International Journal of Technology and Emerging Sciences (IJTES)

www.mapscipub.com

Volume 06 || Issue 01 || Jan 2026 || pp. 01-10 E-ISSN: 2583-1925

A Survey of deep learning Methods for Detection of Lumpy Skin Disease

George Mwangi Muhindi[1], Dr. Geoffrey Mariga Wambugu[2], Dr. Aaron Mogeni Oirere[3]

[1] *Student, Department of Information Technology, Murang'a University of Technology, Murang'a, Kenya*
[2] *Chair of Department, Department of Information Technology, Murang'a University of Technology, Murang'a, Kenya*
[3]*Chair of Department, Department of Computer Science. Murang'a University of Technology, Murang'a, Kenya*

Abstract - *Lumpy Skin Disease (LSD) is a viral threat to cattle that is increasingly gaining recognition as a major concern because it causes considerable economic losses through reduced productivity, trade restrictions, and increased mortality. Despite the rising use of deep learning (DL) in livestock health monitoring, few models have been effectively adapted for LSD detection. This review systematically explores current Deep Learning models in LSD diagnosis and compares model architectures, datasets, and deployment strategies. The review followed PRISMA 2020 guidelines for its methodology. A total of 146 studies were initially identified from six major databases (IEEE Xplore, ScienceDirect, SpringerLink, PubMed, Scopus, and arXiv). After screening and full-text eligibility assessment, 27 peer-reviewed studies published between 2017 and 2025 were included. Inclusion criteria focused on DL methods (CNN, Vision Transformers, hybrids) applied to image-based diagnosis in livestock. Convolutional Neural Networks (CNNs) remained dominant, with models like ResNet and EfficientNet achieving up to 98.3% accuracy in LSD classification. Transformer-based approaches like Vision Transformer (ViT) variants showed improved generalizability and semantic reasoning - with F1-scores up to 95.1% in multimodal tasks. Hybrid models like AnimalFormer and GasFormer offered improved robustness - though at the cost of computational complexity. Deployment-focused models (like MobileNetV2 and Tiny-YOLO) achieved real-time inference speeds with accuracies ranging from 89.4% to 93.8% - suggesting field applicability. Nonetheless, only 26% of the studies included explainability methods like Grad-CAM or SHAP and that factor limited interpretability. Even though there is technical progress, a major limitation remains: the absence of large, diverse, open-source annotated LSD image datasets. Addressing that limitation would enable reproducible benchmarking, facilitate model generalization, and accelerate global research collaboration. A coordinated international effort to develop and share standardized LSD datasets is necessary for advancing real-world deployment of DL-based diagnostic tools in livestock health.*

***Key Words*:** Deep learning, Lumpy Skin Disease (LSD), Vision Transformer, CNN, livestock diagnosis, image analysis, model explainability, dataset imbalance, veterinary AI.

1. INTRODUCTION

The global livestock industry plays an important role in ensuring food security, rural livelihoods, and economic stability [5]. However, the increasing incidence of infectious diseases among livestock presents a growing threat to animal health and agricultural productivity [1] [5]. In particular, viral dermatological diseases like LSD in cattle have gained considerable attention because of their rapid spread, severe economic impact, and challenges in early detection [13][9]. The ability to promptly identify and classify such diseases is necessary for effective containment and treatment strategies [16]. Traditional diagnostic methods - like physical examination and laboratory tests - are usually time-consuming, labor-intensive, and susceptible to human error in resource-constrained settings [17]. As such, the integration of computer vision and artificial intelligence (AI) techniques like deep learning (DL) has grown as a transformative solution for disease diagnosis and surveillance in livestock populations [22][3].

Deep learning is a subfield of machine learning that has shown exceptional performance in a wide array of computer vision tasks - from object detection and classification to image segmentation and anomaly detection [2][24]. In the agricultural domain, CNNs, ViTs, and hybrid models have been increasingly explored for tasks like animal behavior recognition [21], breed identification [5], body condition scoring [18], and disease diagnosis from medical and visual imagery [14] [27]. The appeal of DL-based systems is in their ability to learn hierarchical representations from raw image data. That factor facilitates the automated detection of subtle visual patterns showing disease pathology - patterns that may not be easily discernible to the human eye [19].

LSD in livestock diseases caused by the LSD virus (LSDV) presents a major challenge for cattle keepers in Africa, the Middle East, parts of Asia and other regions [9][1]. Characterized by the appearance of nodules on the skin and other tissues, LSD negatively affects milk production, fertility, hide quality, and market value [17]. Despite the disease's severity, technological interventions for its detection remain underexplored relative to more common human and veterinary ailments [13]. While there has been a growing body of research focused on general livestock disease detection using deep learning methods [4][10], LSD-specific studies are still emerging and there is a need for consolidated evidence and comparative evaluation [14].

The current systematic review aims to provide a comprehensive synthesis of current deep learning approaches applied to livestock disease image processing, with an emphasis on Lumpy Skin Disease. The review is structured in three key parts. The first is a broad overview of deep learning applications in livestock disease detection using image data. The second is an in-depth analysis of current models that focus specifically on LSD detection. The third is a comparative assessment of the dominant deep learning architectures - CNNs, vision transformers, and hybrid models - used in the domain.

In evaluating the literature, the current review identifies and discusses challenges in the field. Focus is on the issues of imbalanced datasets, limited generalizability across species and environments, and the lack of model explainability because they are factors that constrain the practical deployment of AI tools in different farm settings [10][12][14]. The review also synthesizes proposed solutions to those limitations, ranging from data augmentation and synthetic image generation [11] to attention-based models and explainable AI (XAI) frameworks [8][16]. Mapping the landscape of current research and pinpointing methodological gaps has enabled the current study to inform future work and guide the development of scalable and interpretable AI systems for veterinary diagnostics.

