rotated views

Classification models

Tea samples were classified based on sugar adulteration levels using machine learning algorithms, such as Linear Discriminant Analysis (LDA), K-Nearest Neighbors (K-NN), Support Vector Machine (SVM), Feed-Forward Neural Network (FFNN), and Convolutional Neural Network (CNN). LDA provided dimensionality reduction while preserving class separability in high-dimensional spectral datasets. K-NN, a non-parametric method, enabled robust detection of local spectral patterns without distributional assumptions, with MATLAB hyperparameter optimization identified optimal performance at k = 1 with cosine distance (Guo et al., 2003). SVM with a Gaussian kernel separated classes in higher dimensional space using one-vs-rest (OVR) multiclass classification (Zhao et al., 2009; Saha et al., 2012).

FFNN captured complex nonlinear relationships between multispectral features and adulteration concentrations using five hidden layers (1024 to 32 neurons) and the Adam optimizer (learning rate 0.0001, selected through trial-and-error) (Murat, 2006). CNN autonomously extracted spatial spectral features from cropped multispectral images (100 × 100 × 13) using a single convolutional layer (3 × 3), max pooling, and one fully connected hidden layer (Alzubaidi et al., 2021; Gu et al., 2018). The CNN was trained over 100 epochs with a batch size of 16, employing Adaptive Moment Estimation (Adam) optimization and a categorical cross entropy loss function for multiclass probabilistic discrimination.

Regression model

Although the classification algorithms showed high accuracy, they were limited in interpolating sugar concentrations between discrete trained levels due to substantial gaps in the data. To address this issue, a polynomial curve fitting model was implemented to enable continuous estimation of sugar content, establishing a functional relationship between independent and dependent variables (Ostertagová, 2012).

Analysis of FDA projected data demonstrated clear class separability, prompting the use of histograms to evaluate sample distributions across sugar adulteration levels as shown in Figure 8. The bell-shaped distributions displayed varying degrees of overlap, with the controlled sample (2% natural sugar) and 35% samples showing significant dispersion. This led to their exclusion to concentrate on the 5% to 25% range. A scatter plot of sugar levels against the first FDA component, shown in Figure 9, indicated horizontal clusters at 5%, 10%, 15%, 20%, and 25%, with component values increasing with higher adulteration, suggesting a predominantly linear yet slightly nonlinear trend amenable to polynomial fitting.

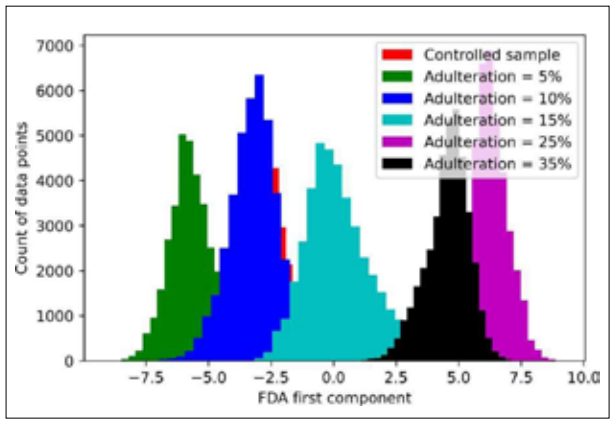


Figure 8: Histogram of the training dataset projected onto the first FDA component

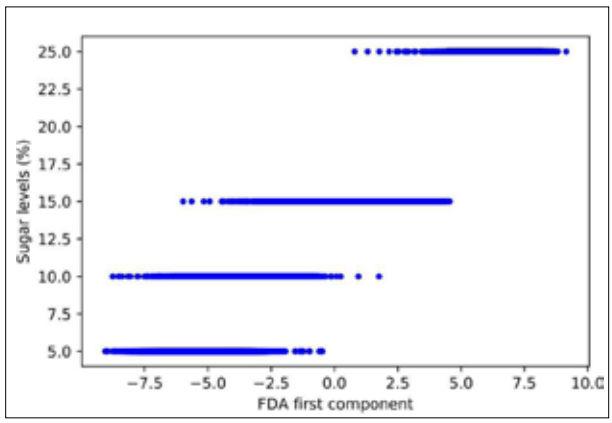


Figure 9: Scatter plot of sugar adulteration levels (5% to 25%) versus the first FDA component, demonstrating clustered data points and the underlying trend for polynomial fitting

The general form of a polynomial function is as follows, with the independent variable X and the dependent variable S.

$$S = \alpha_0 + \alpha_1 X^1 + \alpha_2 X^2 + \cdots + \alpha_n X^n \quad \ldots(18)$$

where n is the order of the polynomial

The mean squared error (MSE) is utilized to evaluate the accuracy of the developed curve, defined as:

$$MSE = \frac{1}{R_{Tr}} \sum_{i=1}^{R_{Tr}} (S_i - \hat{S}_i)^2 \quad ...(19)$$

where S_i is the actual sugar class label and $\hat{S}_i$ is the predicted sugar label.

RESULTS AND DISCUSSION

In this section, the results of both the classification algorithms and the regression algorithm are discussed individually.

Classification algorithms

These algorithms aim to categorize a given data point into a discrete sugar level. Utilizing the data captured by the MSI device and the first five components of the FDA applied data, machine learning algorithms and subsequent neural networks were employed. Results were obtained for both powdered and brewed black tea data to identify the most suitable sample preparation method.

The testing and training accuracies for the K-NN, LDA, SVM, FFNN, and CNN algorithms were evaluated initially. The results for powdered black tea are detailed in Table 2, whereas the results for brewed black tea are presented in Table 3.

Table 2: The training and testing accuracies for the powdered black tea original MSI data

Algorithm	Training accuracy (%)	Testing accuracy (%)
K-NN	99.928	68.470
LDA	80.477	60.870
SVM	89.060	64.889
FFNN	97.800	76.600
CNN	98.750	98.750

Table 3: The training and testing accuracies for the brewed black tea original MSI data

Algorithm	Training accuracy (%)	Testing accuracy (%)
K-NN	96.639	84.695
LDA	99.910	98.603
SVM	99.718	98.852
FFNN	99.771	96.967
CNN	93.229	92.188

Considerably improved validation accuracies were obtained for brewed black tea samples, reflecting the efficacy of multispectral imaging in detecting sugar adulteration signatures. Both LDA and SVM achieved validation accuracies above 98%, emphasizing their robustness for this classification task. However, since SVM demands greater computational resources, LDA emerges as the more efficient choice, particularly under computational constraints.

For neural network-based approaches, FFNN achieved strong validation accuracy, whereas CNN attained 98.75% testing accuracy, closely matching training, demonstrating superior generalization. CNN's convolutional layers more effectively capture complex spectral features in powdered and brewed samples than FFNN's pixel-based approach. This performance suggests robust invariant feature learning, particularly in homogeneous brewed samples. While FFNN outperformed traditional models, CNN's superior handling of high-dimensional data establishes it as the most reliable for unseen sample detection, with LDA offering an efficient alternative when resources are limited.

In theory, the FDA enhances the class separation of the data. To leverage this theoretical advantage, the same algorithms are assessed for training and testing accuracies. The results for powdered black tea and brewed black tea are tabulated in Table 4 and Table 5, respectively.

Table 4: The training and testing accuracies for the powdered black tea FDA applied data

Algorithm	Training accuracy (%)	Testing accuracy (%)
K-NN	99.928	68.498
LDA	88.673	73.200
SVM	69.449	60.510
FFNN	88.900	70.300

Table 5: The training and testing accuracies for the brewed black tea FDA applied data

Algorithm	Training accuracy (%)	Testing accuracy (%)
K-NN	99.877	98.495
LDA	98.695	93.125
SVM	97.146	96.193
FFNN	99.578	93.437

The implementation of the FDA has resulted in decreased testing accuracy, except for LDA in powdered tea and K-NN in brewed black tea. Nonetheless, the overall validation accuracies for brewed black tea remain above 93%.

Brewed black tea images proved more effective for detecting sugar adulteration, as brewing extracts soluble components, including sugar, into the liquid phase, yielding clearer and more uniform spectral changes. In contrast, powdered tea retains heterogeneous particle structures and natural leaf variations, which can obscure subtle adulteration effects. Thus, brewed samples provide more consistent spectral signatures, enabling more accurate classification and highlighting their advantage for multispectral imaging-based detection.

Table 6: The validation accuracy of the fitted polynomial functions of the orders of one to five

Order	R^2 *Score*	Fitted function (P=)
1st order	0.9734	$+1.582641X + 14.63158648$
2nd order	0.9700	$+0.007222X^2 + 1.575401X + 14.476192$
3rd order	0.9739	$-0.003618X^3 + 0.007136X^2 + 1.704775X + 14.582717$
4th order	0.9748	$-0.000200X^4 - 0.003618X^3 + 0.016530X^2 + 1.708238X + 14.534085$
5th order	0.9802	$-0.000652X^5 + 0.000572X^4 + 0.040239X^3 - 0.008211X^2 + 1.064254X + 14.333792$

Regression analysis

Continuous prediction of sugar levels in brewed black tea was performed using regression, with the first FDA principal component as the independent variable (X) and sugar concentration (in percentage) as the dependent variable (P). Polynomial regression models of orders 1st to 5th were trained to capture nonlinear trends, and performance was evaluated using the coefficient of determination (R^2) (Enders, 2022; Zhang et al., 2013).

As shown in Table 6, the linear model (1st order) yielded an $R^2 = 0.9734$, indicating a predominantly linear relationship. Higher order models slightly improved fit, with the 5th order polynomial achieving the highest R^2 of 0.9802. However, coefficients beyond the 3rd order were negligible and added complexity. Given that escalating polynomial order did not yield substantial gains in R^2 and risked overfitting, particularly with data clustered around discrete sugar levels (5%, 10%, 15%, 20%, and 25%) the 3rd order polynomial was selected as the optimal balance between accuracy and parsimony.

