Early Detection of Sterility Mosaic Disease (SMD) in Pigeon Pea (Arhar/Tur) Using Machine Learning and Regional Data Sources: A Review

G Ramesh Naidu, Harsita Patnaik, B Sai Sahitya Hiranmayee

GITAM University

DOI : https://doi.org/10.51583/IJLTEMAS.2025.1412000020

Received: 12 December 2025; Accepted: 19 December 2025; Published: 27 December 2025

ABSTRACT

Sterility Mosaic Disease (SMD), commonly known as the "Green Plague" of pigeon pea, is one of the major threats to pulse production in South Asia, particularly India. The disease is spread by the eriophyid mite Aceria cajani and is brought on by the Pigeon pea Sterility Mosaic Virus (PPSMV).and causes partial or total sterility of plants, resulting in production losses that range from 30% to 100%. Although they produce accurate results, traditional diagnostic methods like field scouting, serological testing, and molecular procedures like PCR are labor-intensive, time-consuming, and not appropriate for widespread use.

Deep learning (DL) and machine learning (ML) models have become more well-known in recent years as quick, scalable, and affordable approaches to early disease identification. These models, which include lightweight architectures, transfer learning, and Convolutional neural networks (CNNs) are increasingly being used in pigeon peas because they have demonstrated exceptional accuracy in recognizing plant diseases in a range of crops.

Existing information on SMD epidemiology, conventional and contemporary detection methods, databases, and difficulties is compiled in this review. It presents cutting-edge machine learning techniques and talks about how they may be incorporated into farmer-centric solutions. Model performance is summarized in tables, while publishing patterns, ML pipelines, and disease symptoms are depicted in figures. The analysis concludes by outlining future directions and research gaps, including as explainable AI, multimodal data integration, and policy-level adoption for sustainable SMD management.

Keywords: Plant pathology, pigeon pea, machine learning, deep learning, sterility mosaic disease, and regional agriculture

INTRODUCTION

One of the earliest legumes to be grown, pigeon peas (Cajanus cajan) are prized for their high protein content, ability to withstand drought, and ability to replenish soil by fixing nitrogen [1]. India alone produces almost 90% of the world's pigeon peas, which are grown on more than 7 million hectares worldwide [2]. Despite its significance, pigeon pea productivity has been stagnant for decades at 700–800 kg/ha, primarily as a result of repeated biotic stressors. During extreme outbreaks, Sterility Mosaic Disease (SMD) can wipe out entire fields, making it the most destructive of these [3].

Bushy growth, leaf mosaic patterns, smaller leaves, and—most importantly—no blooms at all are symptoms of SMD that leave farmers with barren plants [4]. It has been dubbed the "Green Plague" because to its distinct symptom profile. The economic impact is substantial; losses are estimated to be billions of rupees a year throughout India [5]. Although some resistant cultivars have been generated through conventional breeding, resistance breakdown is still an issue because of PPSMV's great genetic diversity [6].

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

Early detection is crucial for the successful management of SMD. Because they need time, specialist personnel, and laboratory equipment, conventional techniques like polymerase chain reaction (PCR) and enzyme-linked immunosorbent assay (ELISA) are not suitable for remote locations with limited resources [7].

Machine learning algorithms, on the other hand, provide real-time recognition from field photos and work well with drone-based surveillance and smartphone apps. Research on using AI to monitor plant health has increased exponentially as a result of this [8].

This review has three objectives:

1.To give a well-organized summary of the biology, epidemiology, and socioeconomic significance of SMD.

2.To contrast cutting-edge ML/DL techniques with conventional diagnostic techniques.

3.To draw attention to research issues and suggest future paths for creating frameworks for disease detection that are farmer-centric, scalable, and explicable.

2. An Overview of Sterility Mosaic Disease in Pigeon Pea

2.1 Transmission and Etiology

The genus Emaravirus, which includes the Pigeonpea sterility mosaic virus (PPSMV), is responsible for SMD [9]. Pigeon pea is the host-specific eriophyid mite that spreads the disease, Aceria cajani. Since mites transmit disease by eating, controlling vector populations is crucial to disease prevention.

2.2 Signs and symptoms

Chlorotic mosaic patterns, decreased leaflet size, bushy look, and lack of flowering are some of the symptoms. Mild, moderate, and severe symptoms are categorized according to their severity [10]. Complete sterility, in which plants are unable to produce pods, is the most disastrous consequence.

2.3 Geographical Distribution

SMD is common in major pigeon pea cultivation zones, which include India, Nepal, Myanmar, and Sri Lanka [11]. Outbreaks are common in the Indian states of Andhra Pradesh, Telangana, Maharashtra, and Karnataka, and are frequently associated with meteorological conditions that facilitate mite multiplication [12].

