Veterinary Medicine and Science

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Veterinary Eduction

Implementation of a Deep Learning System for Detection and Classification of Lumpy Skin Disease in Cattle: Enhancing Precision and Efficiency in Veterinary Diagnostics

Madhumita Pal[1] | Soudamini Behera[2] | Ranjan K. Mohapatra[3] | Francesco Branda[4] | Gaber A. M. Mersal[5] | Snehasish Mishra[6] | Salah M. El-Bahy[7] | Gurudutta Pattnaik[8] | Sovan Pattanaik[9] | Md Sajid Ali[10] | Amiyakanta Mishra[11] | Lawrence Sena Tuglo[12] | Mona Youssef[13]

[1]Department of Electrical Engineering, Government College of Engineering, Keonjhar, Odisha, India | [2]Department of Electrical Engineering, Government College of Engineering, Kalahandi, Odisha, India | [3]Department of Chemistry, Government College of Engineering, Keonjhar, Odisha, India | [4]Unit of Medical Statistics and Molecular Epidemiology, Università Campus Bio-Medico di Roma, Rome, Italy | [5]Department of Chemistry, College of Science, Taif University, Taif, Saudi Arabia | [6]School of Biotechnology, Campus-11, KIIT Deemed-to-be-University, Bhubaneswar, Odisha, India | [7]Department of Chemistry, Turabah University College, Taif University, Taif, Saudi Arabia | [8]School of Pharmacy and Life Sciences, Centurion University of Technology and Management, Khordha, Odisha, India | [9]School of Pharmaceutical Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India | [10]Department of Pharmaceutics, College of Pharmacy, Jazan University, Jazan, Saudi Arabia | [11]College of Pharmaceutical Sciences, Puri, Odisha, India | [12]Department of Nutrition and Dietetics, School of Allied Health Sciences, University of Health and Allied Sciences, Ho, Ghana | [13]Department of Hepatology, Gastroenterology and Infectious Diseases, Benha Teaching Hospital, Benha, Egypt

Correspondence: Lawrence Sena Tuglo (senatuglo@gmail.com)

Received: 14 March 2025 | **Revised:** 24 August 2025 | **Accepted:** 29 September 2025

Funding: This research was funded by Taif University, Saudi Arabia, Project No. (TU-DSPP-2024-21).

Keywords: ConvMixer | deep learning | ensemble | lumpy skin disease | vision transformer

ABSTRACT: Lumpy skin disease (LSD) poses significant threats to cattle health, mainly in the high bovine density locations, leading to severe agro-economic consequences. The present study is a maiden attempt to introduce deep learning (DL)-based systems to detect and classify LSD, leveraging two cutting-edge models: vision transformer (ViT) and ConvMixer. A comprehensive dataset of LSD lesions was developed, incorporating extensive data augmentation techniques for enhanced model robustness and generalization. The approach included the creation of an ensemble model that combined the strengths of both ViT and ConvMixer for improved accuracy of the diagnosis. The models were rigorously evaluated on the basis of the accuracy, recall, precision, *F*1-score and area under the receiver operating characteristic curve (AUC-ROC). The obtained results have demonstrated that the ensemble model outperformed the individual models significantly, achieving high precision and recall rates. The study showed the potentials of advanced DL techniques as reliable and efficient tools in veterinary diagnostics, offering early detection and management interventions to counter LSD outbreaks. Implementing these models could facilitate timely interventions to mitigate

Abbreviations: ARIMA, auto-regressive moving average; AUC-ROC, area under the receiver operating characteristic curve; AuNPs, gold nanoparticles; COVID-19, coronavirus disease-2019; ELISA, enzyme-linked immunosorbent assay; FAT, fluorescent antibody technique; FPR, false positive rate; GANs, generative adversarial networks; iELISA, indirect enzyme-linked immunosorbent assay; LAMP, loop-mediated isothermal amplification; LFIA, lateral flow immunoassay; LSDV, lumpy skin disease virus; mAbs, monoclonal antibodies; MHSA, multi-head self-attention; NGS, next-generation sequencing; NNAR, neural network auto-regressive; RPA, recombinase polymerase amplification; rt-qPCR, real-time quantitative PCR; TPR, true positive rate; WGS, whole genome sequencing.

Madhumita Pal, Soudamini Behera and Ranjan K. Mohapatra contributed equally and will be treated as joint first authors.

the economic impacts of LSD and improve livestock health management measures. This study opens up the scope to diagnose and classify numerous other neglected but critical bovine diseases by using similar digital platforms and employing new-age computerized tools such as artificial intelligence (AI) and machine learning (ML). To further enhance the practical applicability of DL-based diagnostic systems in future, investigations need to focus on expanding and enriching the datasets, refining model architectures and exploring real-time deployment.

1 | Introduction

Home to the major bovine population and a chief milk producer in the world, a severe Lumpy skin disease (LSD) outbreak was witnessed in India in the late 2022. LSD impacted mostly the northern regions, especially in the Rajasthan and Gujarat provinces. It led to the infection of more than 2.4 million cattle and more than 110,000 deaths (BBC 2022), which had significant economic repercussions and threatened food security. The livelihoods of livestock farmers were severely affected, and it raised concerns about the production and consumption patterns of milk among the public locally. Although there was concern that calves could contract LSD from infected lactating mothers, the Food and Agriculture Organization (FAO) of the United Nations assured that the disease was not zoonotic and did not pose human health risks.

The lumpy skin disease virus (LSDV) is part of the Chondropoxvirinae subfamily within the Poxviridae family and primarily affects goats and cattle. It belongs to the genus *Capripoxvirus* and is antigenically similar to sheeppox and goatpox viruses. Although LSDV has been known for decades as the causative agent of LSD in cattle, the recent emergence of a more virulent strain has alarmed veterinarians, animal health experts, administrators and the public. In the wake of the coronavirus disease-2019 (COVID-19) pandemic and the re-emergence of the mpox virus (Mohapatra et al. 2023; Mohapatra, Mahal, et al. 2024), the rise in LSD cases has prompted government action.

