Aviation Electronics, Information Technology, Telecommunications, Electricals, and Controls (AVITEC)
Vol. 8, No. 2, August 2026, pp. 225-236, p-ISSN 2685-2381, e-ISSN 2715-2626
http://dx.doi.org/10.28989/avitec.v8i2.3942

Review Article

Multimodal Machine Learning for Maize Disease Detection: A Systematic Review of Architectures and Deployment Challenges

Mercy Chepkoech Tonui*, John Wachira Kamau, Raymond Wafula Ongus
School of Computing and Informatics, Mount Kenya University, Kenya

Article Info

Article history:
Submitted February 24, 2026
Accepted April 13, 2026
Published April 22, 2026

Keywords:
deep learning;
multimodal machine learning;
edge computing;
intelligent systems;
maize disease detection;
data fusion.

ABSTRACT

Maize diseases continue to threaten agricultural productivity and food security, particularly in developing regions where early diagnosis remains constrained by limited expert access. While deep learning has enabled automated disease detection systems, most existing approaches rely on unimodal image datasets and cloud-dependent architectures, limiting robustness and deployment feasibility in resource-constrained environments. This study presents a structured systematic review of 38 peer-reviewed studies published between 2020 and 2025, focusing on multimodal machine learning approaches integrating visual and environmental data for maize disease detection. Quantitative synthesis reveals that 58% of reviewed studies employ image-only deep learning models, 26% implement multimodal frameworks, and only 29% conduct validation under real or semi-real field conditions. Furthermore, 32% explicitly address deployment considerations, including edge computing and mobile inference. The findings demonstrate that multimodal architectures improve robustness and contextual modeling compared to unimodal systems by integrating phenotypic and environmental drivers of disease expression. However, increased computational complexity, synchronization challenges, and limited edge optimization remain barriers to scalable deployment. This review advances scientific knowledge by providing a computing-centered synthesis of multimodal architectures, fusion strategies, deployment constraints, and explainability gaps, identifying key research priorities in edge efficiency, real-world validation, and interpretable intelligent systems.

Corresponding Author:
Mercy Chepkoech Tonui,
School of Computing and Informatics, Mount Kenya University, Thika, Kenya.
Email: *chepsmercie@gmail.com

1. INTRODUCTION

Maize remains one of the most widely cultivated staple crops globally and plays a central role in food security, economic stability, and agricultural sustainability. However, maize production is highly susceptible to infectious diseases such as Maize Lethal Necrosis (MLN), Gray Leaf Spot (GLS), and Northern Leaf Blight (NLB), which can significantly reduce yield when not detected at early stages. Conventional diagnostic approaches—including manual field inspection and laboratory-based pathogen confirmation—are often subjective, time-consuming, and difficult to scale in smallholder agricultural systems [1][2]. These constraints have accelerated interest in artificial intelligence (AI)–driven diagnostic systems.

While this study adopts a global systematic review methodology, the findings are interpreted within the context of smallholder agricultural systems, particularly in maize-growing regions such as Bomet County, Kenya. This region is characterized by environmental variability, limited access to expert diagnosis, and resource constraints, making it a relevant case for understanding the practical implications of multimodal disease detection systems.

Deep learning, particularly convolutional neural networks (CNNs), has become the dominant methodology for plant disease recognition due to its ability to automatically learn hierarchical spatial representations from raw leaf imagery [3-5] demonstrated that CNN-based architectures outperform traditional machine learning classifiers in plant leaf disease classification. Similarly, Ref. [6] showed that transfer learning–based CNN models achieve high classification accuracy under controlled image acquisition conditions [7].

Broader surveys by Refs. [1][8-10] consistently identify CNN architectures as the prevailing paradigm in agricultural disease detection research.

Despite promising benchmark performance, unimodal image-based CNN systems exhibit inherent limitations when deployed under real agricultural conditions. Variability in illumination, background clutter, leaf orientation, and camera quality significantly affects model generalization [11]. Recent empirical studies further confirm that plant disease detection models trained on controlled datasets often exhibit reduced performance when deployed in heterogeneous agricultural environments [12][13]. Ref. [14] reported performance degradation under uncontrolled field settings, while Ref. [15] emphasized that CNN performance varies across datasets collected under heterogeneous environmental conditions. These findings highlight the gap between laboratory-level accuracy optimization and real-world robustness.

Importantly, plant disease manifestation is inherently multidimensional. Visible leaf symptoms represent phenotypic outcomes influenced by environmental drivers such as temperature, humidity, rainfall, and soil moisture. Ref. [16] demonstrated that environmental time-series variables significantly contribute to disease prediction using LSTM architectures. Similarly, Ref. [17] showed that integrating environmental signals with image features improves classification robustness under climatic variability. Purely visual models therefore fail to incorporate contextual and temporal dependencies that influence disease progression.

Sensor-based and time-series modeling approaches have been proposed to capture these environmental dynamics. LSTM-based environmental modeling enables early disease risk prediction by analyzing climatic patterns [16][17]. However, environmental predictors alone may generate false positives because favorable climatic conditions do not always result in infection [2]. Furthermore, sensor-based systems introduce deployment challenges related to hardware cost, calibration, maintenance, and energy efficiency [18].

To overcome these limitations, multimodal machine learning has emerged as a structurally superior paradigm. Multimodal frameworks integrate heterogeneous data modalities—typically combining CNN-extracted visual features with LSTM-modeled environmental data—to jointly represent phenotypic and contextual disease indicators [8][17][19] .Empirical evidence indicates that multimodal feature fusion improves robustness and reduces misclassification compared to unimodal baselines [17][20]. Hybrid and attention-based fusion strategies further enhance cross-modal interaction and contextual representation [10].