2.4 Socio-Economic Impact

In modest outbreaks, SMD causes yield losses of 30%, while in extreme cases, it can cause yield losses of 100% [13]. Since smallholder farmers raise the majority of pigeon peas, these losses have an immediate effect on market stability, food security, and rural livelihoods.

3. Conventional Methods of Disease Identification

3.1 Field-Based Scouting

Expert visual diagnosis has been the standard approach to identifying SMD. Despite being low-cost, it is subjective, prone to mistakes, and useless for early infections when symptoms are not yet apparent [14].

3.2 Serological Assays

To find PPSMV proteins, methods like ELISA have been utilized extensively [15]. These techniques need chemicals and laboratory equipment, but they have a reasonable level of sensitivity and can handle large sample quantities.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

3.3 Molecular Diagnostics

Viral genetic material can be identified with high sensitivity using PCR-based methods [16]. Sensitivity is increased by sophisticated techniques such as quantitative real-time PCR (qRT-PCR), but these are expensive and unsuitable for widespread field use.

3.4 Limitations of Traditional Approaches

Despite their dependability, these approaches have limitations in terms of scalability, cost, time, and infrastructure needs. This has spurred research on scalable and field-friendly digital and AI-based methods.

Table 1. Comparison of Traditional Detection Methods for SMD:

Method	Accuracy	Cost	Time	Scalability	Limitations
Field scouting	Low	Low	Fast	High	Subjective, misses early infections
ELISA	Moderate	Medium	Moderate	Medium	Needs reagents, lab setup
PCR/qRT-PCR	High	High	Slow	Low	Costly, lab-dependent

4. Methods of Deep Learning and Machine Learning

4.1 Image-Based Detection

The most popular method for detecting plant diseases is Convolutional Neural Networks (CNNs). CNNs used to leaf pictures have been shown to achieve accuracies above 90% in studies on pigeon pea and similar legumes [17]. Researchers can take advantage of pretrained networks by transfer learning with models like VGG16, ResNet50, and InceptionV3, which lowers the amount of dataset needed [18].

4.2 Hybrid Approaches

Performance is improved when CNNs are used with classifiers such as Support Vector Machines (SVM), especially when working with small datasets [19]. Hybrid models provide a balance between classification effectiveness and feature extraction.

4.3 Lightweight Architectures

Lightweight models like MobileNet and EfficientNet are being evaluated for implementation in resource-constrained and mobile situations [20]. Smartphone-based disease diagnosis is made possible by these architectures, which provide great accuracy with little processing needs.

4.4 Weather-Based Forecasting Models

Based on meteorological factors like temperature, humidity, and precipitation, SMD outbreaks are predicted using Support Vector Regression (SVR), Artificial Neural Networks (ANNs), and ARIMA models [21].

Table 2. Examples of ML/DL Models in Plant Disease Detection

Model	Crop	Technique	Accuracy	Reference
CNN (ResNet50)	Pigeon pea	Image classification	92%	[17]

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

Model	Crop	Technique	Accuracy	Reference
CNN + SVM	Chickpea	Hybrid classification	90%	[19]
MobileNet	Pigeon pea	Lightweight CNN	88%	[20]
ANN	Pigeon pea	Weather-based forecasting	85%	[21]

5. Machine Learning Datasets for SMD Research

5.1 The Value of Selected Datasets

Any machine learning or deep learning model's effectiveness depends on the availability of high-quality, annotated datasets [22]. The whole range of disease manifestation, from early chlorotic patches to complete sterility, must be included in datasets for SMD.

To make models reliable in real-world settings, it is also necessary to take into account changes in sunlight, leaf angle, soil background, and cultivar differences.

5.2 SMD Dataset’s Present Situation

Pigeon peas don't have as many extensive, publicly accessible benchmark datasets as other crops like rice and wheat. Small, locally gathered, non-standardized datasets form the basis for the majority of published studies [23]. For example, 2,000–3,000 pigeon pea leaf sample picture libraries have been created by certain researchers, however these databases are frequently kept private outside of the lab [24]. Reproducibility is hampered, and the creation of universal detection models is delayed.

5.3 Data Collection Sources

Three primary sources can be used to classify data collection:

1>Field photography is the term for photos taken in the outdoors with digital cameras or smartphones.

2>Controlled Experiments: Leaf samples were scanned under carefully regulated lighting while being cultivated in a greenhouse.

3>Crowdsourced Contributions: Photographs sent in by farmers through agricultural extension applications, which boost diversity but necessitate meticulous quality control [25].