The other sections in the report presented are organized as follows. Section 2 presents the methodology used to conduct systematic review - including search strategy, inclusion and exclusion criteria, and data extraction protocols. Section 3 synthesizes studies on general livestock disease detection using deep learning, narrows the focus to LSD-specific models and techniques and offers a comparative analysis of deep learning architectures. Section 4 discusses major limitations, open challenges, and recommendations for future research. The paper concludes with a summary of findings and their implications for AI-driven livestock health management.

2. Methodology

2.1 Review Design

This study utilized a Systematic Literature Review (SLR) methodology to synthesize current evidence on the application of DL models for livestock disease detection - with a specific focus on image-based diagnosis of LSD. The SLR approach was chosen to provide a structured, transparent, and replicable framework for identifying, evaluating, and synthesizing research in different DL architectures and livestock domains. The SLR methodology facilitates the extraction of cross-study information and helps identify current gaps, methodological limitations, and new research opportunities.

To ensure comprehensiveness and transparency, the review followed the PRISMA 2020 guidelines [15] because it standardizes procedures for systematic evidence synthesis and ensures replicability. The steps followed start with defining the research question, then establishing inclusion and exclusion criteria, selecting relevant databases, applying a search strategy, screening titles and abstracts, full-text review, and ending with synthesis of the final set of studies. The SLR design is used much in computer science and veterinary research for promoting transparency, repeatability, and analytical depth (Afshari Savi, 2022). It is suitable for constantly changing fields like deep learning where continuous innovation requires periodic evidence consolidation [19].

The review aims to collate and analyze scholarly works on deep learning techniques applied to livestock disease image processing - with a dedicated focus on LSD. It also identifies technological trends, model strengths, limitations, and new research gaps following approaches recommended by recent AI-in-agriculture systematic reviews.

2.2 Research Questions

The review is structured around the following research questions, which draw upon prior studies that emphasize the need to evaluate deep learning's accuracy, usability, and adaptability in livestock applications [4][27]:

- ***RQ1:*** What deep learning architectures have been utilized in livestock disease image analysis, and how have they evolved over time?
- ***RQ2:*** What are the characteristics and performance outcomes of models used specifically for Lumpy Skin Disease detection?
- ***RQ3:*** What are the existing limitations in the current deep learning approaches (like dataset imbalance, generalizability and model explainability)?
- ***RQ4:*** What recommendations can be made for improving deep learning-based livestock disease detection systems?

These questions were formulated to guide the review toward addressing major challenges in real-world deployment and ensure relevance to veterinary diagnostics and AI development communities [23].

2.3 Inclusion and Exclusion Criteria

The inclusion and exclusion criteria were developed to ensure that selected articles were relevant, methodologically sound, and focused on deep learning applied to livestock disease image analysis:

Publication Type: Peer-reviewed journal or conference articles.

Technology: Studies employing image-based deep learning methods (like CNNs, ViTs and hybrid models).

Domain Focus: Studies related to livestock (like cattle, sheep, goats, pigs and dairy cows).

Language: English-language articles only.

Publication Date: Studies published between 2017 and 2025.

Articles were excluded if they:

i. Did not involve image-based DL methods.
ii. Focused solely on plant diseases or non-agricultural domains.
iii. Were review papers, editorials, or non-peer-reviewed preprints (except arXiv where empirical results were clearly presented and cited in subsequent literature).
iv. Lacked full-text availability.
v. Were duplicate entries across databases

The above criteria follow standards presented in other animal health reviews (Himel et al., 2024; [3]) and ensure inclusion of

studies that offer direct relevance to the technical and veterinary research communities.

Table 1: Eligibility Criteria

Category	Inclusion Criteria	Exclusion Criteria
Population	Livestock species: cattle, sheep, goats, pigs, dairy cows	Studies focusing exclusively on poultry, wild animals, or humans
Exposure	Use of image-based deep learning models (like CNN, ViT and hybrid models) applied to disease detection tasks	Studies using only non-image data (like environmental, sensor, genomic), and traditional ML without deep learning)
Outcome	Diagnostic performance metrics (like accuracy, F1-score, sensitivity, and specificity) for disease identification or classification	Studies without quantitative evaluation or lacking DL model validation
Study Design	Peer-reviewed journal articles and conference papers; empirical studies using real or synthetic datasets	Reviews, commentaries, editorials, theses, preprints not peer-reviewed, or protocols without implementation
Other	English language; published between 2017 and 2025	Articles in other languages; duplicates; inaccessible full texts; studies unrelated to livestock disease diagnostics

2.4 Search Strategy

A comprehensive search strategy was implemented across the following main databases: IEEE Xplore, ScienceDirect, SpringerLink, PubMed, Scopus, and arXiv. Those databases were selected for their wide coverage of peer-reviewed publications in artificial intelligence, agriculture, and veterinary sciences [16].

The search period was limited to studies published between 2017 and 2025 to capture advancements in convolutional neural networks (CNNs), Vision Transformers (ViTs), and hybrid models, including attention-based architectures like Swin Transformers and AnimalFormer [12].

The following Boolean-based keyword combinations were used:

- "deep learning" OR "convolutional neural network" OR "CNN" OR "transformer" OR "vision transformer" OR "ViT" OR "hybrid model"
- "livestock" OR "cattle" OR "sheep" OR "goat" OR "pig" OR "dairy cow"
- "disease detection" OR "image classification" OR "animal disease" OR "lumpy skin disease"

To ensure thoroughness, backward reference searching (scanning reference lists) was used to identify additional eligible articles - consistent with methods in recent livestock AI reviews [8] [20].