Although LSD does not pose a direct threat to human health, its potential to infect other domestic and wild animals remains a serious concern, particularly among farmers and veterinarians. The disease primarily affects cattle and water buffaloes and is characterized by high morbidity but typically low mortality. LSD is primarily transmitted through blood-feeding insects such as certain species of flies, mosquitoes and ticks, although the exact modes of transmission are not fully understood. Research by Paslarua et al. (2021) highlighted the role of the cosmopolitan fly species *Stomoxys calcitrans* in potentially spreading LSDV. Clinically, LSD presents with distinctive nodules on the skin and mucous membranes of the mouth, udder, nostrils, genitalia and rectum, often accompanied by fever and anorexia. As the disease progresses, it can become systemic, leading to lesions in the mouth and upper respiratory tract, which may cause anorexia, fever, reduced milk production (dysgalactia) and pneumonia (Davies 1991). The disease can also result in reduced fertility, abortion and, in severe cases, death (Gupta et al. 2020; Azeem et al. 2022). The incubation period ranges from 1 to 4 weeks. The persistence of LSDV, combined with the widespread distribution and long-range flight capabilities of *S. calcitrans*, poses significant challenges for controlling outbreaks and implementing effective containment measures.

Although LSD was first identified in Zambia in 1929, its origins remain uncertain. Initially endemic to sub-Saharan Africa, the disease gradually spread to North Africa, Europe, the Middle East and Asia. LSD was first detected in Asia in 2019, with reports emerging from north-west China, Bangladesh and India (Sudhakar et al. 2022). LSD expanded to Southeast Asia, mostly during the 2020 summer. Genome of the responsible virus from 2022 Indian outbreak was analysed as having significant genetic variations as compared to the reference genome, which indicated a distinct lineage (Bhatt et al. 2022; The Indian Express 2022). Despite the slow global spread of LSD over about nine decades, it could potentially evolve to a significant pandemic. It underscores the urgent need of robust monitoring and surveillance, alongside a comprehensive vaccination program aligned with 'One Health', to effectively manage and mitigate the outbreak risks in future (Mohapatra, Mishra, et al. 2024).

Significant transboundary introduction and spread of animal diseases among exotic animals is worrisome. Diseases like African swine fever, foot-and-mouth disease, avian influenza and peste des petits ruminants (PPR) are increasingly spreading across borders due to the uncontrolled globalisation (Gongal et al. 2022). Thus, robust cross-border animal movement management strategy is crucial given the severe economic repercussions and potential socio-economic impacts. Current transboundary animal disease outbreaks could radically disrupt the global cattle industry. Anwar et al. (2022) used auto-regressive moving average (ARIMA) and neural network auto-regressive (NNAR) models to predict a sharp rise in LSD as early as 2023–2024. Alkhamis and VanderWaal (2016) analysed the outbreak data from 2012 to 2015 and identified Israel and Turkey as potential LSD outbreak high-risk areas. Key environmental factors contributing to a favourable ecological niche for LSDV include land cover, mean diurnal temperature range, annual precipitation, type of livestock production system and livestock density. These insights are valuable for developing risk-based surveillance and control programs and for formulating effective epidemic preparedness strategies in regions adjacent to LSDV-free areas.

Furthermore, effective control of LSD depends not only on vaccination and vector management but also critically on early and accurate diagnosis. Unfortunately, current diagnostic practices have significant shortcomings. In the field, veterinarians typically rely on visual inspection of characteristic skin nodules—a subjective, error-prone approach that struggles with early or mild cases and can easily be confounded by other dermatological conditions. Delayed or inaccurate clinical diagnosis allows further disease

Summary

- Lumpy skin lesion images collected, and the LSD dataset is prepared for public use.
- Detecting and classifying lumpy skin lesions in cattle.
- Implementing DL models to classify lumpy skin lesions.
- Evaluated the efficacy of two select DL models using the balance or imbalance dataset.
- Measured performance of DL models on accuracy, precision, AUC, recall, *F*1-score, precision–recall curve and run time.

spread before control measures can be implemented. The main diagnostic methods for LSD are molecular methods, histology and clinical symptoms. Although more testing is necessary for confirmation, clinical indications, like as fever, skin nodules, lymph node enlargement, mucous membrane and internal organ oedema, can point to LSD. In order to make a laboratory diagnosis of LSD, clinical signs and symptoms are used to raise first suspicion. Viral, molecular and immunohistopathological investigations, including necropsy, virus identification and serological tests, are then used to potentially confirm the diagnosis (Amin et al. 2021; Datten et al. 2023). In order to facilitate prompt detection in outbreak-like scenarios, two monoclonal antibodies (mAbs) that target distinct epitopes of the LSD viral structural protein p32 and AuNPs using a colorimetric sandwich-type lateral flow immunoassay (LFIA) were recently evaluated as rapid diagnostic tools (Cavalera et al. 2022). Molecular techniques like as real-time quantitative PCR (rt-qPCR) and multiplex PCR (mPCR) could be used to characterize the LSD virus from clinical specimens (Zeedan et al. 2024). The identification of mutation events between haplogroups and the continual genetic diversification of the virus, which aligns with established patterns of propagation, can be facilitated by whole genome sequencing (WGS) or next-generation sequencing (NGS) of LSD (Bayyappa et al. 2025).