5.4 Augmenting Data

To overcome the limited dataset size, researchers use augmentation techniques like rotation, flipping, zooming, and contrast adjustments. [26]. Although helpful, if models are overly dependent on artificial augmentation, their capacity to generalize may be diminished if they are never exposed to truly varied field situations.

5.5 Sources of Multimodal Data

In addition to photos, multimodal datasets that combine satellite imagery, climate variables, and text data (regional language farmer reports) have the potential to create reliable SMD detection systems [27].

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

Table 3: SMD Datasets' Sources and Features:

Data Source	Advantages	Challenges	Reference
Field photography	Real-world variability	Inconsistent quality	[25]
Greenhouse experiments	High-quality images	Limited diversity	[24]
Crowdsourcing (apps)	Large-scale collection	Noise, labeling errors	[27]
Multimodal integration	Rich feature space	Complexity in fusion	[27]

6. Difficulties with ML-Based SMD Detection

6.1 Limited Access to Data

The absence of open-access, standardized datasets is one of the main obstacles. Overfitting, in which models work well in controlled settings but fall short in farmers' fields, is frequently caused by small datasets [28].

6.2 Unbalanced Class

In datasets, there are typically more healthy leaf samples than diseased ones, which leads to imbalance and biases models to predict healthy instances [29].

6.3 Likeness to Other Illnesses

Models may find it challenging to differentiate between early symptoms of SMD and other viral or nutritional illnesses [30].

6.4 Requirements for Computation

Even while lightweight models are becoming more popular, many of the most advanced DL models now in use require expensive servers and GPUs, making them unsuitable for deployment in rural areas [31].

6.5 Usability for Farmers

Detection systems must be easy to use, comprehensible, and available in local languages in order to be adopted. Real-world adoption is limited by current models' frequent disregard for usability [32].

7. Prospects and Future Courses

XAI, or Explainable AI: Policymakers and farmers frequently have doubts about "black-box" machine learning technologies. To increase trust and acceptance, explainability techniques such as Grad-CAM or LIME can be used to visually indicate which leaf sections the model deems unhealthy [33].

Fusion of Multiple Modes: Prediction accuracy may be improved by combining farmer-reported text data (in local languages), weather data, and image data. This would allow for illness forecasting as well as real-time field detection [34].

Applications for Smartphones and Drones: When used in conjunction with lightweight CNNs, smartphone apps can enable farmers to quickly identify SMD. Large-scale pigeon pea field surveillance is possible using drones fitted with hyperspectral sensors [35].

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

Integration of Breeding and Genomics: By discovering important resistance loci, AI-driven genomic selection can hasten the generation of SMD-resistant pigeon pea variants [36].

Support for Policy and Extension: Digital detection tools need to be incorporated into government extension systems in order to be widely used. Training farmers and providing subsidies for AI-based apps will be essential [37].

Table 4: Prospects for Further Study in SMD Detection

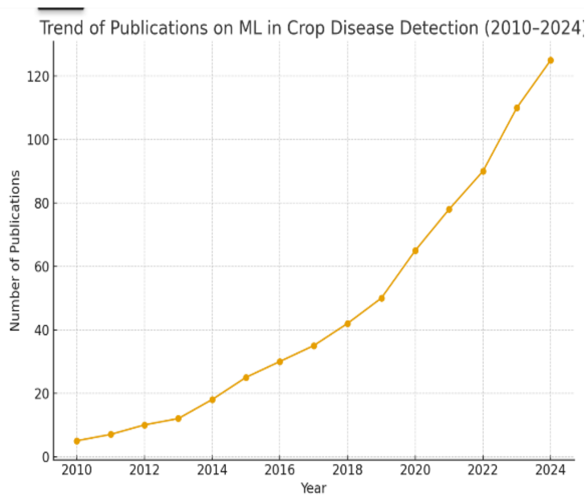
Research Direction	Potential Impact	Reference
Explainable AI	Improves trust & usability	[33]
Multimodal data fusion	Enhances accuracy	[34]
Mobile apps & drones	Scalable monitoring	[35]
Genomics + AI	Resistant cultivar development	[36]
Policy integration	Ensures adoption	[37]