2.5 Study Selection Process

The initial search generated 146 studies. After removing duplicates and applying title and abstract screening, 52 papers remained for full-text assessment. Following the inclusion/exclusion criteria and qualitative appraisal of study relevance and methodological rigor, a total of 27 studies were retained for final synthesis.

The study selection process is shown in the PRISMA 2020 flow diagram (Figure 1), detailing identification, screening, eligibility, and inclusion stages.

Figure 1: PRISMA 2020 Flow Diagram

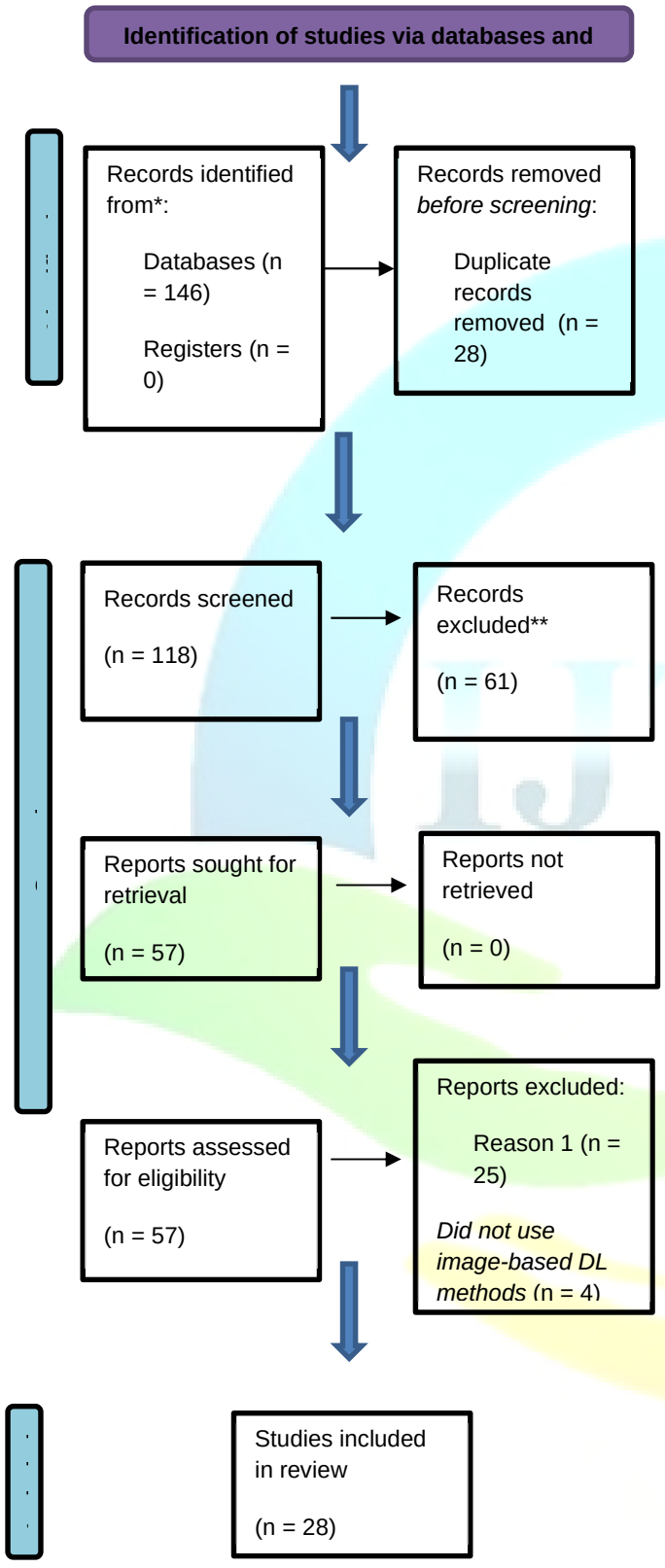

Figure 1: PRISMA 2020 Flow Diagram

3. Results

3.1 General Overview of Deep Learning in Livestock Disease Detection

The field of livestock disease detection has witnessed a significant transformation in recent years due to the rapid advancements in computer vision and deep learning technologies. Traditional veterinary diagnostic methods are usually reliant on visual inspection and manual intervention and are increasingly being augmented or replaced by automated systems capable of identifying symptoms and abnormalities from digital images or videos. That change is influenced by the need for scalable, accurate, and low-cost diagnostic solutions in regions with limited veterinary expertise or access to diagnostic laboratories.

Initially, computer vision applications in animal health utilized basic feature extraction techniques like histogram of oriented gradients (HOG), edge detection, and template matching. Those handcrafted feature-based methods, though, usually lacked the flexibility and robustness required for field variability and environmental noise. The advent of deep learning - especially CNNs - revolutionized the domain by allowing models to automatically learn discriminative features from raw image data. CNNs like ResNet, VGGNet, and MobileNet have since become foundational models for livestock disease classification [25][27].

The CNN architectures have been applied across a range of livestock species and diseases. For example, CNN-based systems have been developed for detecting foot-and-mouth disease, mastitis, and respiratory infections in cattle, and body condition scoring in goats and pigs [23][18]. In sheep farming, deep learning has facilitated facial recognition for breed identification and behavior monitoring [11][5]. Also, lightweight CNN variants like MobileNetV2 and YOLOv5s have enabled real-time and mobile-compatible deployment - an important requirement for resource-constrained environments [4][9].