Newer diagnostic modalities, like immunoperoxidase monolayer assay, high resolution melting assay, nanopore sequencing, TrippleE, immunohistochemical (IHC) and recombinase polymerase amplification (RPA)-Cas12a-fluorescence assay, were found useful in the diagnosis of LSD. Other methods, like RPA assay, loop-mediated isothermal amplification (LAMP), fluorescent antibody technique (FAT), indirect FAT (iFAT), enzyme-linked immunosorbent assay (ELISA) and indirect enzyme-linked immunosorbent assay (iELISA), are also available for diagnosis (Farag et al. 2025). IHC, or immunohistochemistry, is a crucial diagnostic technique for LSD. Using particular anti-LSD virus antibodies, IHC is a direct technique for identifying the pathogenic antigen in the tissue samples obtained from nodules of infected cattle (Sanz-Bernardo et al. 2020). The most reliable method of diagnosis is LSD virus isolation. It is time-consuming, nevertheless, as the procedure to isolate the LSD virus in cell cultures or the chorioallantoic membrane (CAM) of an embryonated chicken egg (ECE) takes several weeks (Awad et al. 2010). Although the lab-based molecular assay PCR is highly specific and sensitive, it is costly, needs specialized equipment and trained personnel and is often inaccessible in remote or resource-limited region. Its turnaround times can be too long for rapid outbreak response, and scaling it for large surveillance efforts is challenging. These limitations create a diagnostic gap that hinders effective LSD management, leading to delayed quarantines, inefficient vaccination strategies and ultimately greater economic and animal health impacts.

To address these challenges, there is a critical need for diagnostic tools that are accurate, rapid, accessible and easy to deploy in the field. ML (machine learning) and deep learning (DL) models have become integral to the diagnosis and classification of a wide range of diseases (Pal, Parija, Mohapatra, et al. 2022; Pal, Tiwari, et al. 2022; Pal, Parija, Panda, et al. 2022; Pal, Parija, et al. 2023; Pal, Mahal, et al. 2023). Recent advances in artificial intelligence (AI) have shown great promise in automating medical and veterinary imaging diagnostics. The DL models can learn complex patterns in images, achieving high diagnostic precision and consistency even in resource-constrained settings. Despite their success in other veterinary applications, advanced DL techniques have not yet been systematically applied to LSD lesion classification. This represents an important knowledge and technological gap in veterinary diagnostics.

Our study aims to fill this gap by developing and evaluating a novel DL-based diagnostic system for LSD lesion detection and classification. We introduce, for the first time in this context, state-of-the-art architectures such as vision transformer (ViT) and ConvMixer, individually and in an ensemble model that combines their strengths for improved accuracy and robustness. To support model training and generalization, we curated and extensively augmented a comprehensive dataset of LSD skin lesion images, addressing the common challenge of class imbalance in veterinary imaging.

The novelty of this research lies in the fact that it focuses on achieving high-precision diagnosis by addressing a critical gap in the existing literature. By delivering an automated, precise and efficient diagnostic tool, our approach has the potential to support veterinarians in early detection of LSD cases, enable faster and more targeted interventions, improve outbreak management and reduce the economic burden of the disease. This work sets a new standard for applying AI technologies to veterinary diagnostics and offers a framework that could be extended to other transboundary animal diseases in the future.

2 | Methodology

2.1 | Data Collection

We adopted a multi-stage approach to develop and validate DL models to automate LSD detection in cattle, as shown in Figure 1. During the *data collection* step, we assembled a comprehensive dataset of lumpy skin images sourced from various online repositories as well as from direct photographic captures in the field. To guarantee the quality and relevance of the dataset, stringent inclusion and exclusion criteria were applied: only images explicitly labelled or verified as confirmed LSD cases—such as those derived from government outbreak reports, peer-reviewed publications or official veterinary health agency

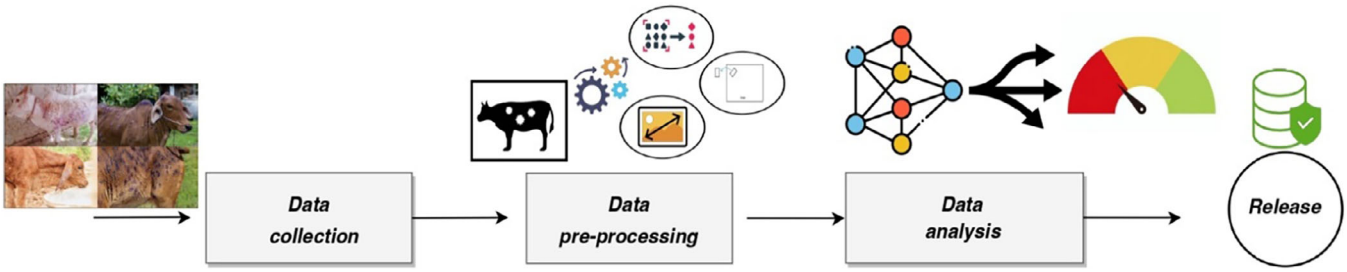

FIGURE 1 | The workflow of the proposed deep learning pipeline for LSD lesion classification. The process includes data collection (from online and field sources), preprocessing and augmentation, balancing (via upsampling), model training with ViT and ConvMixer architectures and ensemble fusion. The pipeline aims to detect and classify skin lesions into three categories: normal, mild and severe.

records—were retained. Images with uncertain or ambiguous diagnostic information were excluded. Furthermore, all images underwent systematic visual screening to confirm consistency with typical LSD lesion morphology, referencing established veterinary guidelines (Food and Agriculture Organization of the United Nations 2017).

The collected images exhibited variable resolutions, typically ranging from 480 × 360 to 1280 × 960 pixels. To standardize input for the models, all images were resized to 128 × 128 × 3 pixels. Annotation was performed manually, categorizing lesions into three severity levels—normal, mild and severe—based on clinical definitions of LSD lesion progression. This labelling was carried out by a veterinary expert with extensive field experience in LSD diagnosis. To enhance annotation reliability, a second independent annotator reviewed a randomly selected 20% subset of the dataset, and any disagreements were resolved through consensus discussions.