Subsequently, the polynomial fitting function of the 3rd order was validated using unseen data from the polynomial fitting process, yielding an R^2 score of 0.9739. The polynomial curve is shown in Figure 10, where the red curve represents the fitted 3rd order polynomial, and the green markings represent the predictions of the validation data. The results demonstrate that the 3rd order polynomial accurately models the relationship between the input spectral features and sugar content, providing reliable predictions for unseen samples.

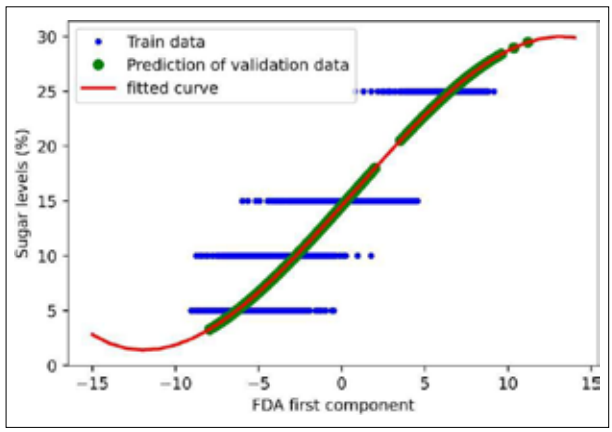


Figure 10: The 3rd order polynomial modeling of sugar content in brewed black tea

The selected 3rd order model achieved robust performance, explaining approximately 97% of the variance in sugar levels. Its high accuracy makes it a reliable tool for quantifying sugar adulteration in brewed black tea, offering improved sensitivity over categorical

classification by enabling continuous predictions within the specified range.

Beyond this study, the MSI setup has demonstrated versatility in detecting turmeric adulteration (Chaminda Bandara et al., 2020), monitoring fungal growth (Darsha Udayanga et al., 2024), quantifying aflatoxin contamination (Madushan et al., 2024), differentiating reheated coconut oil (Ranasinghe et al., 2022), and evaluating soil moisture (Ranasinghe et al., 2025), highlighting its adaptability for interdisciplinary research in food science, agriculture, and construction.

CONCLUSION

The present study evaluated the feasibility of using multispectral imaging to estimate sugar adulteration levels in black tea, using sugar adulterated black tea samples as a case study. Images were captured following standard laboratory conditions of room temperature (21 °C) and atmospheric pressure (0.999 atm) and procedures included controlled sample handling, uniform preparation of powdered and brewed tea, consistent imaging, and proper calibration of instruments to ensure reproducibility and accuracy. Thirteen narrowband LEDs from 365 nm to 940 nm were used for image acquisition. The images were preprocessed via dark current removal and median filtering, followed by FDA based dimensionality reduction. Classification with K-NN, LDA, SVM, and FFNN achieved testing accuracies above 93%, with brewed tea samples showing clearer class separation than powdered tea. Continuous prediction of sugar content in brewed tea was enabled through a 3rd order polynomial regression model, achieving an R^2 of 0.97, with high precision for intermediate sugar levels of 5% to 25%.

This work is novel in integrating a cost effective MSI platform with advanced multistage machine learning and regression models to provide both classification and continuous estimation of sugar adulteration. Unlike conventional chemical methods, the approach is rapid, nondestructive, and requires minimal sample preparation. It has strong applicability in the tea industry and across all sciences and technologies, enabling regulatory bodies, quality assurance laboratories, and exporters to efficiently ensure product authenticity, enhance quality control, and protect consumer trust. The scalable framework can be further refined with higher resolution sensors and expanded spectral ranges, supporting industrial deployment and broader applications in food quality monitoring.

Acknowledgement

The authors gratefully acknowledge the Tea Research Institute (TRI) and the Sri Lanka Tea Board for generously providing the tea samples, laboratory expertise, and support.

Special appreciation is extended to Dr. Keerthi M. Mohotti, Deputy Director of TRI, and Dr. Kumara Kithsiri, Deputy Tea Commissioner of the Sri Lanka Tea Board, for their invaluable guidance and support throughout the research process, which significantly contributed to achieving favorable outcomes.

REFERENCES

Aaqil, M., Peng, C., Kamal, A., Nawaz, T., Zhang, F., & Gong, J. (2023). Tea harvesting and processing techniques and its effect on phytochemical profile and final quality of black tea: A review. *Foods*, *12*(24), 4467. https://doi.org/10.3390/foods12244467

Ahmad, H., Sun, J., Nirere, A., Shaheen, N., Zhou, X., & Yao, K. (2021). Classification of tea varieties based on fluorescence hyperspectral image technology and ABC-SVM algorithm. *Journal of Food Processing and Preservation*, *45*(3). https://doi.org/10.1111/jfpp.15241

Alwis, J. A. P. N., Perera, G. A. A. R., & Abeywickrama, K. R. W. (2020). Determination of Sucrose, Glucose and Fructose Level in Low Grown Black Tea Using High Performance Liquid Chromatography . *International Research Conference of Uva Wellassa University*, 281.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 1–74. https://doi.org/10.1186/s40537-021-00444-8

Cabrera, C., Artacho, R., & Giménez, R. (2006). Beneficial effects of green tea—A review. *Journal of the American College of Nutrition*, *25*(2), 79–99. https://doi.org/10.1080/07315724.2006.10719518

Chaminda Bandara, W. G., Kasun Prabhath, G. W., Sahan Chinthana Bandara Dissanayake, D. W., Herath, V. R., Roshan Indika Godaliyadda, G. M., Bandara Ekanayake, M. P., Demini, D., & Madhujith, T. (2020). Validation of multispectral imaging for the detection of selected adulterants in turmeric samples. *Journal of Food Engineering*, *266*, 109700. https://doi.org/10.1016/j.jfoodeng.2019.109700

Chen, Q., Zhao, J., Zhang, H., & Wang, X. (2006). Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with multivariate calibration. *Analytica Chimica Acta*, *572*(1), 77–84. https://doi.org/10.1016/j.aca.2006.05.007

Ekanayake, E. M. S. L. B., Wickramasinghe, W. A. N. D., Wijesinghe, A. D., Wijedasa, M. A. C. S., Weerasooriya,

H. M. H. K., Ranasinghe, D. Y. L., Ekanayake, M. P. B., Herath, H. M. V. R., Godaliyadda, G. M. R. T., & Madhujith, T. (2021). Multispectral Imaging System to Estimate Sugar Adulteration Level of Black Tea. *2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS)*, 17–22.

Fraser, K., Harrison, S. J., Lane, G. A., Otter, D. E., Hemar, Y., Quek, S.-Y., & Rasmussen, S. (2014). Analysis of low molecular weight metabolites in tea using mass spectrometry-based analytical methods. *Critical Reviews in Food Science and Nutrition*, *54*(7), 924–937. https://doi.org/10.1080/10408398.2011.619670

Kamer, L. (2022, August 1). Tea industry in Kenya-statisticsfacts. Statista.https://www.statista.com/topics/8363/tea-industry-in-kenya/

el Moctar, B. O., Schellin, T. E., & Söding, H. (2021). *Numerical Methods for Seakeeping Problems* (pp. 59–86). Springer Nature.

Enders, F. B. (2022). Coefficient of Determination | Interpretation & Equation | Britannica. In *Encyclopædia Britannica*.

Engelhardt, U. H. (2010). *Comprehensive Natural Products II* (pp. 999–1032). Elsevier. https://doi.org/10.1016/B978-008045382-8.00089-7.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN Model-Based Approach in Classification. *On the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, *2888*, 986–996. https://doi.org/10.1007/978-3-540-39964-3_62

Herath, S., Weerasooriya, H. K., Ranasinghe, D. Y. L., Bandara, W. G. C., Herath, V. R., Godaliyadda, R. I., Ekanayake, M. P. B., & Madhujith, T. (2023). Quantitative assessment of adulteration of coconut oil using transmittance multispectral imaging. *Journal of Food Science and Technology*. https://doi.org/10.1007/s13197-023-05697-0

Hernández-López, A., Sánchez Félix, D. A., Zuñiga Sierra, Z., García Bravo, I., Dinkova, T. D., & Avila-Alejandre, A. X. (2020). Quantification of reducing sugars based on the qualitative technique of benedict. *ACS Omega*, *5*(50), 32403–32410. https://doi.org/10.1021/acsomega.0c04467

Hu, Y., Huang, P., Wang, Y., Sun, J., Wu, Y., & Kang, Z. (2023). Determination of Tibetan tea quality by hyperspectral imaging technology and multivariate analysis. *Journal of Food Composition and Analysis*, *117*, 105136–105136. https://doi.org/10.1016/j.jfca.2023.105136

IBEF. (2021, September 23). The tea industry in India. *India Brand Equity Foundation*. https://www.ibef.org/blogs/the-tea-industry-in-india

ISO 3103. (2019). *Tea — Preparation of liquor for use in sensory tests* (ISO 3103:2019(E)). International Organization for Standardization. htps:/standards.iteh.ai/catalog/standards/sist/a731214d-fbae-4850-9bc65c232bdd73c3/iso-3103-2019.

Jia, X., Ren, J.-N., Fan, G., Reineccius, G. A., Li, X., Zhang, N., An, Q., Wang, Q., & Pan, S. (2022). Citrus juice off-flavor during different processing and storage: Review of odorants, formation pathways, and analytical techniques. *Critical Reviews in Food Science and Nutritio*, *64*, 3018–3043. https://doi.org/10.1080/10408398.2022.2129581

Kamal, R. J., Wadood, A. W. M. R., & Silva, K. D. R. R. (2008). Total antioxidant capacity of selected grades of black tea grown in different geographical elevations in Sri Lanka. *Tropical Agricultural Research*, *20*, 155–161.