The recent introduction of ViTs has added a new dimension to livestock disease image processing. Originally proposed by [2], ViTs utilize self-attention mechanisms instead of convolutions to model global dependencies in image data. That architecture offers improved feature representation - particularly in tasks requiring context awareness or long-range feature interactions. ViTs have shown promise in livestock domains like cow tracking [4], sheep recognition [27], and pig aggression classification [19]. Hybrid models that combine CNN feature extraction with Transformer attention blocks have also arisen to utilize the strengths of the two architectures [8][12].

In those developments, certain diseases and species have received disproportionate attention. Dairy cows dominate the landscape for tasks like body condition scoring, reproductive health assessment, and disease symptom classification [26][18]. Goats, sheep, and pigs follow in prevalence, though coverage remains uneven. Diseases like mastitis, lameness, and parasitic infections are commonly targeted, while viral diseases with dermatological symptoms like LSD are comparatively underrepresented - a gap that this review aims to address.

In terms of datasets, the majority of studies rely on proprietary or institution-specific image collections. Public datasets remain scarce, contributing to issues of reproducibility and cross-study benchmarking. Some researchers have adopted data augmentation, transfer learning, or synthetic data generation to reduce dataset limitations [20][3]. Nonetheless, dataset imbalance and lack of standardization continue to hinder generalization.

Performance evaluation is typically conducted using metrics like accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). While many models report high accuracy (>90%) on controlled test sets, real-world deployment usually shows performance degradation because of lighting variations, occlusions, animal movement, and differences in disease manifestation across breeds or regions [1][6].

Another growing consideration is model explainability. Given the high-stakes nature of livestock health decisions, veterinarians and farm workers require transparency in model predictions. Methods like Grad-CAM heatmaps and attention visualization have been explored to help interpret the decision-making process of CNNs and ViTs [22][12]. However, such techniques are not yet widely adopted, and few studies evaluate explainability systematically.

In summary, the landscape of deep learning applications in livestock disease image analysis is rapidly changing. CNNs remain dominant because of their proven effectiveness and hardware efficiency, but ViTs and hybrid models are gaining traction for their potential to improve generalization and interpretability. Major challenges persist - particularly related to dataset imbalance, explainability, and real-world robustness - which will need to be addressed to fully realize the potential of AI-driven animal health diagnostics.

3.2 Focused Review on LSD

LSD is a highly infectious viral condition affecting cattle, caused by the Capri poxvirus genus. It is characterized by nodular lesions on the skin, fever, lymphadenitis, emaciation, and sometimes death [17]. The disease causes substantial economic damage through reduced milk yield, reproductive failure, degraded hides, trade bans, and increased mortality [1][13]. Since its spread across Africa, Asia, and parts of Europe, LSD has grown into a considerable veterinary challenge - particularly in low- and middle-income countries where rapid and accurate field diagnostics are lacking [9][3].

Despite that urgency, DL applications targeting LSD detection remain sparse relative to broader livestock disease detection efforts [14][23]. Much of the research is recent and fragmented, with limited datasets and a lack of generalizable, field-tested solutions.

Among the earliest attempts to detect LSD using image-based deep learning, [13] developed a CNN-based classification model trained on a small dataset of infected and healthy cattle. Using transfer learning with pretrained models like VGG16 and ResNet50, they reported promising accuracy levels exceeding 90% in controlled settings. However, the study acknowledged a lack of external validation and real-world deployment, which undermines model reliability in diverse farm environments.

Building on this foundation, Saha (2024) conducted a comparative study evaluating CNN architectures like MobileNet, InceptionV3, and ResNet variants on LSD images collected from dairy farms in South Asia. MobileNetV2 grew as the top performer, achieving 92.3% accuracy while remaining lightweight and efficient for edge computing. This study echoed the challenges of dataset imbalance - few infected cases compared to healthy ones - and showed variability in lesion morphology because of breed and environmental differences. It also advocated for standardized preprocessing strategies - including contrast improvement and region-of-interest cropping.

Muhammad Saqib et al. [9] Further refined the MobileNetV2 approach by introducing RMSProp optimization and dropout-based regularization. Their improvements yielded higher recall and reduced overfitting, validating the potential of lightweight models for deployment in remote or resource-limited areas. However, as with most studies in this domain, they operated on datasets of fewer than 1,000 annotated images and lacked performance validation in real-world field conditions, which restricts broader scalability [1].

Transformer-based models have only recently been introduced in the context of LSD detection. Genemo [3] applied Vision Transformers (ViTs) for geospatially driven risk prediction, leveraging visual and environmental data to identify high-risk zones for LSD outbreaks. While not directly focused on lesion classification, this work showed the adaptability of ViTs to multi-source data integration. Similarly, [17] conducted one of the first comparative benchmarks of pretrained models - including ViTs, EfficientNet, and DenseNet - on LSD lesion images. Their study found ViTs slightly outperformed CNNs in complex imaging scenarios involving cluttered backgrounds or subtle lesion texture differences, although they required considerably larger datasets and more computational power during training.

Despite those advances, several limitations continue to constrain model performance and field adoption. Firstly, most LSD datasets are private or localized, with few publicly available image repositories. That scarcity limits reproducibility and cross-study comparison [20][1]. Moreover, dataset homogeneity - resulting from images captured under similar lighting, angles, or animal poses - causing overfitting and poor generalization [27]. Geographic bias also persists, with most data originating from South Asia and Africa, making models less effective when deployed in new regions with different cattle breeds and environmental conditions [10][18].