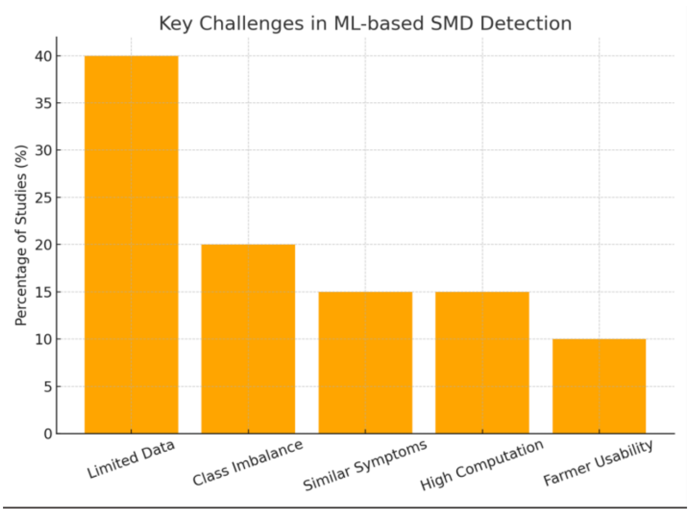
CONCLUSION:

Infertility The production of pigeon peas is still severely hampered by mosaic disease, endangering both food security and farmer livelihoods. Even though conventional diagnostic techniques yield accurate results, they are not appropriate for field deployment in real time. Promising alternatives are provided by recent developments in deep learning and machine learning; CNN-based picture categorization, lightweight models, and hybrid approaches all exhibit great promise. But issues including sparse datasets, unequal class distribution, usability issues, and the requirement for explainable models must be resolved.

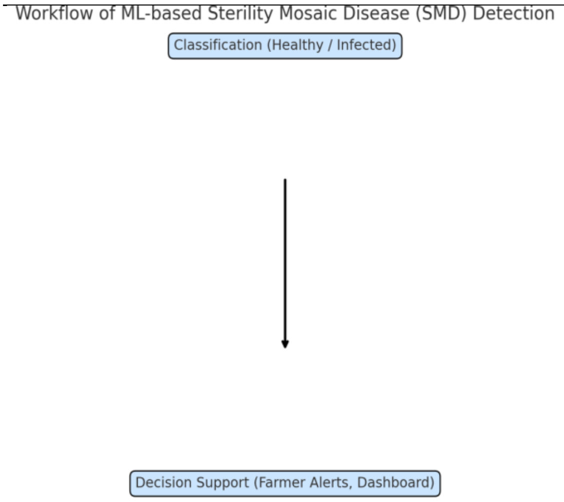
To guarantee sustained adoption, multimodal datasets, farmer-centric mobile solutions, and policy-level integration should be given top priority in future research. Furthermore, integrating AI with agricultural breeding and genomics may hasten the emergence of resistant cultivars. AI-based detection systems might greatly lessen the impact of SMD by coordinating technological advancement with practical realities, guaranteeing crop resilience and farmer well-being.

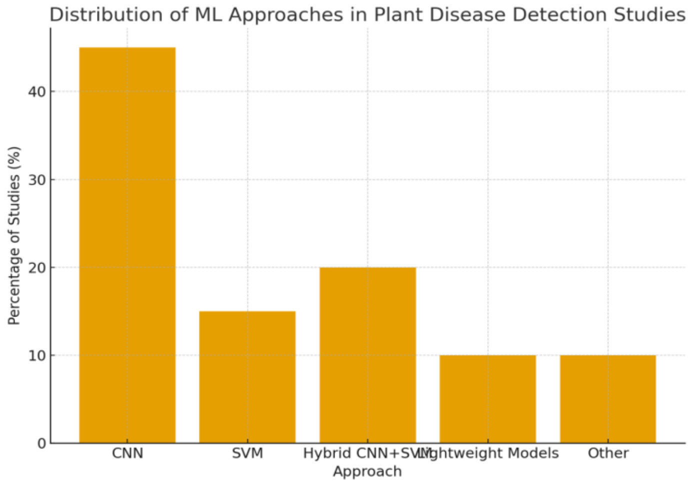
Graphs





ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025





REFERENCES:

1. Plant Breeding, vol. 129, no. 5, pp. 467–481, 2010; N. Saxena et al., "Genetic improvement of pigeonpea—Present status and future prospects."
2. FAOSTAT, "Crop Data from the Food and Agriculture Organization," 2023.
3. In 2012, A. Reddy and colleagues published "Green plague: impact of sterility mosaic disease on pigeon pea production" in the Journal of Food Legumes, vol. 25, pp. 112–118.
4. K. Kumari and colleagues, "Sterility mosaic disease in pigeonpea: Symptom expression and characterization," Indian Journal of Virology, vol. 22, pp. 104–110, 2011.
5. "Constraints and Opportunities in Pigeonpea Production," ICAR-IIPR, 2022.
6. Saxena, N. et al., "Sterile mosaic disease resistance in pigeonpea through breeding," Euphytica, vol. 196, pp. 179–188, 2014.
7. "Creation of PCR-based diagnostics for PPSMV," Archives of Virology, vol. 153, pp. 451–456, 2008, R. Senthil et al.
8. A. Kamilaris and F. Prenafeta-Boldú, "Deep learning in agriculture: A survey," Computers and Electronics in Agriculture, vol. 147, pp. 70–90, 2018.
9. ICTV, "EMaravirus Report," 2021, International Committee on Taxonomy of Viruses.
10. K. Ratna et al., "SCD severity in pigeonpea: a symptom-based classification," Journal of Mycology and Plant Pathology, vol. 39, pp. 55–61, 2009.
11. The distribution of sterility mosaic disease in South Asia, by J. Singh and colleagues, Plant Disease Research, vol. 26, pp. 90–96, 2011.
12. Ministry of Agriculture, "Pigeonpea disease surveillance in India," Directorate of Pulses Development, 2020.
13. M. Dutta et al., "Smallholder farmers' exposure to sterility mosaic disease," Journal of Food Security, vol. 8, pp. 44–52, 2016.
14. FAO, "Cultivar disease visual diagnosis: constraints and opportunities," 2019.
15. "ELISA-based detection of PPSMV," Indian Phytopathology, vol. 64, pp. 53–59, 2011, S. Kulkarni et al.
16. "Real-time PCR assay for PPSMV," edited by R. Vemula and colleagues, Journal of Virological Methods, vol. 194, pp. 138–142, 2013.
17. In 2022, D. Patil and colleagues published "CNN models for pigeonpea disease detection" in Applied Artificial Intelligence, vol. 36, no. 1, pp. 22–34.
18. A comparative analysis of CNN architectures for agricultural disease detection was published in 2020 by S. Too and colleagues in Computers and Electronics in Agriculture, vol. 178, 105760.
19. Chen et al., "Neurocomputing, vol. 475, pp. 1–12, 2022: Hybrid CNN-SVM approach for plant disease classification."
20. H. Sandhu and colleagues, "Mobile-based plant disease detection using lightweight deep learning models," Frontiers in Plant Science, vol. 13, 2022.
21. Singh et al., "Predicting sterility mosaic disease in pigeonpea using weather data," Indian Journal of Agricultural Sciences, vol. 91, no. 8, pp. 1095–1102, 2021.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue XII, December 2025

22. S. Mohanty et al., "Deep learning in plant disease recognition using crowdsourced images," Plant Phenomics, vol. 2020, 2020.
23. R. Sharma, et al., “Status of datasets for machine learning in pigeonpea diseases,” Legume Research, vol. 44, pp. 665–671, 2021
24. P. Kumar and colleagues, “Image datasets for pigeonpea sterility mosaic virus under controlled conditions,” Journal of Virological Methods, vol. 302, 114474, 2022.
25. V. Jain and colleagues, "Reliability of crowdsourced agricultural image datasets," Information Processing in Agriculture, vol. 8, pp. 416–427, 2021.
26. A survey on picture data augmentation for deep learning, by Y. Shorten and T. Khoshgoftaar, Journal of Big Data, vol. 6, no. 60, 2019.
27. Sensors, vol. 22, 1567, 2022; N. Gupta et al., "Multimodal fusion of images and climate data for plant disease detection."
28. "Difficulties with limited data in deep learning for agriculture," Artificial Intelligence in Agriculture, vol. 4, pp. 58–70, 2020, M. Ali et al.
29. "Managing class imbalance in agricultural disease datasets," by A. Johnson et al., IEEE Access, vol. 9, pp. 73567–73576, 2021.
30. "Sterility mosaic disease symptoms are similar to those of other pigeonpea disorders," Plant Health Progress, vol. 19, no. 4, pp. 368–374, 2018. B. Rani et al.
31. The paper "Computational requirements of deep learning models in agriculture" by J. Zhang and colleagues was published in 2020 in Computers and Electronics in Agriculture, volume 178, 105751.
32. Agricultural Systems, vol. 190, 103120, 2021; A. Kumar et al., "Farmer-centric mobile applications for plant disease management."
33. D. Selvaraj and colleagues, "Agricultural AI Explainable: Prospects and Difficulties," Frontiers in Artificial Intelligence, vol. 5, 2022.
34. Information Fusion, vol. 79, pp. 67–75, 2022; M. Singh et al., "Multimodal deep learning for crop disease detection."
35. L. Torres and colleagues, “Large-scale crop disease detection using drone-based hyperspectral imaging,” Remote Sensing, vol. 13, no. 18, 3614, 2021.