Liu, E-Hu., Qi, L.-W., Cao, J., Li, P., Li, C.-Y., & Peng, Y.-B. (2008). Advances of modern chromatographic and electrophoretic methods in separation and analysis of flavonoids. *Molecules*, *13*(10), 2521–2544. https://doi.org/10.3390/molecules13102521

Luqing, L., Lingdong, W., Jingming, N., & Zhengzhu, Z. (2015). Detection and quantification of sugar and glucose syrup in roasted green tea using near infrared spectroscopy. *Journal of near Infrared Spectroscopy*, *23*(5), 317–325. https://doi.org/10.1255/jnirs.1178

Markets and Trade Division - Economic and Social Development stream . (2022). *International tea market: market situation, prospects and emerging issues (FAO document CC0238EN/1/05.22)* (pp. 3–5). Food and Agriculture Organization of the United Nations.

Meng, L., Chen, X., Chen, X., Yuan, L., Shi, W., Cai, Q., & Huang, G. (2019). Linear and nonlinear classification models for tea grade identification based on the elemental profile. *Microchemical Journal*, *153*, 104512–104512. https://doi.org/10.1016/j.microc.2019.104512

Murat H., S. (2006). A brief review of feed-forward neural networks. *Communications Faculty of Science University of Ankara*, *50*(1), 11–17. https://doi.org/10.1501/commua1-2_0000000026

Nadeera, D. (2022, January 9). *Lankan tea exports earned $ 1.3 Bn in 2021*. The Island Online; Upali Newspapers (Pvt) Ltd. https://island.lk/lankan-tea-exports-earned-1-3-bn-in-2021/

National Science Foundation. (2025). National Instrument Database, National Science Foundation of Sri Lanka. https://nid.nsf.gov.lk/instrumentView/219

Nelum P. Piyasena, K. G., Hettiarachchi, L. S. K., D.P.S. Jayawardhane, S. A., U. Edirisinghe, E. N., & Jayasinghe, W. S. (2022). Evaluation of inherent fructose, glucose and sucrose concentrations in tea leaves (*Camellia sinensis* L.) and in black tea. *Applied Food Research*, *2*(1), 100100. https://doi.org/10.1016/j.afres.2022.100100

Nguyen, Q. H., Ly, H.-B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., Prakash, I., & Pham, B. T. (2021). Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Mathematical Problems in Engineering*, *2021*, 1–15. https://doi.org/10.1155/2021/4832864

Ostertagová, E. (2012). Modelling using Polynomial Regression. *Procedia Engineering*, *48*(1), 500–506. https://doi.org/10.1016/j.proeng.2012.09.545

Polat, A., Kalcıoğlu, Z., & Müezzinoğlu, N. (2022). Effect of infusion time on black tea quality, mineral content and sensory properties prepared using traditional Turkish infusion method. *International Journal of Gastronomy and Food Science*, *29*, 100559. https://doi.org/10.1016/

j.ijgfs.2022.100559

Przybylska, A., Gackowski, M., & Koba, M. (2021). Application of capillary electrophoresis to the analysis of bioactive compounds in herbal raw materials. *Molecules*, *26*(8), 2135. https://doi.org/10.3390/molecules26082135

Ren, G., Wang, S., Ning, J., Xu, R., Wang, Y., Xing, Z., Wan, X., & Zhang, Z. (2013). Quantitative analysis and geographical traceability of black tea using Fourier transform near-infrared spectroscopy (FT-NIRS). *Food Research International*, *53*(2), 822–826. https://doi.org/10.1016/j.foodres.2012.10.032

Ridder, M. (2023). *Leading tea exporters worldwide in 2022*. Statista. https://www.statista.com/statistics/264189/main-export-countries-for-tea-worldwide/#statisticContaine

Samanta, S. (2020). Potential bioactive components and health promotional benefits of tea (Camellia sinensis). *Journal of the American College of Nutrition*, *41*(1), 1–29. https://doi.org/10.1080/07315724.2020.1827082

Shashoa, N. A. A., Salem, N. A., Jleta, I. N., & Abusaeeda, O. (2017). Classification depend on linear discriminant analysis using desired outputs. *2016 17th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, 328–332. DOI: 10.1109/STA.2016.7952041.

Shieh, W., & Djordjevic, I. (2010). OFDM for Optical Communications. In *Elsevier eBooks* (pp. 83–87). Academic Press. https://doi.org/10.1016/b978-0-12-374879-9.00003-4. https://doi.org/10.1016/B978-0-12-374879-9.00003-4.

Skauli, T. (2011). Sensor noise informed representation of hyperspectral data, with benefits for image storage and processing. *Optics Express*, *19*(14), 13031. https://doi.org/10.1364/oe.19.013031

Tea Research Institute. (2025). *Mid Country Regional Centre*. Tea Research Institute of Sri Lanka, Talawakelle. https://www.tri.lk/mid-country-regional-centre/

Tea Research Institute, Sri Lanka. (2003, September). *Manufacture of the tea by the CTC process*. (Serial No. 11/72 and T4, Serial No. 12/72).

Wang, H., Hu, L., Zhou, P., Ouyang, L., Chen, B., Li, Y., Chen, Y., Zhang, Y., & Zhou, J. (2021). Simultaneous determination of fructose, glucose and sucrose by solid phase extraction-liquid chromatography-tandem mass spectrometry and its application to source and adulteration analysis of sucrose in tea. *Journal of Food Composition and Analysis*, *96*, 103730. https://doi.org/10.1016/j.jfca.2020.103730

Wedagedara, H., Wanigasuriya, G., Nissanka , S. P., Mohotti, A., Mendis, E., & Botheju, W. S. (2019). Detection of sugar adulteration in black tea and its impact on quality parameters. *22nd Peradeniya University International Research Sessions (IPURSE)*, *22*, 151.

Wickramasinghe, W. A. N. D., Ekanayake, E. M. S. L. B., Wijedasa, M. A. C. S., Wijesinghe, A. D., Madhujith, T., Ekanayake, M. P. B., Godaliyadda, G. M. R. I., & Herath, H. M. V. R. (2021). Validation of multispectral imaging for the detection of sugar adulteration in black tea. *2021 10th International Conference on Information and automation for sustainability (ICIAfS)*, 494–499.

XU, Y., QIAO, F., & HUANG, J. (2022). Black tea markets worldwide: Are they integrated? *Journal of Integrative Agriculture*, *21*(2), 552–565. https://doi.org/10.1016/s2095-3119(21)63850-9

Xu, Y., Zhang, H., Zhang, C., Wu, P., Li, J., Xia, Y., & Fan, S. (2019). Rapid prediction and visualization of moisture content in single cucumber (Cucumis sativus L.) seed using hyperspectral imaging technology. *Infrared Physics & Technology*, *102*, 103034–103034. https://doi.org/10.1016/j.infrared.2019.103034

Yu, H., Qing, L.-W., Yan, D.-T., Xia, G., Zhang, C., Yun, Y., & Zhang, W. (2021). Hyperspectral imaging in combination with data fusion for rapid evaluation of tilapia fillet freshness. *Food Chemistry*, *348*, 129129–129129. https://doi.org/10.1016/j.foodchem.2021.129129

Zhang, R., Xu, P., Guo, L., Zhang, Y., Li, P., & Yao, D. (2013). Z-Score linear discriminant analysis for EEG Based brain-computer interfaces. *PLoS ONE*, *8*(9), e74433. https://doi.org/10.1371/journal.pone.0074433

J.Natn.Sci.Foundation Sri Lanka 2026 54 (1): 143 - 147
DOI: http://dx.doi.org/10.4038/jnsfsr.v54i1.12506

RESEARCH COMMUNICATION

Mycology

First report of *Ramularia coleosporii* associated with rust sori of *Coleosporium plumeriae* on *Plumeria* species in Sri Lanka

WPT Wijenayake[1], R Rienzie[1*] and RHG Ranil[2]
[1] *Department of Agricultural Biology, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka.*
[2] *Department of Crop Science, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka.*

Submitted: 02 May 2025; Revised: 02 January 2026; Accepted: 12 March 2026

Abstract: In this study, we report the first instance of the fungus *Ramularia coleosporii*, isolated in the sori of *Plumeria* leaf rust caused by *Coleosporium plumeriae* in Sri Lanka. The fungus *R. coleosporii* was cultured on potato dextrose agar (PDA) medium, DNA was isolated, and PCR was conducted targeting the internal transcribed spacer region (ITS) of the genome. Phylogenetic analysis using the obtained sequence data and other homologous sequences retrieved from GenBank revealed a close relationship with *Ramularia coleosporii* strains reported from Eastern and Western regions of Asia. The Sri Lankan isolate showed 99% sequence identity with *R. coleosporii* from South Korea, with strong bootstrap support (99%) for this identification. Although *R. coleosporii* has been reported as a hyperparasite of plumeria rust in other regions, hyperparasitism was not experimentally verified in this study.

Keywords: *Coleosporium plumeriae*, ITS phylogeny, *Plumeria* leaf rust, *Ramularia coleosporii*.

INTRODUCTION

Plumeria, commonly grown as an ornamental tree belongs to the family Apocynaceae and is well adapted to tropical and subtropical climates worldwide. This introduced species in Sri Lanka is locally known as "Araliya" or the temple tree. Two species commonly found on the island are *Plumeria rubra* L. (Figure 1a) and *Plumeria obtusa* L. (Figure 1b) with flowers that are often used in religious ceremonies (Manimohan & Mannethody, 2011). The rust disease was first reported in Sri Lanka in 2006 (Weeraratne & Adikaram, 2006), and can be seen on both *Plumeria* species in Sri Lanka.

In the present case, whitish mycelium-like structures containing sori or rusts were observed on the abaxial surface of leaves of *P. rubra* collected from a tree located on the periphery of the Hanthana conservation forest in Peradeniya, Central Province, Sri Lanka (Figure 2). It was noted that the whitish mycelium appears after 4-5 weeks of the rust's initiation. Previous studies have reported the occurrence of hyperparasites on rust fungi in other regions of the world. Mycoparasites are considered as potential biocontrol agents against plant pathogens (Baiswar et al., 2015). Various species of *Ramularia* have been documented as hyperparasites of rust fungi on different host plants across multiple geographic regions (Pirnia et al., 2021).