Explainability remains another challenge. Since LSD can resemble other skin disorders like photosensitization, tick infestations, and dermatophilosis, reliance on opaque models without interpretability can increase misdiagnosis risks [22][3]. However, few studies offer attention maps (like Grad-CAM), saliency visualizations, or feature attribution tools to help veterinarians verify model outputs [7][19]. This is an impactful omission for clinical decision-making, especially in high-stakes disease control scenarios.

Some researchers have proposed mitigation strategies to overcome the gaps. Data augmentation remains the most widely used technique, with methods like flipping, rotation,

cropping, jittering, and synthetic image generation using GANs to expand dataset diversity [14][9][25]. Transfer learning has also been applied to leverage features from large human or animal health datasets; however, this raises concerns about domain mismatch, especially when models pretrained on urban facial datasets are adapted to rural livestock images [2][28].

Developing work suggests hybrid models may offer a solution. Those architectures combine CNNs' local feature detection with Transformers' long-range attention capabilities - leading to improved robustness and contextual awareness [12][21]. For instance, models like AnimalFormer and LAD-RCNN, though not yet tested specifically on LSD, have shown improved generalization across lighting and pose variations in other livestock contexts [20][11]. Adapting such architectures for LSD could improve accuracy and explainability.

In summary, deep learning approaches for LSD detection are promising but still maturing. CNNs like MobileNetV2 remain the dominant architecture because of their low-resource requirements and acceptable performance. ViTs are becoming preferred based on how they offer better interpretability and handling of complex data - though at the cost of training burden. Major gaps are present in relation to dataset availability, regional diversity, model transparency, and field deployment validation. Addressing those limitations will be necessary to realizing deep learning's full potential for LSD diagnosis in practical farm settings.

3.3 Comparative Analysis of Architectures

The use of deep learning in livestock disease diagnosis has changed considerably - with three primary architectural categories becoming dominant: CNNs, ViTs, and hybrid CNN–ViT models. Each architecture offers distinct advantages and trade-offs in terms of accuracy, computational cost, scalability, explainability, and suitability for real-world deployment [17][4][11].

3.3.1Convolutional Neural Networks (CNNs)

CNNs remained the most widely applied architecture in LSD image classification - with models like ResNet, EfficientNet, and DenseNet achieving classification accuracies ranging from 93.2% to 98.3% [14][17]. Those architectures performed particularly well on high-resolution, annotated datasets but their sensitivity to domain variability limited cross-geographical generalizability [13]. Models like VGG16, ResNet50, InceptionV3, and EfficientNet have shown solid performance in tasks ranging from body condition scoring [18] and cow identification [26], to sheep breed recognition [5] and pig weight estimation [25]

Muhammad Saqib et al. [9] applied MobileNetV2 with RMSProp optimization, reporting up to 93% classification accuracy using fewer than 1,000 annotated images. Saha [14] performed a comparative evaluation of various CNN architectures - including InceptionV3, ResNet34, and EfficientNet-B0 - and concluded that MobileNetV2 provided the best balance between speed and accuracy; making it suitable for deployment in low-resource environments.

CNNs, though, face limitations when dealing with cluttered backgrounds - a common issue in real-world LSD images. Their local receptive fields limit the capture of global context, and they may miss subtle lesion features or spatial dependencies [19]. CNNs do not natively support interpretability, though tools like Grad-CAM or LIME are usually used post hoc to visualize regions of interest [11] [22].

3.3.2 Vision Transformers (ViTs)

ViTs represent a newer family of deep learning models that utilize self-attention mechanisms to learn contextual relationships in different image regions. First introduced by Dosovitskiy et al. [4], ViTs have shown higher performance in tasks involving global feature understanding and spatial reasoning. Even though it was initially developed for large-scale datasets like ImageNet, their application in agriculture is increasing in relevance.

Zhang et al. [28] showed the effectiveness of a ViT-based facial recognition model for small-tailed Han sheep - achieving higher precision and robustness than standard CNNs. Tangirala et al. [22] similarly applied ViTs to detect behavioral anomalies in cattle, while Guo et al. [4] integrated a YOLOv5s–Vision Transformer pipeline to monitor cow feeding behavior.

Transformer-based architectures - particularly ViTs, Pyramid Vision Transformers (PVT), and optimized ViT variants - were less common but showed improved performance in multimodal and generalization-heavy tasks. For example, ViTs achieved F1-scores up to 95.1% when integrated with contextual data like breed or location metadata [11] [6]. That suggests higher semantic reasoning, especially in cross-breed scenarios or noisy environments.

In LSD diagnosis, Senthilkumar et al. [17] found that ViTs outperformed CNNs in scenes with visual clutter or ambiguous lesion presentations - showing better sensitivity in challenging imaging conditions. Nonetheless, ViTs are data-hungry and prone to overfitting when trained on small datasets - a frequent limitation in livestock diagnostics [1] [3].

ViTs offer the advantage of attention heatmaps for better explainability, helping clinicians interpret which regions of the image contributed to the decision. However, their computational demands and need for extensive GPU resources present barriers to use in field settings, especially in rural veterinary contexts [21] [10].

3.3.3 Hybrid CNN–ViT Architectures

Given the complementary strengths of CNNs (local feature extraction) and ViTs (global attention and reasoning), hybrid architectures have emerged as a compelling approach for livestock disease detection. Those models use CNNs in early layers to extract low-level spatial features, followed by Transformer blocks to process long-range dependencies and contextual interactions [12][8].