In terms of ethical considerations and data provenance, images obtained from online sources were exclusively drawn from open-access veterinary datasets, public outbreak reports and official documentation permitting academic use and redistribution. The subset of images captured in the field was collected without handling or harming animals, and verbal consent was obtained from livestock owners prior to image acquisition. Given that the study utilized only anonymized, non-invasive visual data without any experimental intervention, formal animal ethics committee approval was not required, in accordance with institutional policies. These protocols were established to uphold ethical standards, ensure transparency and maintain annotation validity, thereby supporting reproducible and responsible AI research.

2.2 | Data Pre-Processing

This step was critical to improve the quality of the data. Data augmentation techniques were used to expand the training set for the model effectively and improve generalization. The techniques included normalization to standardize pixel intensity, uniform scaling to ensure consistent dimensions, random horizontal flipping to introduce reflection invariance, subtle random rotations to simulate variation in camera angles and random zooming to represent real-world distance variations. In addition, to address the class imbalance and prevent skewed model performance, an up-sampling strategy was implemented to equalize the representation of each class within the dataset, ensuring that the number of samples for each class was balanced. By duplicating instances of under-represented classes, the up-sampling method aimed to provide a more equitable distribution of samples that helped the model learn more effectively from all classes for improved the overall performance. This approach mitigates potential biases that could arise from imbalanced data and enhances the ability of the model to generalize across different categories of LSD lesions. The dataset was split into two subsets for training (80%) and evaluation/testing (20%). This division was chosen to ensure that the models were trained on a diverse sample and tested on previously unseen data, thereby assessing generalization performance. To analyse the dataset, advanced DL architectures were employed, aiming at automating the classification of LSD lesions, that involved assigning each image to one of the three predefined categories: 'normal', 'mild' and 'severe' (Figure 2).

2.3 | Data Analysis

For this purpose, we implemented two state-of-the-art models: the ViT, known for capturing long-range dependencies in image data through self-attention mechanisms (Dosovitskiy et al. 2021), and ConvMixer, which combines the strengths of conventional convolutional layers with modern mixing techniques for efficient spatial feature extraction (Trockman and Kolter 2022). These models were selected over traditional convolutional neural network (CNN) architectures (e.g., ResNet, DenseNet and EfficientNet) on the basis of their superior capacity to model both global and local visual patterns, which is critical in distinguishing between subtle lesion severity levels. ViT excels at capturing global context across the entire image—a key factor when lesion spread and morphology must be assessed holistically—as demonstrated in medical image classification tasks requiring fine-grained reasoning (Chen et al. 2021; Hatamizadeh et al. 2022). Conversely, ConvMixer retains CNN-like inductive biases but enhances spatial generalization via patch-wise mixing, offering high performance with reduced computational complexity (Trockman and Kolter 2022; Mehta and Rastegari 2022).

To enhance predictive accuracy and robustness, we also developed an ensemble model that integrates ViT and ConvMixer, leveraging their complementary characteristics. This

FIGURE 2 | Visual examples of three classes of images used for the study. First: normal healthy cattle skin (smooth and uniform in texture and colour). Middle: mild to moderate lesions with early stage nodules. Last: severe lesions with extensive dermal damage. The LSD lesions present as firm nodules on the skin, with varying size. It may also be accompanied by other symptoms like scabs, oedema and potential ulceration.

combination was specifically designed to mitigate the limitations of standalone CNNs, which may struggle to differentiate between visually similar lesion categories due to their reliance on hierarchical local features.

The ensemble approach proved particularly effective in achieving balanced performance across all lesion categories, especially in detecting mild and severe cases, which are often more difficult to classify accurately. Model performance was rigorously assessed using standard evaluation metrics, including accuracy, recall, precision, *F*1-score and area under the receiver operating characteristic curve (AUC-ROC). These metrics allowed us to comprehensively compare the individual and combined models, highlighting the diagnostic strengths of the ensemble model in classifying varying severity levels of LSD. As illustrated in Figure 3, representative examples from each class were used to train and validate the models, providing a visual reference for lesion severity levels.

Our analyses were performed using Python, leveraging its vast ecosystem of open-source libraries within the TensorFlow environment. Simulations were conducted on Google Colab, a cloud-based platform that offers robust computational resources and easy access to GPU acceleration, making it ideal for DL projects. Key libraries used in this project include Pandas for data manipulation and analysis; Keras, integrated with TensorFlow, which facilitates the construction and training of DL models; and NumPy, essential for numerical computations and manipulation of large multidimensional arrays. OpenCV (cv2) was essential for image processing tasks, such as reading, scaling and augmenting images. Seaborn and Matplotlib were used for data visualization, providing tools to create informative graphs to help understand data distribution and model performance. Scikit-learn (sklearn) provided tools for breaking down datasets, evaluating models and calculating metrics. The Glob utility assists in file management by retrieving file paths that correspond to specified models, whereas LabelEncoder (from sklearn.preprocessing) converts categorical labels to numerical values for model input. Pathlib has simplified filesystem path management and file access operations. All data and scripts are openly available at the following link: https://zenodo.org/records/11003967.

2.4 | DL Models Used in the Research

2.4.1 | Vision Transformer

The ViT processes input images by splitting them into fixed-size patches, which are then linearly embedded and enriched with positional encodings to preserve spatial relationships (Pal, Branda, et al. 2025; Pal, Mohapatra, et al. 2025). These patch embeddings are treated as input tokens and passed through a series of transformer blocks, each composed of multi-head self-attention (MHSA) and feed-forward MLP layers, with layer normalization and dropout for regularization. ViT departs from traditional CNNs by modelling global dependencies via self-attention, making it effective in capturing contextual relationships across the entire image—an essential capability for detecting subtle morphological differences in lesion severity (Kameswari et al. 2023). This global receptive field allows the model to evaluate the overall distribution and progression of skin nodules, which is difficult for purely local filters. Originally proposed for natural language processing, the transformer architecture has demonstrated strong performance on image classification tasks when sufficient data and regularization are used (Han et al. 2023). In this study, ViT was configured with moderate complexity (see Table 1) to balance model capacity and training stability given the size of the LSD dataset. Key hyperparameters include eight transformer layers, four attention heads and an MLP head with 2048 and 1024 U.