MATERIALS AND METHODS

Rust spores from sori were collected using a paintbrush, suspended in sterile distilled water, and transferred to Potato Dextrose Agar (PDA) medium. After three weeks of incubation at 27 °C (relative humidity 78%), a cottony fungal mycelium was observed and single spores were isolated and then incubated under the same conditions (Figure 3).

* Corresponding author (ryanrienzie@agri.pdn.ac.lk; https://orcid.org/0000-0002-1774-858X)

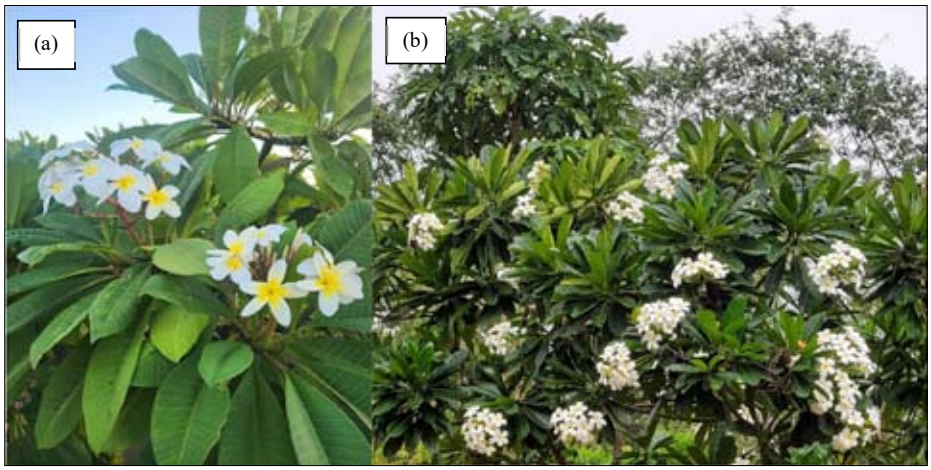


Figure 1: *Plumeria* species (a) *P. rubra* L.; (b) *P. obtusa* L.

Figure 2: The white mycelia of *R. coleosporii*, colonizing the rust sori found on the abaxial surface of the leaves of *P. rubra* L.

Figure 3: Culture characteristics of *R. coleosporii* isolate PQ142333 grown on Potato Dextrose Agar (PDA) medium showing cottony mycelium after single spore isolation; Left half; upper surface; right half; lower surface.

Fresh mycelia (100 mg) were prepared by scraping colonies from a two-week old culture grown on PDA medium at 27 °C in 1.5 mL Eppendorf tubes for the isolation of genomic DNA. Mycelia were homogenized for 20 s with liquid nitrogen using a pestle and mortar. To this homogenate, 500 µL of extraction buffer [200 mmol/L of Tris-HCl (pH 7.5)], 250 mmol/ L of NaCl, 25 mmol/L of EDTA, and 0.5% of SDS (Edwards et al., 1991) were added, and the solutions were mixed and homogenized for five minutes with a Vortex Genie 2 mixer at the maximum speed. Then, 350 µL of phenol was added to each tube and gently inverted four times followed by the addition of 250 µL of chloroform, and the tubes were mixed by inverting them 40 times. The samples were then centrifuged at 2834 x g for 30 min at 4 °C. Subsequently, the aqueous phase of all the samples was transferred to new tubes. The samples were treated with 2 µL RNase (20 µg/mL TE) and allowed to incubate at 37 °C for 10 min, followed by mixing equal volumes of chloroform. The tubes were centrifuged at 4 °C at 2834 x g for 10 min, and the aqueous layer was transferred to new Eppendorf tubes containing 0.5 volume of cold

isopropanol to precipitate the DNA. Tubes with the samples were spun at 4 °C for 2834 x g for 5 min, and the supernatant was discarded by gently pouring it into a sink. Cold 70% ethanol (100 µL) was added to the pellet, and the samples were centrifuged at 4 °C, 2834 x g for 5 min, and the ethanol was discarded. Each pellet was then incubated at 37 °C for 30 min, and suspended in 50 µL of sterile distilled water. The optimum DNA yields and quality of the subsequent DNA were analyzed by the ethidium bromide-stained 1% (w/v) agarose gel method as described by Edwards et al. (1991).

PCR reactions were done using the primers ITS1 (5'-TCCGTAGGTGAACCTGCGG-3') and ITS4 (5'-TCCTCCGCTTATTGATATGC-3') (White et al., 1990). The amplification reaction for each sample was obtained with 0.20 mM of each dNTP, 0.4 mM each of the forward and reverse primers, 0.5U Taq polymerase (Thermo Fisher Scientific, Waltham, MA, USA), 2 µL of genomic DNA solution, and 10 × Easy Taq buffer, bringing the volume up to 50 µL. A typical reaction was as follows: denaturation at 95 °C for 5 min, 35 cycles of denaturation at 95 °C for 50 s, annealing at 52 °C for 50 s, extension

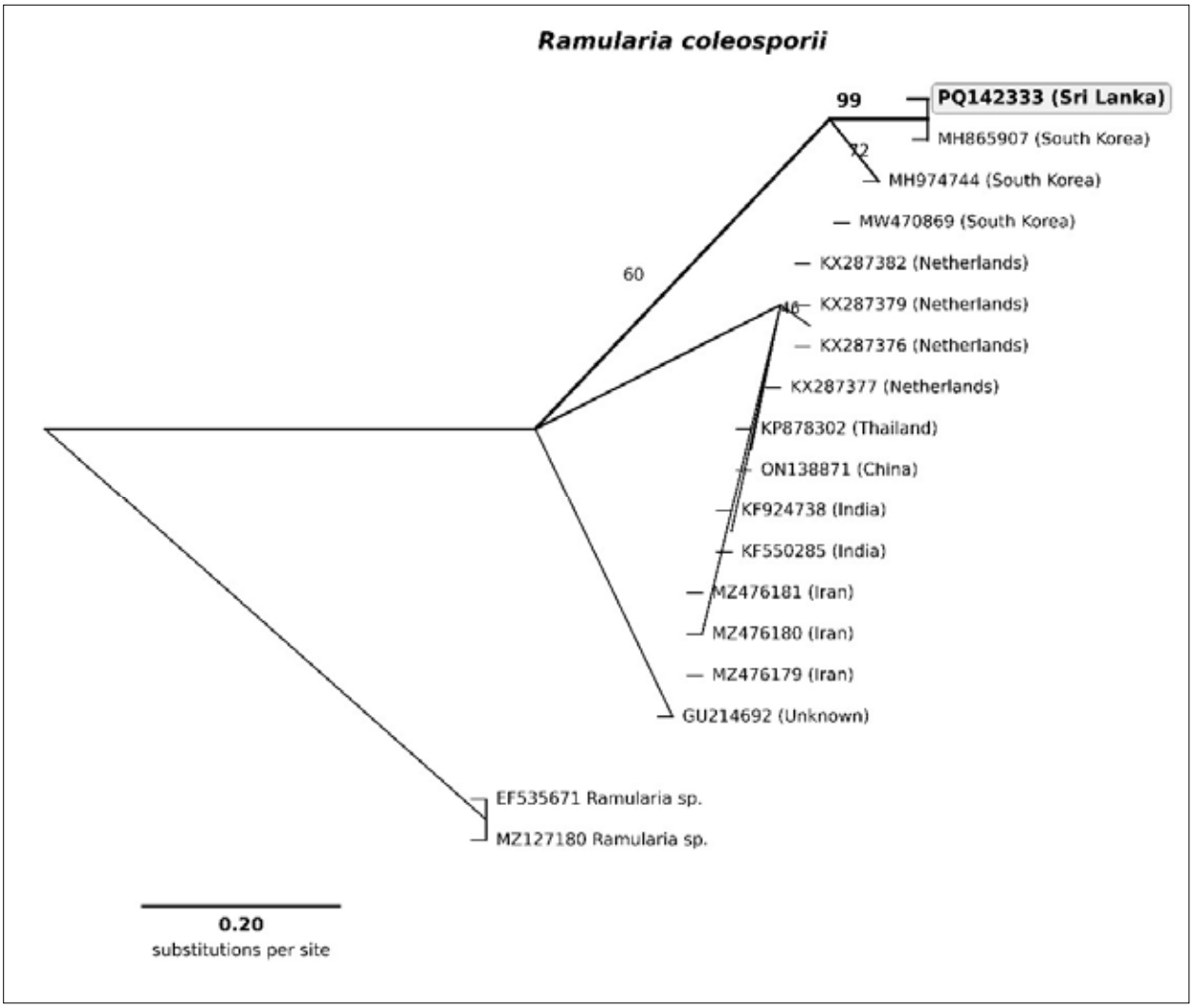


Figure 4: Maximum Likelihood phylogenetic tree of *Ramularia coleosporii* based on ITS sequence data. The tree was constructed using the Tamura-Nei model with 1000 bootstrap replications in MEGA11. Bootstrap support values ≥45% are shown at nodes. The Sri Lankan isolate (PQ142333) is highlighted and clusters with the South Korean isolate MH865907 with 99% bootstrap support. *Ramularia* spp. sequences serve as outgroups. The scale bar indicates 0.20 substitutions per site.

at 72 °C for 60 s, and final extension for another 10 min at 72°C. The PCR products were purified using a PCR purification kit (Qiagen) and sequenced in both directions using the same primers at Macrogen Inc., South Korea.