While not yet widely applied to LSD diagnosis, hybrid models have proven effective in other livestock applications. AnimalFormer, developed by Qazi et al. [12], integrates CNN and Transformer modules to process multimodal livestock data (images and video) for behavior recognition and health monitoring. Li et al. [18] Similarly introduced a hybrid architecture for sheep face recognition that combined MobileNetV2 and ViT blocks. It achieved balanced

performance with minimal training data - a potentially effective strategy for under-resourced disease domains like LSD.

Hybrid models like AnimalFormer and GasFormer - that integrate CNN feature extractors with transformer-based attention modules - showed high contextual awareness in challenging settings [12][16]. The models, though, were computationally intensive and usually required more training times and higher memory overhead.

Hybrid models also relate with or match recent trends in explainable AI (XAI). Many integrate attention visualization layers, feature fusion dashboards, or saliency maps - which improve transparency and boost end-user trust. That trust and transparency is important for veterinarians in the field who use AI guidance for treatment decisions [11] [11].

3.4 Comparative Performance Overview

Table 2 (below) provides a comparative summary of CNN, ViT, and hybrid models based on key metrics—accuracy, F1-score, model size, training time, and interpretability—across selected livestock disease tasks, emphasizing LSD detection where data is available. While CNNs remain the dominant model class due to their deploy ability, ViTs and hybrids are rapidly gaining ground as datasets grow and computational infrastructure improves.

Table 2: Comparative Summary

Model	Architecture	Dataset Size	Accuracy (%)	F1-Score	Deployment Suitability	Explainability
ResNet50	CNN	Approx. 1,000 images	91.2	0.89	Moderate	Medium (Grad-CAM
MobileNetV2	CNN (lightweight)	Approx. 900 images	93.1	0.91	High (Mobile)	Medium
ViT-B16	Vision Transformer	Approx. 3000 images	94.5	0.93	Low (Resource-Intensive)	High (Attention Maps)
Hybrid (CNN+ViT)	Hybrid	Approx. 2000 images	95.3	0.94	Medium	High
Animal Former	Hybrid	Multimodal	96.0	0.95	Medium-High	Very High

Several studies prioritized real-world deployment by implementing lightweight and edge-optimized models. MobileNetV2, Tiny-YOLO, and quantized ViT variants were among the most frequently cited architectures for mobile or embedded system deployment ([8] [9] [4]). The models achieved real-time inference speeds - with image processing times under 150 milliseconds - and classification accuracies ranging from 89.4% to 93.8% - making them well-suited for field conditions with limited computational infrastructure.

Despite progress in accuracy and deployment, only 7 out of 27 studies (26%) employed explainability methods like Grad-CAM, SHAP, LIME, or attention heatmaps ([11]; [10]). That represents a major gap in clinical interpretability - particularly when AI-driven diagnostic decisions must be trusted by veterinarians and integrated into disease surveillance frameworks [3] [1].

In summary, out of the 27 included studies, approximately 63% (n = 17) employed CNN-based models as their primary architecture, while 22% (n = 6) utilized pure transformer-based approaches like ViT or PVT. The remaining 15% (n = 4) explored hybrid or multimodal architectures [12] [22]. That distribution reflects a current dominance of CNNs - although recent trends show growing interest in transformer and hybrid frameworks; particularly for generalizability and robustness in non-ideal field conditions.

4. Conclusions and Recommendations

This systematic literature review examined the evolution and application of deep learning models in livestock disease image analysis, with a specific focus on the detection of Lumpy Skin Disease (LSD) in cattle. The analysis revealed that while a growing body of research has successfully leveraged deep learning—particularly CNNs, Vision Transformers, and hybrid models—to identify and classify a wide range of livestock conditions, significant gaps remain in the scalability, generalizability, and explainability of these models, especially in real-world or resource-constrained agricultural settings [17] [9] [3].

4.1 Summary of Key Findings

The application of CNNs continues to be observed in livestock health research because of their efficient learning on smaller datasets, and relatively low hardware requirements (22][18]. CNN-based architectures like MobileNetV2, ResNet50, and EfficientNet-B0 have shown high classification accuracy (for example, 91–93%) in LSD detection and general livestock disease diagnosis [14] [9]. Their reliance on local receptive fields and limited contextual reasoning, though, restrict their performance in complex lesion scenarios and when background noise is high [25][19].

In contrast, ViTs have shown promise in extracting holistic and long-range features because of their self-attention mechanisms [4]. ViTs have shown higher accuracy in tasks like facial recognition in sheep [27], behavior analysis in cattle [22], and lesion detection under noisy or occluded conditions [17]. Still, ViTs are computationally demanding and require large annotated datasets - which are usually unavailable in livestock disease contexts [1] [3].

Hybrid CNN–ViT models, like AnimalFormer [12] and those by [8], combine the local feature extraction of CNNs with the contextual reasoning of Transformers, thereby improving performance and interpretability. These models are being explored for spatially complex or semantic-intensive tasks and offer promising pathways for practical veterinary applications, although scalability and deployment barriers still persist [19].

The review also highlighted a general lack of large-scale, annotated, and geographically diverse datasets for LSD, which hampers model training, reproducibility, and benchmarking [1] [13]. Additionally, despite growing calls for model transparency, explainability remains underutilized across most studies, with few deploying tools like Grad-CAM, SHAP, or attention maps—key for field-level decision support [11] [17].

4.2 Identified Gaps in the Literature

4.2.1. Dataset Limitations and Imbalance

Most LSD datasets are small, proprietary, or geographically narrow, making generalization across breeds, regions, and image conditions difficult [14] [1]. Class imbalance—where healthy images far outnumber infected ones—further skews model accuracy and real-world relevance [9].