Beyond RGB imagery, multispectral and hyperspectral integration enhances disease discrimination by capturing reflectance characteristics outside the visible spectrum. Recent UAV-based multispectral imaging studies demonstrate earlier detection of crop stress and disease symptoms compared with traditional RGB-based monitoring systems [21]. Ref. [22] demonstrated improved disease diagnosis using UAV-based hyperspectral remote sensing. Ref. [23] provided a systematic comparative analysis of RGB and hyperspectral approaches, reporting enhanced sensitivity with spectral–spatial fusion. However, Ref. [24] emphasized that hyperspectral systems increase computational and hardware complexity, limiting scalability in resource-constrained agricultural settings.

Parallel research has focused on deployment feasibility. Edge computing architectures enable local inference on mobile and embedded devices, reducing cloud dependency and latency [18][25]. Recent research also explores edge-AI frameworks and lightweight neural architectures designed for deployment on IoT-enabled agricultural sensing platforms [13][17][26].Mobile-based frameworks for low-resource environments have demonstrated the feasibility of real-time crop disease detection [27][28]. Lightweight CNN models optimized through pruning and quantization further support edge deployment while maintaining acceptable accuracy [29]. Nevertheless, integrated multimodal systems optimized specifically for edge environments remain relatively underexplored.

Several recent review articles provide comprehensive overviews of deep learning applications in plant disease detection [4][8-10][30-33]. While these studies summarize CNN architectures and comparative accuracy metrics across crops, they predominantly emphasize unimodal image-based approaches. Limited attention has been given to systematically quantifying the distribution of unimodal versus multimodal frameworks, evaluating fusion strategies from a computing perspective, and analyzing deployment trade-offs within maize-specific disease detection research.

Furthermore, explainable AI remains under-integrated in multimodal agricultural systems. Although [9] and [10] highlight the importance of interpretability mechanisms for trustworthy AI deployment, empirical adoption within maize disease detection studies remains limited. Recent studies applying explainable artificial intelligence techniques such as Grad-CAM, SHAP, and attention visualization demonstrate improved interpretability and user trust in AI-driven agricultural diagnostics[34-36].

Therefore, a critical research gap persists in the consolidated, computing-centered synthesis of multimodal maize disease detection architectures and deployment constraints. Specifically, insufficient integration exists between quantitative methodological trend analysis, comparative evaluation of unimodal and multimodal frameworks, and system-level assessment of deployment readiness in resource-constrained environments.

Despite the rapid advancement of deep learning techniques for plant disease detection, several critical research gaps remain. First, most existing studies rely on unimodal image-based convolutional neural network (CNN) architectures, which focus solely on visual leaf symptoms and fail to incorporate contextual environmental variables that influence disease development. This methodological limitation restricts the robustness of disease detection systems under heterogeneous field conditions where illumination variability, background noise, and camera differences affect model performance. Second, the majority of existing studies emphasize laboratory-based accuracy optimization while overlooking deployment constraints such as computational cost, edge-device compatibility, energy consumption, and real-time inference requirements in smallholder farming environments. Third, explainability mechanisms remain underexplored in multimodal agricultural AI systems, limiting transparency and trust in automated disease diagnosis models used by farmers and agricultural extension services. These limitations highlight the need for a structured synthesis of multimodal machine learning approaches that integrates methodological design, deployment feasibility, and explainability considerations within maize disease detection research.

Therefore, this study aims to provide a systematic synthesis of multimodal machine learning approaches for maize disease detection published between 2020 and 2025. Specifically, the review seeks to (1) analyze methodological trends in maize disease detection research, (2) evaluate the distribution and effectiveness of unimodal and multimodal learning architectures, (3) examine multimodal data fusion strategies and their implications for model robustness, and (4) assess deployment considerations including edge computing feasibility and real-world field validation. The primary contribution of this review is the provision of a computing-centered synthesis that integrates architectural analysis, quantitative methodological trends, and deployment constraints. By explicitly quantifying modality distributions, fusion strategies, and validation practices across the literature, this study advances understanding of maize disease detection as an integrated intelligent systems challenge and provides guidance for the development of scalable, field-ready multimodal agricultural AI systems.

2. REVIEW METHODOLOGY

2.1 Review Design and Framework

This study adopted a structured systematic literature review methodology to synthesize advances in machine learning–based maize disease detection between 2020 and 2025, with emphasis on multimodal architectures, fusion mechanisms, deployment feasibility, and explainability.

The review followed an adapted PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework to ensure methodological transparency, reproducibility, and structured study selection. Although PRISMA was originally developed for clinical research, its staged identification–screening–eligibility–inclusion structure was applied to guide this computing-oriented review. The review protocol was defined prior to literature screening to ensure transparency and reproducibility of the search and selection process.

Although the included studies are drawn from global literature, the analytical interpretation of findings is contextualized toward agricultural environments similar to those found in sub-Saharan Africa, particularly Bomet County, Kenya, to enhance the practical relevance of the review.

2.2 Information Sources

A comprehensive search was conducted across seven major academic databases between 20 August 2025 and 25 August 2025 to capture relevant peer-reviewed literature. Table 1 presents the academic databases used for literature retrieval. Detailed information on the included studies, including titles, methods, and contributions, is provided in Table 1.

Table 1. Databases searched

No.	Database	Publisher	Scope	Access Date
1	Scopus	Elsevier	Multidisciplinary indexed journals	20–25 Aug 2025
2	IEEE Xplore	IEEE	Engineering & AI research	20–25 Aug 2025
3	Web of Science	Clarivate	High-impact indexed journals	20–25 Aug 2025
4	ScienceDirect	Elsevier	Applied computing & agriculture	20–25 Aug 2025
5	SpringerLink	Springer Nature	AI & ML journals	20–25 Aug 2025
6	MDPI	MDPI	Open-access agricultural journals	20–25 Aug 2025
7	Google Scholar	Google	Supplementary coverage	20–25 Aug 2025

Only publications written in English and published between January 2020 and August 2025 were considered.