RESULTS AND DISCUSSION

The sequence was deposited in GenBank under the accession number PQ142333. BLAST analysis showed that the analyzed isolate was 99% homologous to *R. coleosporii* isolate accessioned as MH865907 reported from South Korea. Phylogenetic analysis was conducted in MEGA11 using the maximum likelihood method with the Tamura-Nei model (Tamura & Nei, 1993). Bootstrap replications were set to 1000 to enhance the credibility of the branching patterns. The phylogenetic analysis indicated that the Sri Lankan isolate clusters with *R. coleosporii* sequences from various geographical regions with strong statistical support. The main identification node (Sri Lankan isolate + South Korean isolate MH865907) showed 99% bootstrap support, providing reliable species identification. Internal nodes showed moderate support values (45-72%), which is typical for closely related intraspecific sequences with minimal genetic divergence (Figure 4).

The use of ITS sequences for fungal species identification follows established protocols, as ITS has been validated as the primary DNA barcode for fungi (Schoch et al., 2012). For *Ramularia* species specifically, ITS-based phylogeny has proven effective for species delimitation and has been extensively used in taxonomic studies (Videira et al., 2015; Bakhshi, 2017). While multi-locus phylogenetics could provide additional resolution for population genetic studies, the combination of high sequence similarity and strong bootstrap support observed in this study provides sufficient evidence for species-level identification.

The phylogenetic analysis places the Sri Lankan isolate within a well-supported clade of *R. coleosporii* sequences from multiple geographic origins including South Korea, Netherlands, India, Iran, Thailand, and China. This cosmopolitan distribution suggests that *R. coleosporii* is widely distributed as a fungicolous fungus associated with rust diseases across diverse climatic zones.

It is important to note that while *R. coleospori*i has been reported as a hyperparasite of rust fungi in other studies, the hyperparasitic nature of this interaction was not experimentally verified in the present study through parasitism assays or microscopic examination of host-parasite interactions. Our observations were limited to the morphological association of *R. coleosporii* with rust sori on *Plumeria* leaves. Further studies, including detailed microscopic observations of fungal penetration into rust structures and pathogenicity tests, are needed to conclusively demonstrate hyperparasitism in the Sri Lankan context. This represents the first confirmed report of *R. coleosporii* associated with *Plumeria* rust in Sri Lanka, expanding the known geographic distribution of this fungicolous species. The finding contributes to our understanding of the mycobiota associated with rust diseases in tropical environments and highlights the need for comprehensive studies of fungal interactions.

Acknowledgement

The authors hereby acknowledge the University Research Grant (URG/2022/06/Ag) of the University of Peradeniya.

REFERENCES

Aktaruzzaman, M., Afroz, T., Choi, H. W., & Kim, B. S. (2021). First report of *Ramularia coleosporii* causing leaf spot on *Perilla frutescens* in Korea. *Plant Disease*, *105*(9), 2727. https://doi.org/10.1094/PDIS-01-21-0223-PDN

Baiswar, P., Chandra, S., Ngachan, S. V., & Kumar, R. (2015). First report of *Ramularia coleosporii* causing hyperparasitism on *Coleosporium asterum* in India. *Plant Disease*, *99*(2), 289.

Bakhshi, M. (2017). Multi-gene phylogeny reveals new species in *Ramularia*. *Mycological Progress*, *16*, 703–712. https://doi.org/10.1007/s11557-017-1308-y

Edwards, K., Johnstone, C., & Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research*, *19*(6), 1349. https://doi.org/10.1093/nar/19.6.1349

Manimohan, P., & Mannethody, S. (2011). *Zygosporium gibbum*: A new and remarkable rust hyperparasite. *Mycosphere*, *2*(3), 219-222.

Pirnia, M., Brazauskiene, I., & Shams, E. (2021). Morphological and molecular characterization of *Ramularia* species associated with cereal crops in Iran. *Mycological Progress*, *20*(3), 277-289.

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., & Fungal Barcoding Consortium. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, *109*(16), 6241-6246. https://doi.org/10.1073/pnas.1117018109

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*(3), 512-526.

Videira, S. I. R., Groenewald, J. Z., Verkley, G. J. M., Braun, U., & Crous, P. W. (2015). The rise of *Ramularia* from the *Mycosphaerella* labyrinth. *Fungal Biology*, *119*(9), 823-843. https://doi.org/10.1016/j.funbio.2015.06.003

Weeraratne, T. P., & Adikaram, N. K. B. (2006). Biology of *Plumeria* leaf rust disease caused by *Coleosporium plumeriae*. *Ceylon Journal of Science (Biological Sciences)*, *35*(2), 157-162.

White, T. J., Bruns, T. D., Lee, S. B., & Taylor, J. W. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. A. Innis, D. H. Gelfand, J. J. Sninsky, & T. J. White (Eds.), *PCR protocols: A Guide to Methods and Applications* (pp. 315-322). Academic Press. https://doi.org/10.1016/b978-0-12-372180-8.50042-1

J.Natn.Sci.Foundation Sri Lanka 2026 54 (1): 149 - 153
DOI: http://dx.doi.org/10.4038/jnsfsr.v54i1.12686

RESEARCH COMMUNICATION

Biotechnology

Impact of *Gliricidia sepium* mulch enriched with native Sri Lankan bacterial strains on cinnamon (*Cinnamomum verum*)

PKD Sandamini[1], IM Mulec[2], E Grabner[2], HWKSL Kumara[1], S Ranaweera[3], NVT Jayaprada[4], RAAK Ranawaka[5] and S Geekiyanage[6*]

[1] *Faculty of Graduate Studies, University of Ruhuna, Matara, Sri Lanka.*
[2] *Department of Microbiology, Faculty of Biotechnical, University of Ljubljana, Slovenia.*
[3] *Department of Chemistry, Faculty of Science, University of Ruhuna, Matara, Sri Lanka.*
[4] *Department of Agricultural Technology, Faculty of Technology, University of Colombo, Pitipana, Sri Lanka.*
[5] *Department of Cinnamon Development, Gunarathna Weerakoon Mawatha, Karandeniya, Sri Lanka.*
[6] *Department of Agricultural Biology, Faculty of Agriculture, University of Ruhuna, Mapalana, Kamburupitiya, Sri Lanka.*

Submitted: 29 May 2025; Revised: 23 February 2026; Accepted: 09 March 2026

Abstract: The integration of organic mulches with beneficial microorganisms offers a sustainable strategy for enhancing soil fertility and plant productivity. This study evaluated the effects of *Gliricidia sepium* mulch enriched with two native bacterial strains (I-I and I-II) from Sri Lanka on the vegetative growth, essential oil yield, and soil nutrient status of *Cinnamomum verum* variety *Sri Gemunu* under greenhouse conditions. A 12-month experiment was conducted in a completely randomized design with three treatments: control (mulch only), bacterial strain I-I enriched mulch (G I-I), and bacterial strain I-II enriched mulch (G I-II). Vegetative growth parameters, essential oil yield, and soil nutrient levels were assessed. Both bacterial treatments significantly improved plant height, stem circumference, leaf number, and root length compared to the control, with G I-II exhibiting significant performance in most growth metrics. Soil analysis indicated significantly elevated phosphorus levels in both inoculated treatments, while nitrogen levels increased only in the G I-I treatment. These findings highlight the potential of plant growth promoting bacteria-enriched *Gliricidia sepium* mulch to enhance vegetative growth and soil fertility in cinnamon cultivation, suggesting potential gains in overall oil productivity, supporting its relevance in sustainable cinnamon production systems.

Keywords: Bacterial strains, *Cinnamomum verum*, *Gliricidia sepium* mulch, soil nutrient content, vegetative growth.

INTRODUCTION

Sustainable agricultural practices are increasingly being explored to enhance crop productivity while minimizing environmental degradation. Organic mulching has gained attention due to its ability to improve soil fertility, moisture retention, and microbial activity (Rossi et al., 2024). Among various organic mulches, *Gliricidia sepium* has been widely recognized for its nitrogen-fixing ability and as an effective soil enhancer by contributing to increase total soil carbon, nitrogen, available phosphorus, pH balance, cation exchange capacity, and organic matter (Alamu et al., 2023). However, there are no reports on enhancing the efficacy of *Gliricidia sepium* mulch through microbial inoculation to improve nutrient availability for plant growth.

The introduction of plant growth promoting bacterial strains helps protect plants from abiotic stresses and biotic stresses, while enhancing plant growth through various direct and indirect mechanisms, including nitrogen fixation, phosphorus solubilization, production of phytohormones, and competing for space on plant roots

* Corresponding author (sudarshanee@agbio.ruh.ac.lk; https://orcid.org/0000-0002-3771-2680)

(De Souza et al., 2015). The incorporation of beneficial bacterial strains in mulch treatments has been reported to enhance soil microbial diversity and nutrient cycling. Recently, Ebrahimi et al. (2023) explored the effect of combined application of mycorrhiza fungi, growth-promoting bacteria, and mulch of plant residues on the enhancement of plant growth and essential oil yield of the medicinal plant anise (*Pimpinella anisum* L.).

True cinnamon (*Cinnamomum verum*) is the most valuable spice crop cultivated in Sri Lanka for export income (Sri Lanka Export Development Board, 2023). The variety *Sri Gemunu* has gained prominence due to its high essential oil content and superior growth characteristics (Department of Cinnamon Development, 2024). Enhancing its growth through organic methods, such as *Gliricidia sepium* mulch enriched with plant growth-promoting bacteria, could be suggested as an innovative alternative to heavy use of synthetic fertilizers in cinnamon cultivation.

This study investigates the effects of *Gliricidia sepium* mulch enriched with bacterial strains I-I and I-II on the growth, essential oil yield, and soil nutrient composition of *Cinnamomum verum* variety *Sri Gemunu*. The research aims to introduce a novel approach for cinnamon fertilizer use, contributing to sustainable cinnamon cultivation.

MATERIALS AND METHODS

Bacterial strains

Sri Lankan bacterial strains (I-I and I-II) available in the collection at the Department of Agricultural Biology, Faculty of Agriculture, University of Ruhuna, were used for the experiment. Of these, the *16s rRNA* (GenBank: PV174539 and PV174565) and *Gyrase A* (GenBank: PV253967 and PV253968) gene sequences are available. According to Sandamini et al. (2025), these strains were partially identified in a previous study: strains I-I and I-II belonged to the genus *Bacillus* based on the *gyr*A sequencing having over 99 % BLAST identity with matches identified by others as *Bacillus cereus*.