Figure 4 illustrates the patch encoding process employed in the ViT architecture. The image is divided into a grid of non-overlapping patches, as seen in panel (b), which serves as the foundational input to the model. The patch encoder layer performs two critical operations: (i) Linear projection: Each image patch is flattened and linearly projected into a lower dimensional embedding space. This projection is achieved through a learnable weight matrix, transforming the raw pixel values of each patch into a dense vector representation of size projection_dim. This step effectively reduces the dimensionality of the input while preserving essential visual information. (ii) Position embedding addition: To retain spatial information that would otherwise be lost in the linear projection, learnable position embeddings are added to each patch embedding. These position embeddings

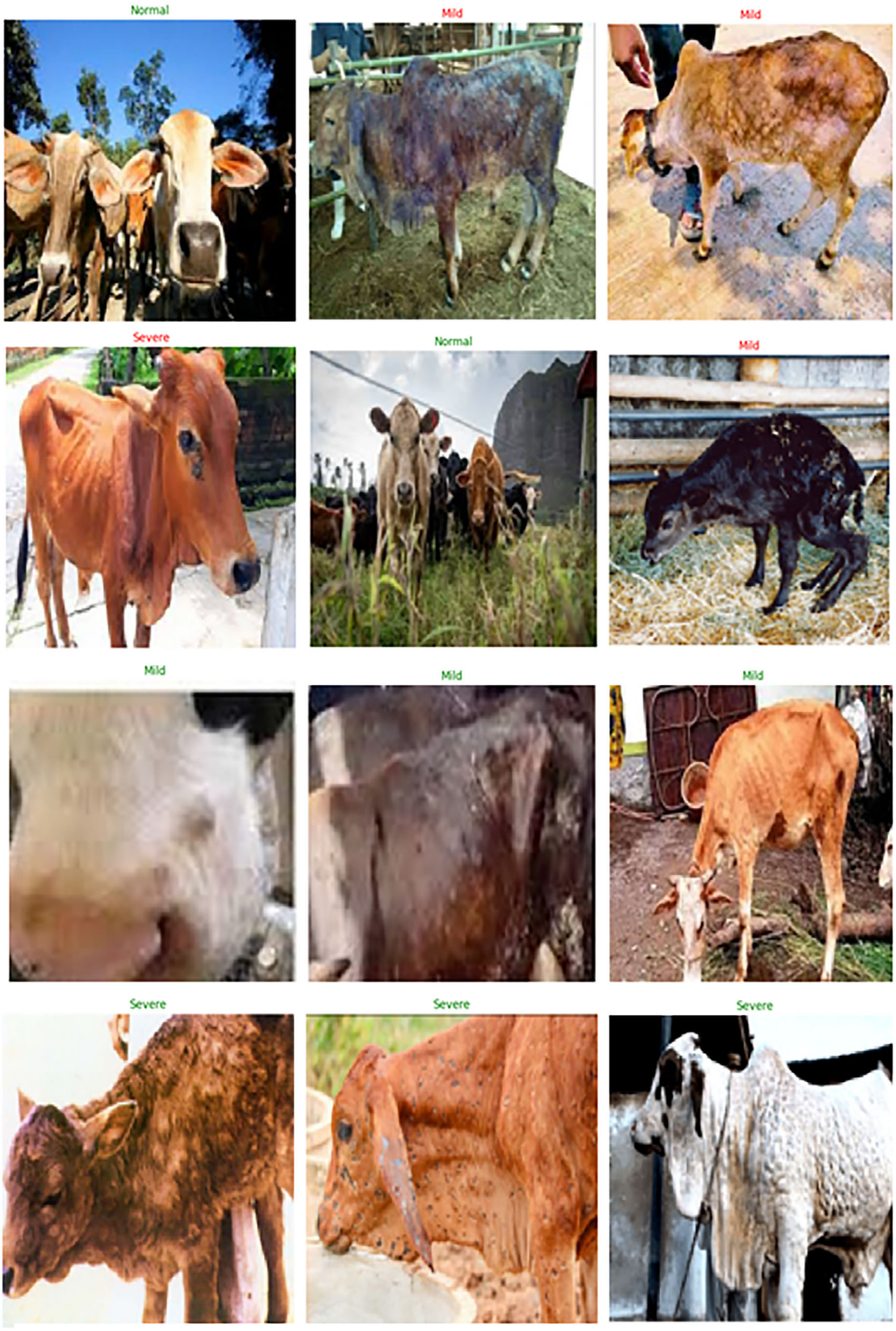

FIGURE 3 | Visual examples of LSD severity classes obtained from the output files. The grid displays 12 images showcasing various LSD infection stages. Each image corresponds to one of the three target classes (normal, mild and severe), helping the model learn visual differences between the severity levels.

TABLE 1 | Hyper-parameters used to construct the vision transformer (ViT) model.

Parameter	Value
Learning rate	0.001
Weight decay	0.0001
Batch size	161
Number of epochs	100
Image size	$128 \times 128 \times 3$
Patch size	6
Patches per image	144
Elements per patch	108
Number of heads	4
Transformer units	8
Transformer layers	8
MLP head units	2048 and 1024
Optimiser	Adam
Loss function	Sparse categorical cross-entropy
Drop out	0.5

encode the relative or absolute position of each patch within the original image, allowing the model to understand spatial relationships between patches. The resulting output is a sequence of patch embeddings, each enriched with positional information. This representation allows the subsequent transformer layers to process the image as a sequence of tokens, similar to how words are processed in natural language tasks. The combination of content information (from the linear projection) and positional information enables the model to capture both local features within patches and global relationships across the entire image.