2.3 Search Strategy

A structured Boolean search strategy was developed to capture studies integrating maize disease detection with machine learning, deep learning, multimodal learning, environmental data, and edge computing.

The following search string was applied in Scopus:

```
TITLE-ABS-KEY (
     ("maize disease detection" OR "plant disease detection" OR "crop disease detection")
     AND
     ("machine learning" OR "deep learning" OR "convolutional neural network" OR "CNN" OR "LSTM")
     AND
     ("multimodal" OR "data fusion" OR "environmental data" OR "sensor data" OR "edge computing")
)
AND PUBYEAR > 2019 AND PUBYEAR < 2026
AND (LIMIT-TO (LANGUAGE, "English"))
```

Equivalent syntax adjustments were applied in IEEE Xplore and Web of Science using database-specific filtering systems. The structured Boolean queries and database-specific search strategies used in this review are summarized in Table 2. The table also presents the number of records retrieved from each database, providing a quantitative overview of the literature identification process.

Table 2. Boolean queries and retrieval summary

Database	Boolean Query (Exact Syntax Used)	Filters Applied	Records Retrieved
Scopus	TITLE-ABS-KEY (("maize disease detection" OR "plant disease detection") AND ("machine learning" OR "deep learning" OR "CNN" OR "LSTM") AND ("multimodal" OR "data fusion" OR "environmental data" OR "edge computing"))	2020–2025; English; Article & Conference	46
IEEE Xplore	("maize disease detection") AND ("deep learning" OR "CNN" OR "LSTM") AND ("multimodal" OR "edge computing")	2020–2025; Journals & Conferences	34
Web of Science	TS=("maize disease detection") AND TS=("machine learning") AND TS=("multimodal")	2020–2025; English	22
ScienceDirect	("maize disease detection") AND ("deep learning") AND ("multimodal")	2020–2025	31
SpringerLink	("maize disease" AND "machine learning") AND ("multimodal")	2020–2025	29
Google Scholar	("maize disease detection" AND "machine learning" AND "multimodal")	2020–2025	58
Total	—	—	**220**

2.4 Study Selection Procedure

The database search yielded a total of 220 records. All retrieved records were exported in RIS format and imported into Mendeley Reference Manager (Version 2.110) for reference management and duplicate removal.

Deduplication was performed within Mendeley using automatic duplicate detection, followed by manual verification to ensure accuracy. After removal of duplicate entries, the remaining records proceeded to screening.

Screening was conducted in two sequential stages. First, title and abstract screening was performed to exclude clearly irrelevant studies. Second, full-text eligibility assessment was carried out to determine compliance with the predefined inclusion and exclusion criteria. The primary author conducted the initial screening process. To enhance methodological reliability, a second reviewer independently evaluated a randomly selected 20% subset of screened articles. Any discrepancies were resolved through discussion and consensus to ensure consistency in study selection.

No automated scripts, machine learning tools, or AI-assisted screening systems were used during the identification, screening, or eligibility assessment stages. The study selection process, including retrieval, screening, eligibility assessment, and final inclusion, is summarized in Table 3. This structured workflow ensures transparency and reproducibility in the selection of studies included in the review.

Table 3. Study selection workflow

Stage	Description	Records (n)
Initial Retrieval	Database search results	220
After Deduplication	Unique records	182
Title/Abstract Screening	Potentially relevant	96
Full-Text Eligibility	Meets criteria	52
Final Included Studies	Included in synthesis	**38**

The structured selection process is illustrated in Figure 1 (PRISMA flow diagram).

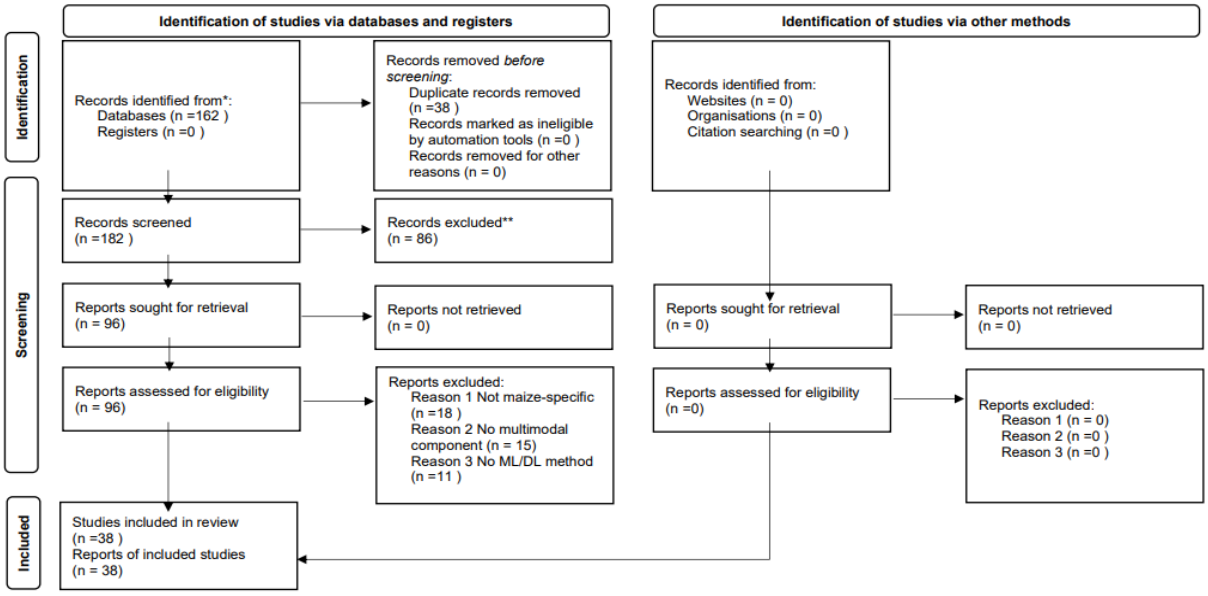

Figure 1. Adapted PRISMA flow diagram illustrating the structured study selection process