Experimental setup and treatment application

The study was carried out at the greenhouse of the Faculty of Agriculture, University of Ruhuna, Mapalana, Sri Lanka, for 12 months. The experiment followed a completely randomized design (CRD) with five replicates per treatment.

Each experimental unit consisted of a UV treated polyethylene grow bag from the local market (60.96 × 60.96 cm), each filled with 20 kg of growth substrate, composed of an equal mixture of topsoil, cow dung, coir dust, and soil (1:1:1:1). Four-month-old vegetative-propagated *Cinnamomum verum* (variety *Sri Gemunu*) plants from National Cinnamon Research and Training Centre, Palolpitiya, Sri Lanka, were used as planting material. The experiment included three treatments: a control (C) with only *Gliricidia sepium* mulch, *Gliricidia sepium* mulch inoculated with strain I-I (G I-I), and *Gliricidia sepium* mulch inoculated with strain I-II (G I-II). The strains I–I and I–II were preserved in glycerol stocks at –20°C in the bacterial collection at the Faculty of Agriculture, University of Ruhuna. Bacterial strains were cultured in LB broth (10 g/L tryptone, 10 g/L NaCl, and 5 g/L yeast extract) and incubated overnight on a shaker at 250 rpm. The culture was diluted with sterile distilled water to a final concentration of 10^9 CFU mL^{-1}. These bacterial suspensions were applied to fresh *Gliricidia sepium* mulch (250 g) every 14 days, using 200 mL per pot per application.

Collection of vegetative data

The following vegetative growth parameters were evaluated at the 12th month of plant establishment: Plant height, stem circumference (at 10 cm above ground level), leaf count per plant, internodal length, root length, and chlorophyll content index (CCI) using a CCM-200 Plus Chlorophyll Content Meter (Opti-Sciences).

Leaf essential oil extraction and measuring oil yield

After assessing plant growth, mature, healthy leaves free from disease were collected and air-dried in a shaded environment at room temperature for two weeks. Two replicates of essential oil were extracted through hydro-distillation following a standardized procedure. Dried leaves (80 g) were distilled in 250 mL of distilled water for four hours. The total oil yield was measured and expressed in microliters per gram (µL/g) of dried leaf material.

Soil nutrient analysis

The total nitrogen (N) and available phosphorus (P) of composite samples prepared by soil collected from the different layers of the pot were analyzed, as described by Sparks (1996).

Data analysis

Plant growth parameters, and soil nutrients were analyzed through one-way ANOVA. Treatment means were compared with the Dunnett test at a statistical significance level of $p < 0.05$ using only *Gliricidia sepium* mulch treatment as a reference. IBM SPSS Statistics (Version 25) was used for the analysis.

RESULTS AND DISCUSSION

Vegetative growth performance

The results indicate that both treatments of *Gliricidia sepium* mulch inoculated with bacterial strains I-I and I-II produced significant improvements in several vegetative growth parameters compared to the control treatment, which is only *Gliricidia sepium* mulch (Table 1). G I-II treatments resulted in significantly greater plant height and root length compared to the control. In terms of stem circumference and number of leaves per plant, both G I-I and G I-II exhibited a significant increase over the control. However, no significant differences were found in the internodal length and CCI. Overall, Dunnett's test results demonstrate that the G I-I and G I-II significantly enhanced vegetative growth parameters of *C. verum* compared to the control, with G I-II generally outperforming G I-I, particularly in plant height, leaf number, and root length.

These findings are in accordance with inoculation of these strains to rice, which produced significant results over negative control (Chathurika et al., 2024). The improved growth of cinnamon under bacteria-enriched *Gliricidia sepium* mulch must be linked to the plant growth-promoting traits of strains I-I and I-II, including phosphate solubilization and IAA production (Chathurika et al., 2024; Sandamini et al., 2025). The significant increase in the soil nitrogen nutrient by bacterial strain I-I treatment suggests the efficient decomposition of *Gliricidia sepium* mulch due to the known acid forming ability of I-I (unpublished data). Although strain I-I is not proven for the possession of *Nif* genes based on our unpublished data, it may promote the nitrogen fixing bacteria in the environment by making nutrient available. IAA production is suggested as a strategy attributed to the two strains to enhance root development and nutrient uptake according to previous reports (Çakmakçı et al., 2023; De Souza et al., 2015).

Table 1: Effect of treatments on the vegetative growth of *Cinnamomum verum* at the 12th month of plant establishment.

Treatment	Plant height (cm)	Stem circumference (cm)	Number of leaves per plant	Internodal length (cm)	Chlorophyll content index (CCI)	Root length (cm)
C	68.20 ± 1.32	2.18 ± 0.08	124.60 ± 11.37	4.18 ± 0.14	48.41 ± 4.48	77.00 ± 2.14
G I-I	88.40 ± 6.93	2.84 ± 0.25*	168.00 ± 8.29*	4.28 ± 0.25	52.81 ± 2.44	96.00 ± 6.59
G I-II	90.00 ± 10.00*	2.81 ± 0.12*	190.40 ± 17.77**	4.00 ± 0.12	52.93 ± 1.94	102.40 ± 9.39*

The mean value of five replicates ± SE. Asterisks represent the statistical significance of the difference between the mean values of each treatment and its associated control based on Dunnett's test. The significance levels: $**0.001 < p < 0.01$; $*0.01 < p < 0.05$.

Essential leaf oil yield

There was no significant difference in oil yield extracted from the leaves of *C.verum* among three treatments. The *Gliricidia sepium* mulch was inoculated with strain I-I, yielded 21.34 ± 0.82 µL/g, the *Gliricidia sepium* mulch inoculated with strain I-II, yielded 16.41 ± 1.09 µL/g while control produced 18.75 ± 2.53 µL/g of essential oil. Although essential oil concentration (%) was not significantly affected by either bacterial treatment, the enhanced vegetative biomass observed in inoculated plants suggests a potential increase in total essential oil yield per plant due to greater leaf biomass. PGPB inoculation generally enhances plant growth and secondary metabolite production, although responses may vary depending on bacterial strain and plant species (Banchio *et al.*, 2010; Çakmakçı *et al.*, 2023).

Figure 1: Comparison of vegetative growth parameters of *C. verum* at the 12th month of plant establishment; C: *Gliricidia sepium* mulch, G I-I: *Gliricidia sepium* mulch inoculated with bacterial strain I-I, and G I-II: *Gliricidia sepium* mulch inoculated with bacterial strain I-II (G I-II). (a) Aboveground plant growth, (b) Root system development. Scale bars represent 30 cm.

Available phosphorus and total nitrogen content of soil

Table 2: Effect of treatments on the available phosphorus and total nitrogen in pots after 12 months under greenhouse conditions.

Treatment	Available phosphorus (ppm)	Total nitrogen (%)
C	0.3125 ± 0.0025	0.2832 ± 0.0032
G I-I	0.3617 ± 0.0017**	0.3438 ± 0.0038**
G I-II	0.3443 ± 0.0043*	0.2705 ± 0.0005

Values represent the mean of two replicates ± SE. Asterisks indicate statistical significance of differences between each treatment and the control, as determined by Dunnett's test. Significance levels: **0.001 < p < 0.01; *0.01 < p < 0.05.

Table 2 summarizes the effects of the two bacteria-inoculated mulch treatments (I–I and I–II) compared with the control on soil nutrient status after 12 months under greenhouse conditions. Both inoculated mulch treatments resulted in significantly higher available phosphorus concentrations relative to the control, indicating enhanced phosphorus availability in the soil. In contrast, a significant increase in total nitrogen content was observed only in the G I–I treatment, whereas G I–II did not differ significantly from the control. These findings suggest strain-specific effects on soil nutrient dynamics, with both strains contributing to phosphorus enrichment, while only strain I–I enhanced soil nitrogen levels.

This study explores an underutilized approach in cinnamon cultivation by inoculating nitrogen-rich *Gliricidia sepium* mulch with plant growth-promoting bacteria, rather than applying bacteria directly to soil or foliage. This strategy was associated with improved soil nutrient status and supported vegetative growth, likely reflecting effects of *Bacillus* strains I-I and I-II-mediated nutrient mineralization during mulch decomposition. The nutrient-rich *Gliricidia* biomass provides a substrate that may facilitate these outcomes, consistent with previous reports of enhanced cinnamon growth with mycorrhiza inoculation (Priya *et al*., 2026).

The bacterial strain-enriched mulch demonstrated potential to enhance vegetative growth and soil nutrient content under greenhouse conditions. Future research should validate these findings under field conditions, assess nutrient cycling and microbial community dynamics, examine essential oil composition, and explore practical mulch application strategies, including cost–benefit analyses, before suggesting it as a sustainable agriculture strategy.