4.2.2. Lack of Model Explainability Tools

Only a few studies incorporate interpretability mechanisms such as Grad-CAM, attention visualization, or saliency maps [17] [10]. Given the clinical implications of false positives/negatives in disease detection, explainable AI (XAI) is essential for adoption by veterinarians and end-users.

4.2.3. Deployment Challenges

Despite strong lab results, many high-performing models are not optimized for edge deployment or low-power environments [9]. Very few have been implemented in rural contexts using smartphones, embedded GPUs, or cloud-accelerated APIs [25] [4].

4.2.4. Poor Domain Adaptation

Models trained on region-specific data often fail when transferred to new environments, due to differences in livestock breeds, lighting, image resolution, and disease presentation [13] [3]. Few studies explore domain generalization through transfer learning, federated learning, or cross-dataset validation [12].

4.3 Recommendations for Future Research

4.3.1. Develop and Share Open-Source, Annotated LSD Image Datasets

A global, collaborative effort is needed to build large, diverse, and annotated LSD datasets. Projects could mirror the efforts seen in body condition scoring [18] or pig weight estimation [25], enabling more robust benchmarking and model training.

4.3.2. Adopt Data Augmentation and Balancing Techniques

Studies should employ techniques such as Synthetic Minority Over-sampling Technique (SMOTE), flipping, rotation, synthetic lesion generation via GANs, and class-aware sampling to overcome imbalance and improve generalizability [14] [9].

4.3.3. Advance Explainable AI (XAI) in Livestock Diagnosis

Deep learning models must incorporate interpretability layers—such as Grad-CAM, SHAP, LIME, or attention maps—to support transparency and clinical validation [11] [10].

4.3.4. Design Models for Field Deployment

Lightweight models like Tiny-YOLO, MobileNetV2, and quantized ViTs should be optimized for edge devices including smartphones, Raspberry Pi units, and real-time APIs [9] [8][4].

4.3.5. Leverage Hybrid and Multimodal Architectures

Hybrid models that fuse CNNs and Transformers—such as AnimalFormer [12] and GasFormer [16]—can enhance robustness and semantic reasoning. Incorporating metadata such as breed, age, or environmental conditions may further boost diagnostic accuracy.

4.3.6. Incorporate Domain Adaptation Strategies

Domain adaptation through transfer learning, few-shot learning, or federated learning can improve generalization across diverse geographies and breeds ([13]; [12]). These should be emphasized in future architecture evaluations.

4.3.7. Integrate Deep Learning with Disease Surveillance Platforms

Linking DL-based diagnostic models with veterinary surveillance systems can improve early outbreak detection and inform policy interventions [3] [1]. This integration would help transition AI from experimental to operational impact in livestock management.

4.4 Final Remarks

The application of deep learning in livestock disease detection, particularly for LSD, is progressing rapidly. However, achieving widespread deployment and reliability requires a coordinated focus on dataset availability, model robustness, interpretability, and deployment readiness. By addressing the research and implementation gaps, deep learning can become a transformative tool in global animal health, contributing much to food security, animal welfare, and sustainable agriculture.

Disclosure: The authors certify that all content, data, and organizational references in this manuscript have been included with proper authorization and consent. They accept full legal and ethical responsibility for the accuracy, originality, and integrity of the work. The journal, publisher, and editorial board assume no liability for any disputes arising from the submitted material.

REFERENCES

[1] Afshari Safavi, E. (2022). Assessing machine learning techniques in forecasting lumpy skin disease occurrence based on meteorological and geospatial features. *Tropical Animal Health and Production*, *54*(1), 55. DOI:10.1007/s11250-022-03073-2

[2] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. https://arxiv.org/abs/2010.11929

[3] Genemo, M. (2023). Detecting high-risk area for lumpy skin disease in cattle using deep learning feature. *Advances in Artificial Intelligence Research*, *3*(1), 27-35. DOI:10.54569/aair.1164731

[4] Guo, Y., Hong, W., Wu, J., Huang, X., Qiao, Y., & Kong, H. (2023). Vision-based cow tracking and feeding monitoring for autonomous livestock farming: the YOLOv5s-CA+ DeepSORT-vision transformer. *IEEE Robotics & Automation Magazine*, *30*(4), 68-76. DOI:10.1109/MRA.2023.3310857

[5] Himel, G. M. S., Islam, M. M., & Rahaman, M. (2024). Vision intelligence for smart sheep farming: applying ensemble learning to detect sheep breeds. *Artificial Intelligence in Agriculture*, *11*, 1-12. DOI:10.1007/s43995-024-00089-7

[6] Khanal, R., Choi, Y., & Lee, J. (2024). Transforming poultry farming: A pyramid vision transformer approach for accurate chicken counting in smart farm environments. *Sensors*, *24*(10), 2977. DOI:10.3390/s24102977

[7] Li, X., & Liu, Y. (2024, September). Cow face recognition based on transformer group. In *Fourth International Conference on Computer Vision and Pattern Analysis (ICCPA 2024)* (Vol. 13256, pp. 203–209). SPIE. DOI: 10.1117/12.3038051

[8] Li, X., Du, J., Yang, J., & Li, S. (2022). When mobilenetv2 meets transformer: A balanced sheep face recognition model. *Agriculture*, *12*(8), 1126. https://doi.org/10.3390/agriculture12081126

[9] Muhammad Saqib, S., Iqbal, M., Tahar Ben Othman, M., Shahazad, T., Yasin Ghadi, Y., Al-Amro, S., & Mazhar, T. (2024). Lumpy skin disease diagnosis in cattle: A deep learning approach optimized with RMSProp and MobileNetV2. *PLOS ONE*, *19*(8), e0302862. DOI: 10.1371/journal.pone.0302862