2.5 Eligibility Criteria

The inclusion and exclusion criteria applied during the screening and eligibility stages are presented in Table 4. These criteria guided the selection of studies to ensure relevance, quality, and consistency with the objectives of this review.

Table 4. Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Peer-reviewed journal or full conference paper	Non-peer-reviewed sources
Published between 2020–2025	Published before 2020
Applies ML/DL to maize disease detection	Pure agronomic/biological study
Uses image, environmental, or multimodal data	Traditional visual diagnosis only
Full-text available in English	Abstract-only papers

2.6 Data Extraction

A structured data extraction template was developed to ensure consistency and comparability. The variables extracted from each selected study are summarized in Table 5. These variables were designed to capture key methodological, architectural, and deployment-related characteristics relevant to multimodal maize disease detection. Data extraction was performed manually to ensure contextual interpretation accuracy.

Table 5. Data extraction variables

Category	Variables Collected
Bibliographic	Author(s), Year, Country
Algorithm Type	CNN, LSTM, Transformer, Hybrid
Data Modality	Image-only, Environmental, Multimodal
Fusion Strategy	Early, Late, Feature-level, Hybrid
Deployment Orientation	Cloud-based, Edge-based, Mobile
Field Validation	Controlled / Real-field
Explainability	XAI applied (Yes/No)
Performance Metrics	Accuracy, robustness indicators

2.7 Data Synthesis

A mixed qualitative–quantitative synthesis approach was employed to analyze the included studies. The quantitative synthesis involved frequency distribution analysis, percentage computation of modality categories, evaluation of deployment orientation patterns, and assessment of explainability adoption rates across the selected literature. These computations were used to identify methodological imbalances and structural trends within maize disease detection research.

The qualitative synthesis focused on comparative evaluation of unimodal and multimodal frameworks, analysis of architectural trade-offs, and identification of recurring research gaps. Emphasis was placed on understanding the relationship between model design choices, robustness under field variability, and deployment feasibility in resource-constrained environments. Due to heterogeneity in datasets, experimental protocols, and evaluation metrics across studies, formal statistical meta-analysis was not conducted.

The selected time frame (2020–2025) reflects the most recent advancements in deep learning, multimodal machine learning, and edge computing within agricultural applications. A total of 38 studies were included following strict inclusion and exclusion criteria. While the number of studies may appear moderate, it represents a high-quality and focused subset of peer-reviewed research directly addressing maize disease detection using machine learning techniques. The studies were sourced from multiple major academic databases, ensuring diversity in methodologies and geographical representation. This structured selection ensures that the dataset is sufficiently representative for both quantitative and qualitative synthesis.

3. RESULTS AND DISCUSSION

To enhance transparency and traceability of the reviewed literature, representative included studies—along with their titles, methods, and contributions—are summarized in Table 6.

Table 6. Representative included studies with titles, methods, and contributions

Ref.	Author(s)	Year	Title	Method	Modality	Key Contribution
[1]	Benos et al.	2021	Machine learning in agriculture: A comprehensive updated review	CNN	Image	Provides comprehensive overview of ML in agriculture
[8]	Zhou et al.	2021	Crop disease identification and interpretation using multimodal deep learning	CNN + LSTM	Multimodal	Improves robustness through multimodal fusion
[27]	Khan et al.	2023	A mobile-based system for maize plant leaf disease detection and classification using deep learning	CNN	Image	Enables real-time mobile-based disease detection
[20]	De Silva & Brown	2023	Multispectral plant disease detection with vision transformer CNN hybrid approaches	CNN + Transformer	Multimodal	Enhances spectral–spatial feature extraction
[28]	Askale et al.	2025	Mobile-based deep CNN model for maize leaf disease detection	CNN	Image	Supports edge-based deployment in low-resource settings
[19]	Lee et al.	2024	Deep learning-based crop disease diagnosis using multimodal mixup augmentation	CNN + Fusion	Multimodal	Improves generalization using multimodal augmentation
[16]	Gafurov et al.	2023	Application of LSTM networks in agricultural crop recognition and prediction	LSTM	Environmental	Captures temporal environmental patterns
[13]	Fu et al.	2022	Maize disease detection based on spectral recovery from RGB images	CNN	Image	Improves disease detection using spectral recovery

Table 6 presents a representative subset of the 38 included studies. All studies included in the systematic review were used in the quantitative and qualitative synthesis.

3.1 Principal Quantitative Findings

The systematic synthesis of the 38 included studies reveals a clear methodological imbalance in maize disease detection research published between 2020 and 2025. Image-based deep learning approaches dominate the field, accounting for 58% (22/38) of reviewed studies. Multimodal frameworks integrating image and environmental data represent 26% (10/38), while 16% (6/38) rely primarily on environmental or time-series modeling without visual fusion.