REFERENCES

Alamu, E. O., Adesokan, M., Fawole, S., Maziya-Dixon, B., Mehreteab, T., & Chikoye, D. (2023). *Gliricidia sepium* (Jacq.) Walp applications for enhancing soil fertility and crop nutritional qualities: A review. *Forests, 14*(3), 635. https://doi.org/10.3390/f14030635

Banchio, E., Bogino, P. C., Santoro, M., Torres, L., Zygadlo, J., & Giordano, W. (2010). Systemic induction of monoterpene biosynthesis in *Origanum* × *majoricum* by soil bacteria. *Journal of Agricultural and Food Chemistry, 58*(1), 650–654. https://doi.org/10.1021/jf9030629

Çakmakçı, R., Haliloğlu, K., Türkoğlu, A., Özkan, G., Kutlu, M., Varmazyari, A., Molnár, Z., Jamshidi, B., Pour-Aboughadareh, A., & Bocianowski, J. (2023). Effect of different plant growth-promoting rhizobacteria on biological soil properties, growth, yield and quality of oregano (*Origanum onites* L.). *Agronomy, 13*(10), 2511. https://doi.org/10.3390/agronomy13102511

Chathurika, L.G.I., Mandic-Mulec, I., Greenberg, E.P., Senanayake, G., & Geekiyanage, S. (2024). Effect of two bacterial isolates on selected improved rice varieties under *in vitro* and greenhouse condition. *Tropical Agricultural Research and Extension*, **27**(1). https://doi.org/10.4038/tare.v27i1.5699

De Souza, R., Ambrosini, A., & Passaglia, L. M. (2015). Plant growth-promoting bacteria as inoculants in agricultural soils. *Genetics and Molecular Biology, 38*(4), 401–419. https://doi.org/10.1590/S1415-475738420150053

Department of Cinnamon Development. (2024). *Cinnamon handbook*. http://www.cinnamon.gov.lk

Ebrahimi, Z., Fateh, E., Soorestani, M. M., & Ghorbanpour, M. (2023). Inoculating mycorrhiza fungi and growth-promoting bacteria and mulch of plant residues improve yield and essential oil production of anise (*Pimpinella anisum* L.). *Acta Scientiarum Polonorum, Hortorum Cultus, 22*(6). https://doi.org/10.24326/asphc.2023.5162

Priya, R. S., Prem, J. J., Gailce, J. L. C., & Jayasekhar, M. (2026). Effect of biofertilizers on the growth and yield of cinnamon (*Cinnamomum zeylanicum* Blume). *Indian Journal of Agricultural Research, 41*(4), 310–312.

Rossi, G., Beni, C., & Neri, U. (2024). Organic mulching: A sustainable technique to improve soil quality. *Sustainability, 16*(23), 10261. https://doi.org/10.3390/su162310261

Sandamini, P. K. D., Mandić Mulec, I., Grabner, E., Kumara, H. W. K. S., Jayaprada, N. V. T., & Geekiyanage, S. (2025). Enhancing vegetative growth and nutrient content of *Gliricidia sepium* (Jacq.) Kunth ex Walp using two Sri Lankan *Bacillus* species. *Vegetos*, https://doi.org/10.1007/s42535-025-01507-4

Sparks, D. L. (Ed.). (1996). *Methods of soil analysis: Part 3 – Chemical methods*. Soil Science Society of America. https://doi.org/10.2136/sssabookser5.3

Sri Lanka Export Development Board. (2023). *Export Performance Indicators 2023*. https://www.srilankabusiness.com/ebooks/export-performance-indicators-of-sri-lanka-2023.pdf

JOURNAL OF THE
NATIONAL SCIENCE FOUNDATION
OF SRI LANKA

GUIDANCE TO CONTRIBUTORS

GENERAL INFORMATION

Scope

The Journal of the National Science Foundation of Sri Lanka publishes the results of research in all aspects of Science and Technology. It is open for publication of Research Articles, Reviews, Research Communications and Correspondence.

Use of AI Tools in Manuscript Preparation

To ensure transparency and uphold scholarly standards, authors must disclose any use of Artificial Intelligence (AI) tools during the research and manuscript preparation process.
The use of AI tools (e.g., for language editing, summarization, data analysis, or figure generation) **must be declared in the manuscript, typically before the Acknowledgements**.
AI tools must not be credited as authors. Human authors remain fully responsible for the integrity, accuracy, and originality of the submitted work.
AI-generated content must be critically reviewed by the authors to ensure quality and ethical compliance, including data privacy and bias mitigation.
AI must not be used to fabricate or manipulate original research data, unless it is part of the approved research methodology and explicitly described.
Non-compliance with these guidelines may lead to rejection of the manuscript or post-publication corrective actions.

IT related and other non-empirical articles

The JNSF is a journal primarily devoted to natural sciences. It also considers for publication significant and novel contributions from formal sciences. Authors of emerging sub-disciplines of Computing and related areas such as Machine Learning, Artificial Intelligence and Data Sciences are requested to carefully adhere to the following guidelines when submitting manuscripts for this journal.

- Clear formulation of outcome-oriented **Research Objective/s** for targeted knowledge (sub)domain/s or (sub)discipline/s.
- Selection and comprehensive summarization of **appropriate Research Method/s** adopted to achieve the stated Research Objective/s.
- Reporting a sound **(Empirical) Evaluation** of the research finding/s thereby arguing reliability, validity, and generalizability of research claim/s.

Categories of manuscripts

Research Articles: Research Articles are papers that present complete descriptions of original research. Research Articles should include an Abstract, Keywords, Introduction, Methodology, Results and Discussion, Conclusion and Recommendations where relevant. References should be prepared according to the "Guidelines for the preparation of manuscripts". Maximum length of the article should be limited to 30 pages (Times New Roman, font size 12 , line space 1.5) with a word count less than 10,000 including references, figures and tables. Any articles above this limit will be returned. Please refer author guidelines for further details (https://jnsfsl.sljol.info/about/submissions#author-guidelines)

Reviews: Reviews are critical presentations on selected topics of Science or Technology. They should be well focused and organized and avoid general "textbook" style. As reviews are intended to be critical presentations on selected topics, the author (or the principal author in a multi-author review) need to have had substantial leadership in research supported by a publication track record in the areas covered by the review. A person/s wishing to submit a Review Article should obtain prior approval from the Editorial Board by submitting a concise summary of the intended article, along with a list of the author's publications in the related area (jnsf@nsf.gov.lk). Maximum length of the article should be limited to 40 pages (Times New Roman, font size 12 , line space 1.5) with a word count of 12,000 including references, figures and tables. Any articles beyond this limit will be returned.

Research Communications: Research Communications are intended to communicate important new findings in a specific area of limited scope that are worthy of rapid dissemination among the scientific community. Authors are required to provide a statement justifying the suitability of the submission for a Research Communication. The article should include an Abstract, Keywords, Introduction, Methodology, Results & Discussion, Conclusion and References. Maximum length of the article should be limited to 10 pages (Times New Roman, font size 12 , line space 1.5) with a word count of 2,500 including references, figures and tables. Any articles beyond this limit will be returned.

Correspondence: Correspondence will be accepted regarding one or more articles in the preceding four issues of the Journal, as well as Letters to the Editor. Articles covering important scientific events or any other news of interest to scientists, reviews of books of scientific nature, articles presenting views on issues related to science and scientific activity will also be considered. Publication will be made at the discretion of the Editor-in Chief. Maximum length of the article should be limited to 05 pages (Times New Roman, font size 12 , line space 1.5) with a word count of 1,500 including references, figures and tables. Any articles beyond this limit will be returned.

SUBMISSION OF MANUSCRIPT

Authors submitting articles to the JNSF should first create an account in the Sri Lanka Journals Online System (https://jnsfsl.sljol.info/). All manuscripts should be in MS Word format and must be submitted to the journal's online platform at https://jnsfsl.sljol.info/submit/start/. Submissions *via* emails are not encouraged. Please make sure that **no** author information is mentioned in the article submitted other than in the title page. Complete names and details of affiliations of all authors must be fed into the system during the online submission process. Authors are required to provide their personal, validated ORCID ID (by obtaining an ORCID ID from https:// orcid.org/) when submitting the manuscript. No change to the

authors or order of authors will be accepted after the submission. All those who have made significant contributions should be listed as co-authors. The corresponding author should ensure that all contributing co-authors are included in the author list and have approved the final version of the paper and have agreed to its submission for publication.

All submissions should be in English. If the manuscript conforms to the guidelines specified, the date received will be the date that the manuscript was submitted to the online system.

Submissions are accepted for processing on the understanding that they will be reviewed and that they have not been submitted for publication elsewhere (including publication as a full paper or extended abstract as a part of Conference Proceedings).

Suggesting potential reviewers by authors
The authors are requested to suggest three names of referees when submitting their manuscript, in the Cover Letter space provided at the bottom of the page in the first stage of online submission. Referees should not be from the institution where the work was carried out and should not have been co-authors in previous publications. The address, institutional affiliation and e-mail of the suggested referees should be provided. Please note that the JNSF is not bound to select all or any of the suggested referees for sending the manuscript for reviewing

Authorship
All authors designated as authors should be eligible for authorship. Those who have made a substantial contribution to the concept or design of the work; or acquisition, analysis or interpretation of data are recognized as Authors.
The corresponding author should be prompt and ensure adherence to timelines when responding to requests, queries and recommendation of reviewers conveyed by or on behalf of the Editor-in Chief and Editorial Board.

Avoiding predatory references in manuscripts

Authors of the manuscripts should identify and not include predatory journal articles in the reference list.
If any paper that is cited is published in an Open Access journal, the authors are advised to ensure that the journal is listed in the DOAJ (Directory of Open Access Journals) or by the COPE (Committee on Publication Ethics). Please note that authors should check with DOAJ and COPE by themselves, and not assume that the statements given in the journal websites are accurate.
If the journal is included either in DOAJ or COPE, it can be presumed as legitimate. If not, authors need to review the journal website and published articles for characteristics of predatory journals. At the same time, authors can check whether the journal or relevant publisher is included in the directory of predatory journals and publishers.
If it is revealed that the manuscript is supported by articles published in predatory journals, it could be rejected by the JNSF at any stage of the publication process.

Supplementary materials
Any experimental data necessary to evaluate the claims made in the paper but not included in the paper should be provided as supplementary materials. Supplementary materials will be sent to the reviewers and published online with the manuscript if accepted. The supplementary materials should conform to Journal guidelines and should be uploaded as separate files. Authors should number Supplementary Tables and Figures as, for example, 'Supplementary Table S1'. Refer to each piece of supplementary material at the appropriate point(s) in the main article. Supplementary Materials may include description of the materials and methods, controls, or tabulated data presented in Tables or Figures, and programming codes.