2.4.2 | ConvMixer

ConvMixer combines the advantages of patch-based image representation with the inductive biases of convolutional networks (Pal, Mohapatra, et al. 2025). The model begins by embedding the input image into patches using a convolutional layer with large kernel size and stride equal to the patch size. This produces a spatially structured feature map, which is then processed through a series of ConvMixer blocks. Each block applies depthwise convolutions to independently extract spatial features within each channel, followed by pointwise convolutions (1×1) to mix information across channels. Non-linear activation functions (e.g., ReLU) and batch normalization follow each step, promoting convergence and regularization. The use of large convolution kernels enables the model to capture extended spatial context, whereas the uniform representation size across layers simplifies the architecture. Unlike ViT, which relies on global attention mechanisms, ConvMixer exploits local patterns efficiently with fewer parameters and lower computational cost. This makes it particularly suitable for mobile or low-resource veterinary settings. The model's performance in our study confirms its ability to extract discriminative features across lesion severity classes while maintaining training efficiency.

2.4.3 | Ensemble Model Architecture

The proposed ensemble model integrates two complementary DL architectures—ViT and ConvMixer—to leverage their individual strengths and improve overall classification performance, as schematized in Figure 5. This ensemble approach is designed to enhance robustness, accuracy and generalization capabilities in detecting and classifying LSD lesions, addressing the challenges posed by intra-class variability and subtle inter-class differences. The ensemble system operated as follows:

- Independent training: Both ViT and ConvMixer models are trained independently on the same training dataset, employing identical preprocessing pipelines, data augmentation techniques and class balancing strategies to mitigate class imbalance issues. By training the models separately, the ensemble benefits from diverse feature representations and decision boundaries, which is crucial for reducing correlated errors and improving generalization. This diversity enables the ensemble to capture both global contex-

FIGURE 4 | Visualization of the patch coding process in ViT: (a) original image of cattle and (b) grid representation showing the image divided into non-overlapping patches, illustrating the first stage of ViT's patch encoding mechanism. Each grid cell represented a patch that was encoded and embedded with positional information independently before being processed by the transformation layers.

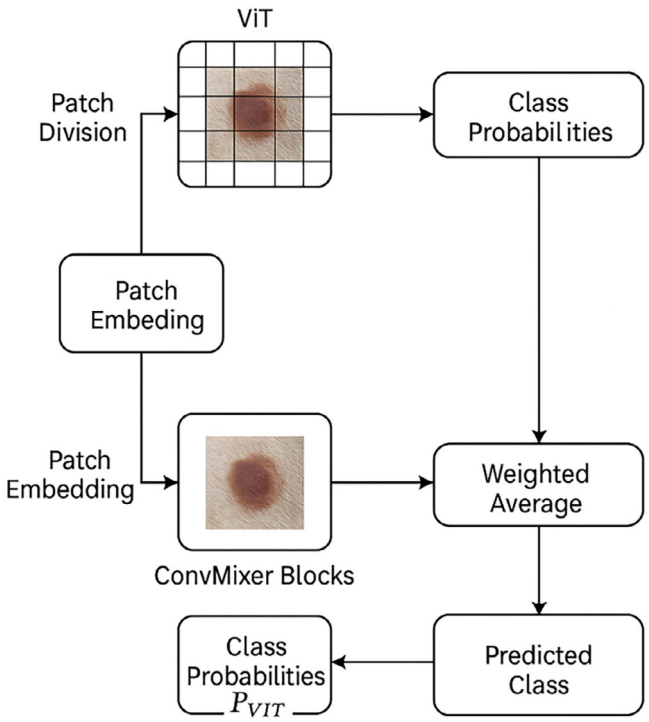

FIGURE 5 | Ensemble architecture of the proposed model. The input image is processed in parallel by two complementary deep networks: the ViT and the ConvMixer. Each branch extracts features and produces a probability distribution over the classes. The predictions are then combined by weighted-average fusion, generating the final prediction that assigns the most likely class to the LSD. ViT, vision transformer.

tual information (via ViT) and fine-grained local patterns (via ConvMixer).

- Feature extraction and prediction: During inference, each model processes the input images separately. The ViT model divides the input image into fixed-size patches, projects them into an embedding space and passes these embeddings through a series of transformer layers to capture long-range dependencies and global context. Simultaneously, ConvMixer employs convolutional layers designed to efficiently mix spatial and channel information, emphasizing local texture and pattern recognition. Both models output class probability distributions through their respective classification heads, representing the likelihood of each lesion severity class.
- Prediction fusion: To combine the complementary predictions from the two models, a weighted average fusion strategy is employed. The individual class probability vectors, $(P_{\text{ViT}}(x))$ and $(P_{\text{ConvMixer}}(x))$, are aggregated with weights (w_{ViT}) and $(w_{\text{ConvMixer}})$, respectively, satisfying $w_{\text{ViT}} + w_{\text{ConvMixer}} = 1$.

These weights are empirically optimized via cross-validation on the validation set, ensuring the fusion maximizes classification performance metrics such as accuracy and *F*1-score. The ensemble prediction $(P_{\text{ensemble}}(x))$ for an input image x is formally computed as

$$P_{\text{ensemble}}(x) = w_{\text{ViT}} \times P_{\text{ViT}}(x) + w_{\text{ConvMixer}} \times P_{\text{ConvMixer}}(x)$$

- Final classification: The predicted label is selected as the class corresponding to the highest probability in the fused vector $(P_{\text{ensemble}}(x))$. This fusion process effectively balances the complementary strengths of both models, resulting in more reliable and accurate classification outcomes.

This ensemble methodology balances the global context modelling capacity of the ViT with the local feature extraction efficiency of ConvMixer, leading to improved discrimination between lesion severity classes. Moreover, by combining two distinct architectures, the ensemble mitigates individual model biases and reduces the risk of overfitting, contributing to more stable and generalizable performance across diverse lesion presentations.

2.4.4 | Performance Parameters Used for Evaluation of DL Models

The definition of performance parameters used to evaluate DL models is summarized in Table 2.