From a deployment perspective, 68% (26/38) of studies focus on cloud-based or laboratory evaluation environments, whereas only 32% (12/38) explicitly address edge computing or mobile deployment. Furthermore, real or semi-real field validation is conducted in just 29% (11/38) of cases, with the majority (71%) relying exclusively on controlled datasets. These findings are particularly significant for regions such as Bomet County, Kenya, where agricultural systems operate under variable environmental conditions and limited technological infrastructure, necessitating robust and deployable disease detection solutions.These distributions are summarized in Table 7

Table 7. Quantitative Synthesis of Included Studies (n = 38)

Category	Subcategory	Count	Percentage (%)
Primary Focus	Image-based models	22	58%
	Multimodal (image + environmental)	10	26%
	Environmental/time-series only	6	16%
Deployment Orientation	Cloud/laboratory evaluation	26	68%
	Edge/mobile deployment	12	32%
Validation Setting	Controlled dataset only	27	71%
	Real/semi-real field validation	11	29%
Method Type	Experimental deep learning	24	63%
	Analytical/comparative/review	14	37%

A comparative analysis between unimodal and multimodal approaches reveals important performance and robustness differences. Unimodal CNN models generally achieve high classification accuracy under controlled datasets; however, their performance deteriorates when exposed to environmental variability such as illumination changes, background noise, and leaf occlusion. In contrast, multimodal systems that integrate environmental or temporal data demonstrate improved robustness by incorporating contextual disease drivers such as temperature, humidity, and soil moisture. These additional data streams enable models to capture disease development dynamics rather than relying solely on visual symptom recognition. However, multimodal systems introduce higher computational requirements and increased architectural complexity, which may limit their deployment on resource-constrained edge devices.

3.2 Comparison with Prior Review Studies

The dominance of image-only CNN approaches identified in this review (58% of included studies) aligns with earlier surveys by Ref. [4] and [1] , who reported convolutional neural networks as the prevailing paradigm in plant disease detection research. Similarly, Ref. [9] and [10] observed that most agricultural AI studies prioritize visual classification accuracy, often evaluated under controlled experimental conditions.

However, unlike these broader reviews, the present study provides a maize-specific quantitative synthesis and explicitly measures the relative adoption of multimodal frameworks (26%) and real-field validation practices (29%). Previous reviews describe multimodal learning conceptually and highlight its potential advantages, yet they do not systematically quantify its prevalence within maize-focused disease detection research. By providing percentage-based methodological distributions, this study advances beyond narrative summarization toward structured empirical synthesis.

Furthermore, consistent with Ref. [17] and [20], the multimodal systems reviewed here demonstrate improved robustness under environmental variability compared to image-only models. The findings reinforce the argument that contextual integration of climatic or sensor data mitigates limitations associated with unimodal CNN architectures, particularly under heterogeneous field conditions. Nevertheless, the relatively low adoption rate of multimodal systems confirms that the transition from unimodal classification toward integrated intelligent systems remains incomplete.

3.3 Fusion Strategies in Multimodal Architectures

Among the 10 multimodal studies identified, feature-level (intermediate) fusion was the most common strategy (60%), followed by late decision fusion (30%) and early data-level fusion (10%). The distribution of fusion strategies across the identified multimodal studies is summarized in Table 8. The table highlights the prevalence of feature-level, late decision, and early data-level fusion approaches and their associated characteristics.

Table 8. Fusion Strategies in Multimodal Studies (n = 10)

Fusion Type	Studies (n)	Percentage (%)	Characteristics
Feature-level fusion	6	60%	Independent encoders, joint embedding
Late decision fusion	3	30%	Separate predictions combined
Early data-level fusion	1	10%	Raw concatenation before feature extraction

Feature-level fusion allows independent extraction of spatial features through CNN encoders and temporal or environmental features through LSTM or recurrent networks before joint embedding. This architecture enables complementary representation learning and reduces feature interference.

In contrast, early fusion approaches concatenate raw inputs, which may introduce dimensional imbalance and increase computational complexity. Late fusion combines independent predictions but may fail to capture cross-modal dependencies.

These structural differences influence both performance robustness and computational cost.

3.4 Why Multimodal Architectures Outperform Unimodal Systems

The improved performance of multimodal models can be explained through complementary representation mechanisms.

Unimodal CNN systems rely exclusively on spatial phenotypic features extracted from leaf imagery. While effective under controlled conditions, such models are sensitive to illumination variability, background clutter, and camera differences. As reported by Ref. [6], CNN performance declines under uncontrolled field conditions.

Multimodal systems, however, incorporate environmental drivers such as temperature, humidity, and soil moisture. Since disease manifestation is influenced by climatic conditions, contextual modeling reduces false positives and improves generalization across locations. Experimental evaluations further confirm that multimodal deep learning architectures combining image features with environmental sensor data improve generalization across heterogeneous agricultural environments [37][38].

Environmental and time-series modeling approaches have demonstrated the potential of recurrent architectures such as LSTM for predicting fungal disease progression based on climatic signals [39]. Similarly, in-field maize-specific evaluations highlight the performance benefits of contextual modeling under heterogeneous environmental conditions [40]. Given the temporal dynamics of climatic drivers in plant disease manifestation, recent deep learning time-series approaches demonstrate the importance of recurrent modeling in agricultural prediction contexts. Such approaches reinforce the potential of temporal environmental integration to improve robustness in maize disease forecasting [9][21]. These findings reinforce the argument that multimodal integration mitigates the sensitivity of unimodal CNN models to illumination variability and background noise.

As demonstrated by Ref. [16], LSTM-based environmental modeling captures temporal disease patterns, while Ref. [17] showed that joint image–environment fusion enhances robustness under climatic variation.