Peer review
The manuscripts submitted to the JNSF will initially be screened by the Editorial Board and, if suitable, will be referred to at least two subject experts in the relevant field. The peer-review process of the JNSF is double-blind.
When revision of a manuscript has been requested, the revised manuscript should be submitted on or before the stated deadline. The authors' response to the comments of referees should be tabulated with the comment and response. The decision of the Editorial Board shall be final.
Accepted papers are subject to editing. The date of acceptance will be the date the Editorial Board accept the paper for publication.

Article processing fee
Article processing fee of US$ 250 will be levied for each manuscript in two stages, except when the corresponding author is affiliated with a Sri Lankan institution,

- An initial processing fee of US$ 20 will be levied for each manuscript at the peer-review stage.
- The remaining US$ 230 will be charged for accepted articles at the time of publication.

Payments can be made online via NSF Payment Portal (http://pg.nsf.gov.lk/)

Authors' declaration
The authors are required to accept the conditions indicated in the online author declaration statement.

Copyright

Post-publication corrections
The Editorial Board reserves the right to take action on publishing an erratum or corrigendum. If serious errors are identified in a published article, the Journal may consider a retraction or publishing a correction.

STRUCTURE OF MANUSCRIPT

Manuscript
The manuscript should be free of errors and prepared in single column, using 1.5 line spaced text of Times New Roman 12 font leaving 1 inch margins. Pages should be numbered consecutively.

a. Style
The paper should be written clearly and concisely. The style of writing should conform to scholarly writing. Slang, jargon, unauthorized abbreviations, abbreviated phrasings should not be used. In general, the impersonal form should be used. Poor usage of language will result in rejection of the manuscript during initial screening.

b. Layout
Manuscripts other than review articles should be generally organized as follows: Title, Abstract, Keywords, Introduction, Methodology, Results and Discussion, Conclusions and Recommendations (where relevant), Acknowledgements and References. Pages should be arranged in the following order:

First page should include the title of the manuscript and **complete names and affiliations** of all authors. **Author information should not be mentioned in other places of the manuscripts.** Any statement (including acknowledgment) which can reveal author identity should be removed. If a major part of the research has been published as an abstract in conference proceedings, it should be mentioned with citation in the space provided for "Comments for Editor". Authors must also indicate the **general and specific research area** of the manuscript in the title page.

***Title*:** Should accurately and concisely reflect the contents of the article.

***Running title*:** Should be a shortened title (limited to a maximum of 50 characters) that could be printed at the top of every other page of the Journal article.

***Abstract*:** Should be between 200 - 250 for research articles and 200 - 300 for reviews. It should not contain any references and should be able to stand on its own. It should outline objectives and methodology together with important results and conclusions.

***Keywords*:** Include a maximum of six keywords, which may include the names of organisms (common or scientific), methods or other important words or phrases specific to the study.

***Introduction*:** This should state the reasons for performing the work with a brief review of related research studies in the context of the work described in the paper. Objectives of the study should be clearly stated.

***Materials and Methods*:** This section should give the details of how you conducted your study. New methods may be described in detail with an indication of their limitations. Established methods can be mentioned with appropriate references. Sufficient details should be included to allow direct repetition of the work by others. Where human subjects are involved, they should be referred to by numbers or fictitious names. A paper reporting the results of investigations on human subjects or on animals must include a statement to the effect that the relevant national or other administrative and ethical guidelines have been adhered to, and a copy of the ethical clearance certificate should be submitted. Methods of statistical analyses used should be mentioned where relevant.

Results and Discussion
Results: the results should be concisely and logically presented. Repetition of the same results in figures, tables or text should be avoided.
Discussion: data essential for the conclusions emerging from the study should be discussed. Long, rambling discussions should be avoided. The discussion should deal with the interpretation of results. It should logically relate new findings to earlier ones. Unqualified statements and conclusions not completely supported by data should be avoided.
Molecular sequence data, such as gene or rDNA sequences, genome sequences, metagenomic sequences etc. must be deposited in a public molecular sequence repository, such as GenBank, that is part of the International Nucleotide Sequence Database Collaboration (INSDC). The accession numbers obtained must be cited in the text, Table or on Figures of phylogenetic trees of the manuscript.

***Conclusion*:** The conclusion should be brief, highlight the outcomes of the study and should be aligned with the objectives of the study. It should not contain references.

Competing Interest statement: The authors should include a statement on conflict of interest disclosing any financial or other substantive conflicts of interest that may influence the results or interpretation of the research in the space provided in the online article submission form. All sources of financial support for the project should also be disclosed.

***Acknowledgement*:** Should be brief and made for specific scientific, financial and technical assistance only. If a significant part of the research was performed in an institution other than the authors' affiliations should be acknowledged. All those who have made substantial contribution to the research but do not qualify to be authors should be acknowledged.

***References* :**
The JNSF uses APA (7th Edition) reference style
All research work of other authors, when used or referred to or cited, should be correctly acknowledged in the text and in the References.
All the references in the text should be in the list and vice versa

Citing references in the text:
- References to the literature must be indicated in the text and tables as per the Author-Year System, by the author's last name and year, in parenthesis (i.e. Able, 1997) or (Able & Thompson, 1998).
- Citation to work by more than two authors should be abbreviated with the use of et al. (i.e. Able *et al*., 1997).
- Multiple publications by the same first author in the same year should be coded by letters, (i.e. Thompson, 1991a, 1991b,1992,1993).
- Multiple citations of different authors should be made in chronological order and separated by a semicolon, (i.e. Zimmerman *et al*., 1986; Able *et al*., 1997).

Citing references in the List of references:
- The list of References should be arranged in alphabetical order based on the last name of the first author.
- In APA 7th ed., **up to 20 authors** should be included in a reference list entry. Write out the last name and first initial(s) for each contributor.

Example for 2–20 authors:

Wright, A., Komal, G., Siddharth, D., Boyd, G., Cayson, N., Beverley, K., Travers, K., Begum, A., Redmond, M., Mills, M., Cherry, D., Finley, B., Fox, M., Ferry, F., Almond, B., Howell, E., Gould, T., Berger, B., Bostock, T., & Fountain, A. (2020). Styling royalty. London Bridge Press.

- For references with more than 20 authors, after listing the 19th author replace any additional author names with an ellipsis (…) followed by the final listed author's last name and first initial(s).

Example for 21+ authors:

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woolen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., … Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society, 77(3), 437-471. http://doi.org/fg6rf9

- All the initials of the author must be given after the last name and the year of publication should follow in parentheses.
- This should be followed by the full title of the referred publication.
- When journal articles are listed, the journal name should be given in full and in italics and followed by the volume number, issue number in parentheses and then the inclusive pages.
- Where there are several publications by the same author(s) and published in the same year they should be differentiated by adding a lower-case letter after the year.

Example

Clarke, P. N., & Fawcett, J. (2014a). Life as a mentor. *Nursing Science Quarterly*, 27(3), 213-215. https://doi.org/10.1177/0894318414534492

Clarke, P. N., & Fawcett, J. (2014b). Life as a nurse researcher. *Nursing Science Quarterly*, 27(1), 37-41. https://doi.org/10.1177/0894318413509708

- Digital object identifiers (DOIs) should be included for all references where available.

Details about this reference style can be obtained from below links

- https://apastyle.apa.org/style-grammar-guidelines/references
- https://apastyle.apa.org/style-grammar-guidelines/references/examples
- https://libguides.jcu.edu.au/apa

Abbreviations and Symbols: Unless common, these should be defined when first used, and not included in the abstract. The SI System of units should be used wherever possible. lf measurements were made in units other than SI, the data should be reported in the same units followed by SI units in brackets, e.g. 5290 ft (1610 m).

Formulae and Equations: Equations should be typewritten and quadruple spaced. They should be started on the left margin and the number placed in parentheses to the right of the equation.

Nomenclature: Scientific names of plants and animals should be printed in italics. In the first citation, genus, species and authority must be given. e.g. *Borassus flabellifer* Linn. In latter citations, the generic name may be abbreviated, for example, *B. flabellifer* L.

Tables and figures: Tables and Figures should be clear and intelligible and kept to a minimum, and should not repeat data available elsewhere in the paper. Any reproduction of illustrations, tabulations, pictures etc. in the manuscript should be acknowledged.

Tables: Tables should be numbered consecutively with Arabic numerals and placed at the appropriate position in the manuscript. If a Table must be continued, a second sheet should be used and all the headings repeated. The number of columns or rows in each Table should be minimized. Each Table should have a title, which makes its general meaning clear, without reference to the text. All Table columns should have explanatory headings. Units of measurement, if any, should be indicated in parentheses in the heading of each column. Vertical lines should not be used and horizontal lines should be used only in the heading and at the bottom of the table. Footnotes to Tables should be placed directly below the Table and should be indicated by superscript lower case italic letters (^{a}, ^{b}, ^{c}, etc.).

Figures: All illustrations are considered as figures, and each graph, drawing or photograph should be numbered consecutively with Arabic numerals and placed at the appropriate position in the manuscript. Any lettering to appear on the illustrations should be of a suitable size for reproduction and uniform lettering should be used in all the Figures of the manuscript. Scanned figures or photographs should be of high quality (**300 dpi**), to fit the proportions of the printed page (12 × 17 cm). Each figure should carry a legend so that the general meaning of the figure can be understood without reference to the text. Where magnifications are used, they should be stated.

Units of measurement

Length: km, m, mm, µm, nm
Area: ha, km^2, m^2
Capacity: kL, L, mL, µL
Volume: km^3, m^3, cm^3
Mass: t, kg, g, mg, µg
Time: year(s), month(s), week(s), day(s), hour(s), minute(s), second(s)
Concentration: M, mM, N, %, g/L, mg/L, ppm
Temperature: °C, K
Gravity: x g
Molecular weight: mol wt
Others: Radio-isotopes: 32P
Radiation dose: Bq
Oxidation-reduction potential: rH
Hydrogen ion concentration: pH