3 | Results and Discussion

3.1 | Performance Evaluation of DL Models for an Imbalanced Dataset

The dataset comprises a total of 1019 images divided into three categories, as shown in Figure 6: normal (697 images), mild (231 images) and severe (91 images). After data augmentation, we have increased the images to 2091. The distribution is notably imbalanced, with normal skin lesions making up about 68% of the samples, mild lesions representing approximately 22% and severe lesions accounting for only 9%. This imbalance presents challenges, such as the risk of model bias towards the majority class and the potential difficulty in accurately identifying severe cases due to limited samples. It also highlights the need for evaluation metrics that address class imbalance. However, it reflects the real-world distribution of LSD cases, where severe instances are less common, and offers opportunities to develop robust models capable of handling imbalanced data.

As shown in Figure 7, the ViT model achieved a training loss of 0.11 and an accuracy of 96.59%. For the validation set, the ViT model recorded a loss of 2.33 and an accuracy of 67.64%. The total CPU wall time for executing the ViT model was 4 min and 32 s. On the other hand, the ConvMixer model reported a training loss of 0.19 with an accuracy of 93.04%. The validation loss and accuracy for the ConvMixer model were 2.09% and 71.95%, respectively. The total CPU wall time for running the ConvMixer model was 13 min and 17 s.

The average performance metrics for the ViT model are as follows: precision at 67.24%, recall at 67.65%, accuracy at 68% and *F*1-score at 66.94%. For the ConvMixer model, the average metrics are precision of 78.76%, recall of 79.90%, accuracy of 80% and *F*1-score of 79.02%. The ensemble model outperforms both, with average precision of 80.12%, recall of 80.88%, accuracy of 81% and *F*1-score of 80.37%. The total CPU wall time for the models is 18.3 ms.

TABLE 2 | Definitions of performance metrics to evaluate classification models.

Performance parameter	Definition
Confusion matrix	It is a comprehensive table used to evaluate the performance of a DL classification model, consisting of four key parameters: true positive (TP), false positive (FP), false negative (FN) and true negative (TN)
TP	This metric indicates the number of data samples that the model correctly identified as positive. It reflects the model's ability to correctly identify positive instances, which is crucial in applications such as disease detection, where identification of positive instances is critical
TN	This metric measures how many data samples are correctly classified as negative by the model. It is essential to ensure that non-positive cases are not misreported, which is important to maintain specificity in model predictions
FP	This metric represents the number of samples misclassified as positive by the model. High FP rates can lead to unnecessary interventions or actions, especially in critical applications such as medical diagnostics or security systems
FN	This metric indicates the number of true positive samples that were incorrectly predicted as negative. Minimization of FN is critical in scenarios where failure to predict a positive instance could have serious consequences, such as failure to diagnose a disease
Accuracy	This metric is calculated as $\frac{TP+TN}{TP+TN+FP+FN}$ It measures the overall accuracy of the model's predictions. Although accuracy is a useful metric, it can be misleading in cases of unbalanced datasets in which the majority class dominates the predictions
Precision (p)	This metric is calculated as $\frac{TP}{TP+FP}$ It indicates the proportion of positive identifications that were actually correct. It is particularly important in applications where the cost of FPs is high, such as fraud detection
Recall (r)	This metric is calculated as $\frac{TP}{TP+FN}$ It measures the ability of the model to identify all actual positive cases. High recall is crucial in fields like healthcare, where failing to identify positive cases could lead to adverse outcomes
$F1$-score	This metric is the harmonic mean of precision and recall, calculated as $2\times\frac{p\times r}{p+r}$ It provides a single metric that balances both precision and recall, offering a more comprehensive measure of a model's performance, especially when dealing with imbalanced datasets
AUC	AUC refers to the area under the ROC curve, which contrasts the TPR with FPR. It ranges from 0 to 1, with higher values indicating better diagnostic performance and the ability of the model to distinguish between positive and negative classes
Precision–recall curve	This curve plots precision versus recall for different threshold values. It is particularly useful when evaluating models on datasets with significant class imbalance, as it provides insight into the tradeoffs between precision and recall

Note: It included accuracy, recall, precision, $F1$-score, AUC and confusion matrix elements (TP, TN, FP and FN). These metrics are essential to assess classification quality, particularly under class imbalance.

Abbreviations: FPR, false positive rate; TPR, true positive rate.

Figure 8 shows the confusion matrix for the ViT and ConvMixer models. For the ViT model, the confusion matrix reveals 107 true positives (TPs), 45 true negatives (TNs), 33 false negatives (FNs) and 19 false positives (FPs) for normal skin lesions; 31 TPs, 111 TNs, 15 FNs and 47 FPs for mild lumpy skin lesions and 0 TPs, 18 FNs, 0 FPs and 186 TNs for severe lumpy skin lesions. For the ConvMixer model, the confusion matrix results are as follows: (i) For normal skin lesions, there are 130 TPs, 44 TNs, 20 FPs and 11 FNs; (ii) for mild lumpy skin lesions, the model identified 22 TPs, 145 TNs, 13 FPs and 24 FNs; and (iii) for severe lumpy skin

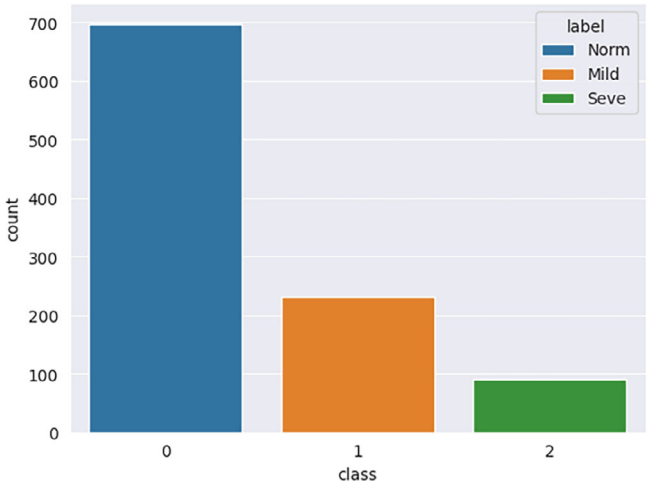

FIGURE 6 | Number of animals affected with lumpy skin lesions for the imbalanced dataset.