Mechanistically, multimodal architectures reduce feature ambiguity by integrating complementary representational streams. Spatial symptom characteristics are extracted through CNNs, while temporal and environmental dynamics are modeled using recurrent architectures such as LSTM networks. These heterogeneous feature representations are subsequently projected into a shared embedding space, enabling joint learning of phenotypic and contextual disease indicators. This integrated representation mitigates dataset-specific visual bias, reduces sensitivity to illumination variability, and enhances cross-domain stability across heterogeneous agricultural environments. However, the computational cost increases due to parallel encoders and fusion layers, which may limit edge deployment feasibility.

The mathematical foundation of the multimodal architecture is based on the integration of visual feature extraction, temporal environmental modeling, feature fusion, and probabilistic classification.

The CNN used for extracting spatial features from maize leaf images is defined as in Equation (1).

$$F_v = \sigma(W * X + b) \quad (1)$$

where: F_v : the extracted visual feature maps
X : the input maize leaf image
W : denotes convolutional filters, bis the bias term
σ : the nonlinear activation function.

The LSTM model used for environmental time-series data is expressed as Equation (2).

$$h_t = LSTM(x_t, h_{t-1}) \quad (2)$$

where: x_t : environmental inputs (e.g., temperature, humidity, soil moisture) at time step t
h_{t-1} : the previous hidden state, h_t is the updated hidden state capturing temporal dependencies.

The multimodal fusion process combining both feature representations is defined as Equation (3).

$$F = F_v \oplus F_e \quad (3)$$

where: F_v : visual features
F_e : represents environmental features
$\oplus$: denotes feature-level fusion (concatenation) used to combine heterogeneous data representations.

The final classification decision is obtained using the softmax function as Equation (4).

$$P(y_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (4)$$

where: $P(y_i)$:the probability of class i
z_i : the output score of the final fully connected layer.

These formulations provide a theoretical foundation for understanding how multimodal architectures integrate heterogeneous data sources to enhance robustness and classification performance.

3.5 Deployment Challenges and Edge Computing Trends

Although 32% of studies explicitly address edge or mobile deployment, most multimodal architectures remain cloud-dependent due to computational overhead.

Edge computing integration, as discussed by Refs. [18] and [27], enables local inference and reduces latency. Lightweight CNN models optimized through pruning and quantization further support embedded deployment [29].

Nevertheless, multimodal models introduce synchronization complexity between image and sensor streams, increasing memory requirements and inference latency. This explains the current research imbalance: unimodal systems dominate due to simplicity, while multimodal systems offer robustness but at increased architectural cost.

3.6 Limitations Identified in Current Literature

The review identified several recurring limitations within the existing body of maize disease detection research. First, real-field validation remains limited, with only 29% of studies conducting evaluation under real or semi-real agricultural conditions. Second, cross-regional dataset testing is insufficient, restricting the generalizability of reported model performance across diverse climatic and geographic contexts. Third, the adoption of explainable artificial intelligence mechanisms remains sparse, limiting transparency and trust in automated disease detection systems. Fourth, benchmarking of multimodal fusion strategies under deployment constraints is minimal, with few studies systematically evaluating trade-offs between robustness, computational cost, and edge feasibility.

Additionally, existing review articles often aggregate crop-agnostic findings without isolating maize-specific methodological distributions [9][41-43]. Consequently, the relative adoption of multimodal architectures and real-field validation practices remains under-quantified within maize-centered research contexts.

3.7 Synthesis of Research Gap

The quantitative findings and comparative analysis reveal a clear structural imbalance in current maize disease detection research. The literature remains predominantly dominated by image-only CNN models, with limited integration of environmental and contextual modalities. Although multimodal systems demonstrate improved robustness and contextual awareness, their architectural optimization for scalable deployment remains insufficient. In particular, edge-aware multimodal frameworks are scarce, reflecting a disconnect between algorithmic innovation and practical implementation in resource-constrained agricultural environments. Furthermore, explainability mechanisms are inconsistently incorporated, limiting transparency and trust in automated diagnostic systems. Collectively, these observations confirm that maize disease detection research must transition from isolated, accuracy-focused modeling toward integrated intelligent system design that balances robustness, interpretability, and deployment feasibility.

4. CONCLUSION

This systematic review synthesized 38 peer-reviewed studies published between 2020 and 2025 to evaluate methodological trends, architectural strategies, and deployment considerations in maize disease detection. Despite the structured synthesis presented in this study, several limitations should be acknowledged. First, the review was restricted to English-language publications indexed in selected academic databases between 2020 and 2025, which may exclude relevant studies published in other languages or regional outlets. Second, the

reviewed studies exhibit heterogeneity in datasets, evaluation protocols, and performance metrics, limiting the possibility of conducting a formal quantitative meta-analysis. Finally, many studies rely on controlled experimental datasets, which may not fully represent real agricultural environments. The quantitative analysis revealed that research remains dominated by image-only deep learning approaches (58%), while multimodal frameworks integrating environmental data account for only 26% of studies and real-field validation is conducted in just 29% of cases. Although multimodal architectures demonstrate structural robustness advantages through complementary spatial–temporal feature representation, their adoption is constrained by computational complexity and limited edge optimization. By explicitly quantifying modality distribution, fusion strategies, deployment orientation, and validation practices, this review advances scientific knowledge beyond descriptive surveys and reframes maize disease detection as an integrated intelligent systems challenge rather than a purely image classification task. Although the reviewed studies are global in scope, the findings are highly applicable to maize-producing regions such as Bomet County, Kenya, where multimodal and edge-based intelligent systems can address real-world agricultural challenges and improve early disease detection. Future research should focus on developing lightweight multimodal architectures optimized for edge computing environments to enable real-time disease detection in smallholder agricultural systems. In addition, the integration of explainable artificial intelligence techniques such as Grad-CAM, SHAP, and attention-based interpretability mechanisms will be essential to improve transparency and trust in automated diagnostic systems. Further research should also prioritize large-scale field validation across diverse climatic regions to enhance model generalizability. Future studies should also investigate standardized multimodal benchmarking datasets to enable consistent comparison of fusion architectures and deployment strategies. Finally, interdisciplinary collaboration between computer scientists, agronomists, and agricultural extension officers will be necessary to translate multimodal intelligent systems from experimental prototypes into practical decision-support tools for farmers.