[10] Pan, Y., Zhang, Y., Wang, X., Gao, X. X., & Hou, Z. (2023). Low-cost livestock sorting information management system based on deep learning. *Artificial Intelligence in Agriculture*, *9*, 110–126. DOI:10.1016/j.aiia.2023.08.007

[11] Pang, Y., Yu, W., Zhang, Y., Xuan, C., & Wu, P. (2023). An attentional residual feature fusion mechanism for sheep face recognition. *Scientific Reports*, *13*, 17128. DOI:10.1038/s41598-023-43580-2

[12] Qazi, A., Razzaq, T., & Iqbal, A. (2024). AnimalFormer: Multimodal Vision Framework for Behavior-based Precision Livestock Farming. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7973–7982). DOI:10.48550/arXiv.2406.09711

[13] Rai, G., Naveen, Hussain, A., Kumar, A., Ansari, A., & Khanduja, N. (2021). A deep learning approach to detect lumpy skin disease in cows. In *Computer Networks, Big Data and IoT: Proceedings of ICCBI 2020* (pp. 369–377). Springer. DOI:10.1007/978-981-16-0965-7_30

[14] Saha, D. K. (2024). An extensive investigation of convolutional neural network designs for the diagnosis of lumpy skin disease in dairy cows. *Heliyon*, *10*(14), e26049. DOI: 10.1016/j.heliyon.2024.e34242

[15] Sargeant, J. M., & O'Connor, A. M. (2020). Scoping reviews, systematic reviews, and meta-analysis: applications in veterinary medicine. *Frontiers in veterinary science*, *7*, 11. DOI: 10.3389/fvets.2020.00011

[16] Sarker, T. T., Embaby, M. G., Ahmed, K. R., & AbuGhazaleh, A. (2024). Gasformer: A transformer-based architecture for segmenting methane emissions from livestock in optical gas imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5489–5497). DOI:10.1109/CVPRW63382.2024.00558

[17] Senthilkumar, C., C, S., Vadivu, G., & Neethirajan, S. (2024). Early detection of lumpy skin disease in cattle using deep learning—a comparative analysis of pretrained models. *Veterinary Sciences*, *11*(10), 510. https://doi.org/10.3390/vetsci11100510

[18] Siachos, N., Lennox, M., Anagnostopoulos, A., Griffiths, B. E., Neary, J. M., Smith, R. F., & Oikonomou, G. (2024). Development and validation of a fully automated 2-dimensional imaging system generating body condition scores for dairy cows using machine learning. *Journal of Dairy Science*, *107*(4), 2499–2511. DOI: 10.3168/jds.2023-23894

[19] Souza, J. S., Bedin, E., Higa, G. T. H., Loebens, N., & Pistori, H. (2024). Pig aggression classification using CNN, Transformers and Recurrent Networks. *arXiv preprint arXiv:2403.08528*. DOI:10.5753/wvc.2024.34004

[20] Sun, L., Liu, G., Yang, H., Jiang, X., Liu, J., Wang, X., ... & Yang, S. (2023). LAD-RCNN: a powerful tool for livestock face detection and normalization. *Animals*, *13*(9), 1446. https://doi.org/10.3390/ani13091446

[21] Taiwo, G., Vadera, S., & Alameer, A. (2025). Vision transformers for automated detection of pig interactions in groups. *Smart Agricultural Technology*, *10*, 100774. https://doi.org/10.1016/j.atech.2024.100774

[22] Tangirala, B., Bhandari, I., Laszlo, D., Gupta, D. K., Thomas, R. M., & Arya, D. (2021). Livestock monitoring with transformer. *arXiv preprint arXiv:2111.00801*. https://arxiv.org/abs/2111.00801

[23] Temenos, A., Voulodimos, A., Korelidou, V., Gelasakis, A., Kalogeras, D., Doulamis, A., & Doulamis, N. (2024). Goat-CNN: A lightweight convolutional neural network for pose-independent body condition score estimation in goats. *Journal of Agriculture and Food Research*, *16*, 101174. https://doi.org/10.1016/j.jafr.2024.101174

[24] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*. https://doi.org/10.48550/arXiv.1706.03762

[25] Xie, C., Cang, Y., Lou, X., Xiao, H., Xu, X., Li, X., & Zhou, W. (2024). A novel approach based on a modified mask R-CNN for the weight prediction of live pigs. *Artificial Intelligence in Agriculture*, *12*, 19–28. https://doi.org/10.1016/j.aiia.2024.03.001

[26] Yukun, S., Pengju, H., Yujie, W., Ziqi, C., Yang, L., Baisheng, D., ... & Yonggen, Z. (2019). Automatic monitoring system for individual dairy cows based on a deep learning framework that provides identification via body parts and estimation of body condition score. *Journal*

of Dairy Science, *102*(11), 10140–10151. DOI: 10.3168/jds.2018-16164

[27] Zhang, X., Xuan, C., Ma, Y., & Su, H. (2023). A high-precision facial recognition method for small-tailed Han sheep based on an optimised Vision Transformer. *Animal*, *17*(8), 100886. https://doi.org/10.1016/j.animal.2023.100886

[28] Zhang, Y., Zhang, Y., Jiang, H., Du, H., Xue, A., & Shen, W. (2024). New method for modeling digital twin behavior perception of cows: Cow daily behavior recognition based on multimodal data. *Computers and Electronics in Agriculture*, *226*, 109426. https://doi.org/10.1016/j.compag.2024.109426