Author Contributions: M.C.T. conceptualized the study, designed the review methodology, conducted data collection, screening, and analysis, and developed the multimodal synthesis framework. J.K. contributed to methodological validation, supervised the research process, and provided critical review of the manuscript. R.W.O. supported the interpretation of results, contributed to the discussion on deployment challenges, and reviewed the manuscript for technical accuracy. All authors read and approved the final manuscript.

Data and Supplementary Materials: The data supporting the findings of this study are derived from publicly available peer-reviewed publications included in the systematic review. Extracted datasets, synthesis tables, and supplementary materials such as classification summaries, fusion strategy categorizations, and methodological distributions are available from the corresponding author upon reasonable request for academic purposes.

Funding and Acknowledgment: This research received no external funding. The author gratefully acknowledges the academic guidance, mentorship, and constructive feedback provided by the supervisors at the School of Computing and Informatics, Mount Kenya University, during the development of this work. Their insights and support contributed significantly to the refinement of the research focus and overall quality of this review paper.

Conflict of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCE

[1] L. Benos, A. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, "Machine learning in agriculture: A comprehensive updated review," Sensors, vol. 21, no. 11, p. 3758, 2021. https://doi.org/10.3390/s21113758

[2] T. Dibbern, L. A. Santos Romani, and S. M. F. S. Massruhá, "Main drivers and barriers to the adoption of digital agriculture technologies," Smart Agric. Technol., vol. 7, p. 100459, 2024. https://doi.org/10.1016/j.atech.2024.100459

[3] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, "Using deep transfer learning for image-based plant disease identification," Comput. Electron. Agric., vol. 173, p. 105393, 2020. https://doi.org/10.1016/j.compag.2020.105393

[4] J. Lu, L. Tan, and H. Jiang, "Review on convolutional neural network applied to plant leaf disease classification," Agriculture, vol. 11, no. 8, p. 707, 2021. https://doi.org/10.3390/agriculture11080707

[5] G. Shandilya et al., "Hybrid CNN–ViT architecture for maize leaf disease classification," Food Sci. Nutr., early access, 2025. https://doi.org/10.1002/fsn3.70513

[6] Z. Ma, Y. Wang, T. S. Zhang, H. G. Wang, Y. J. Jia, and R. Gao, "Maize leaf disease identification using deep transfer convolutional neural networks," Int. J. Agric. Biol. Eng., vol. 15, no. 5, pp. 187–195, 2022. https://doi.org/10.25165/j.ijabe.20221505.6658

[7] A. Singla et al., "Exploration of machine learning approaches for automated crop disease detection," Curr. Plant Biol., vol. 30, p. 100382, 2024. https://doi.org/10.1016/j.cpb.2024.100382
[8] J. Zhou, J. Li, and C. Wang, "Crop disease identification and interpretation method based on multimodal deep learning," Comput. Electron. Agric., vol. 189, p. 106408, 2021. https://doi.org/10.1016/j.compag.2021.106408
[9] A. Upadhyay, "Deep learning and computer vision in plant disease detection: A comprehensive review of techniques, models and trends in precision agriculture," Artif. Intell. Rev., vol. 58, 2025. https://doi.org/10.1007/s10462-024-11100-x
[10] S. Wang et al., "Advances in deep learning applications for plant disease and pest detection: A review," Remote Sens., vol. 17, no. 4, p. 698, 2025. https://doi.org/10.3390/rs17040698
[11] S. S. Chouhan, U. P. Singh, and S. Jain, "Applications of computer vision in plant pathology: A survey," Arch. Comput. Methods Eng., vol. 27, pp. 611–632, 2020. https://doi.org/10.1007/s11831-019-09324-0
[12] H. N. Ngugi, A. A. Akinyelu, and A. E. Ezugwu, "Machine learning and deep learning for crop disease diagnosis: Performance analysis and review," Agronomy, vol. 14, no. 12, p. 3001, 2024. https://doi.org/10.3390/agronomy14123001
[13] J. Fu et al., "Maize disease detection based on spectral recovery from RGB images," Front. Plant Sci., vol. 13, p. 1056842, 2022. https://doi.org/10.3389/fpls.2022.1056842
[14] M. Shoaib et al., "Advanced deep learning models-based plant disease detection: A review of recent research," Front. Plant Sci., vol. 14, p. 1158933, 2023. https://doi.org/10.3389/fpls.2023.1158933
[15] W. Shafik et al., "Deep learning technique for plant disease classification and pest detection and model explainability elevating agricultural sustainability," BMC Plant Biol., vol. 25, p. 1491, 2025. https://doi.org/10.1186/s12870-025-07377-x
[16] A. Gafurov, S. Mukharamova, A. Saveliev, and O. Yermolaev, "Advancing agricultural crop recognition: The application of LSTM networks and spatial generalization in satellite data analysis," Agriculture, vol. 13, no. 9, p. 1672, 2023. https://doi.org/10.3390/agriculture13091672
[17] B. Prashanthi and C. M. Rao, "A comparative study of fine-tuning deep learning models for leaf disease identification and classification," Eng. Technol. Appl. Sci. Res., vol. 15, no. 1, pp. 19661–19669, 2025. https://doi.org/10.48084/etasr.9017
[18] J. Zhang and D. Tao, "Empowering things with intelligence: A survey of the progress, challenges, and opportunities in artificial intelligence of things," IEEE Internet Things J., vol. 8, no. 10, pp. 7789–7817, 2021. https://doi.org/10.1109/JIOT.2020.3039359
[19] H. Lee et al., "A deep learning-based crop disease diagnosis method using multimodal mixup augmentation," Appl. Sci., vol. 14, no. 10, p. 4322, 2024. https://doi.org/10.3390/app14104322
[20] M. De Silva and D. Brown, "Multispectral plant disease detection with vision transformer convolutional neural network hybrid approaches," Sensors, vol. 23, no. 20, p. 8531, 2023. https://doi.org/10.3390/s23208531
[21] C. K. Sunil et al., "Deep learning-based plant disease detection: A systematic study," Artif. Intell. Rev., vol. 57, p. 10517, 2023. https://doi.org/10.1007/s10462-023-10517-0
[22] L. W. Kuswidiyanto, H.-H. Noh, and X. Han, "Plant disease diagnosis using deep learning based on aerial hyperspectral images: A review," Remote Sens., vol. 14, no. 23, p. 6031, 2022. https://doi.org/10.3390/rs14236031
[23] M. Shafay et al., "Recent advances in plant disease detection: Challenges and opportunities," Plant Methods, vol. 21, p. 140, 2025. https://doi.org/10.1186/s13007-025-01450-0
[24] D. Yang et al., "Non-destructive detection of defective maize kernels using hyperspectral imaging and convolutional neural network with attention module," Spectrochim. Acta A, vol. 309, p. 124166, 2024. https://doi.org/10.1016/j.saa.2024.124166
[25] X. Zhang, Z. Cao, and W. Dong, "Overview of edge computing in the agricultural Internet of Things," IEEE Access, vol. 8, pp. 141748–141761, 2020. https://doi.org/10.1109/ACCESS.2020.3013005
[26] S. Sakka et al., "CNN applications in smart agriculture using multimodal data," Sensors, vol. 25, no. 2, p. 472, 2025. https://doi.org/10.3390/s25020472
[27] F. Khan et al., "A mobile-based system for maize plant leaf disease detection and classification using deep learning," Front. Plant Sci., vol. 14, p. 1079366, 2023. https://doi.org/10.3389/fpls.2023.1079366
[28] G. T. Askale et al., "Mobile-based deep CNN model for maize leaf disease detection," Plant Methods, vol. 21, p. 72, 2025. https://doi.org/10.1186/s13007-025-01386-5
[29] T. O'Halloran, J. T. Byrne, and M. O'Neill, "A deep learning approach for maize lethal necrosis and maize streak virus detection using field imagery," Mach. Learn. Appl., vol. 16, p. 100556, 2024. https://doi.org/10.1016/j.mlwa.2024.100556
[30] J. Liu and X. Wang, "Plant diseases and pests detection based on deep learning: A review," Plant Methods, vol. 17, p. 22, 2021. https://doi.org/10.1186/s13007-021-00722-9

[31] I. Paçal et al., "A systematic review of deep learning techniques for plant diseases," Artif. Intell. Rev., vol. 57, p. 304, 2024. https://doi.org/10.1007/s10462-024-10944-7
[32] J. Zhao et al., "A review of plant leaf disease identification by deep learning algorithms," Front. Plant Sci., vol. 16, p. 1637241, 2025. https://doi.org/10.3389/fpls.2025.1637241
[33] M. Murugan, S. Arumugam, and R. Rajendran, "Research advances in maize crop disease detection using machine learning techniques," Computers, vol. 15, no. 2, p. 99, 2026, https://doi.org/10.3390/computers15020099
[34] R. Zhang et al., "A bibliometric review of deep learning in crop monitoring: Trends, challenges, and future perspectives," Front. Artif. Intell., vol. 8, p. 1636898, 2025. https://doi.org/10.3389/frai.2025.1636898
[35] B. N. Hassan and M. T. Somashekara, "A CNN-LSTM-based approach for the early detection of rice seed diseases," Eng. Technol. Appl. Sci. Res., vol. 15, no. 6, pp. 30643–30648, 2025. https://doi.org/10.48084/etasr.13630
[36] T. He, M. Li, and D. Jin, "Deep learning-based time series prediction for precision agriculture," Front. Plant Sci., vol. 16, p. 1575796, 2025. https://doi.org/10.3389/fpls.2025.1575796
[37] T. J. Maginga et al., "Design and implementation of IoT sensors for nonvisual symptoms detection on maize inoculated with Exserohilum turcicum," Smart Agric. Technol., vol. 5, p. 100260, 2023. https://doi.org/10.1016/j.atech.2023.100260
[38] J. Nakatumba-Nabende and S. Murindanyi, "Deep learning models for enhanced in-field maize leaf disease diagnosis," Mach. Learn. Appl., vol. 19, p. 100673, 2025. https://doi.org/10.1016/j.mlwa.2025.100673
[39] A. Jafar et al., "Revolutionizing agriculture with artificial intelligence: Plant disease detection methods, applications, and their limitations," Front. Plant Sci., vol. 15, p. 1356260, 2024. https://doi.org/10.3389/fpls.2024.1356260
[40] W. B. Demilie, "Plant disease detection and classification techniques: A comparative study," J. Big Data, vol. 11, p. 63, 2024. https://doi.org/10.1186/s40537-023-00863-9
[41] S. U. Khan, "A review on automated plant disease detection: Motivation, limitations, challenges, and recent advancements for future research," Plant Cell Tissue Organ Cult., 2025. https://doi.org/10.1007/s44443-025-00040-3