lesions, the model correctly predicted 11 TPs, 178 TNs, 8 FPs and 7 FNs. The ensemble model's confusion matrix reveals: (i) for normal skin lesions, it correctly identified 127 TPs, 44 TNs, 20 FPs and 13 FNs; (ii) for mild lumpy skin lesions, the ensemble model reported 25 TPs, 144 TNs, 14 FPs and 21 FNs; and (iii) for severe lumpy skin lesions, it achieved 13 TPs, 181 TNs, 5 FPs and 5 FNs.

Using the ViT model, the AUC values achieved are 0.73 for normal skin lesions, 0.69 for mild lumpy skin lesions and 0.50 for severe lumpy skin lesions. In comparison, the ConvMixer model demonstrates AUC values of 0.81 for normal, 0.70 for mild and 0.78 for severe lumpy skin lesions. The ensemble model shows the highest performance with AUC values of 0.80 for normal, 0.73 for mild and 0.85 for severe lumpy skin lesions. These AUC metrics, illustrated in Figure 9, reflect the models' varying capabilities in distinguishing between different severity levels of LSD, highlighting the ensemble model's superior overall diagnostic performance.

The precision–recall curve values for the ViT model are 0.81 for normal skin lesions, 0.34 for mild lumpy skin lesions and 0.09 for severe lumpy skin lesions. In contrast, the ConvMixer model demonstrates higher precision–recall values with 0.85 for normal, 0.85 for mild and 0.42 for severe lumpy skin lesions. The ensemble model achieves precision–recall values of 0.85 for normal, 0.45 for mild and 0.55 for severe lumpy skin lesions. These results, depicted in Figure 10, highlight the varying effectiveness of each model in balancing precision and recall across different types of lumpy skin lesions, with the ensemble model showing improved

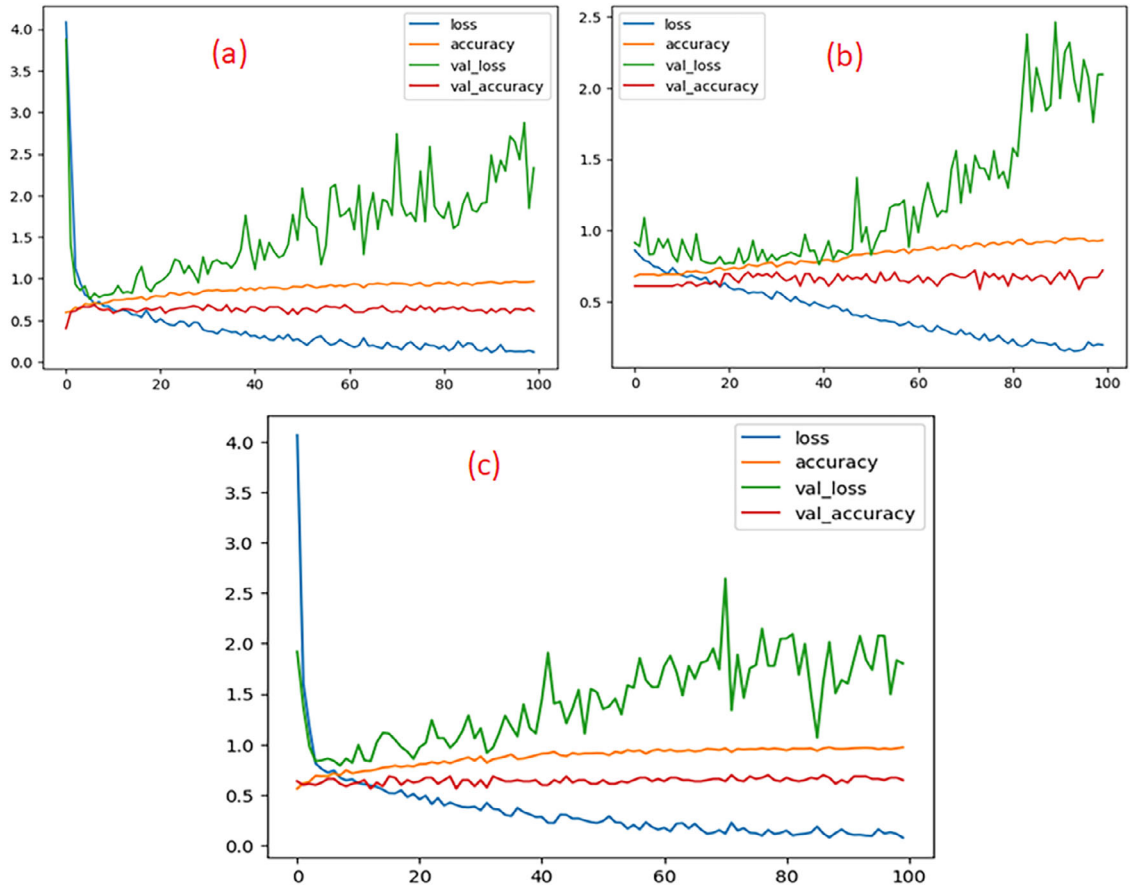

FIGURE 7 | Training and validation accuracy and loss curves for the three deep learning models on the imbalanced dataset: (a) ViT exhibits high training accuracy but significantly lower validation performance, indicating overfitting; (b) ConvMixer shows improved generalization, with validation curves closer to training ones; and (c) the ensemble model achieves the most balanced and stable performance, with reduced loss and better validation accuracy. *X*-axis: epochs; *Y*-axis: accuracy (top) and loss (bottom).

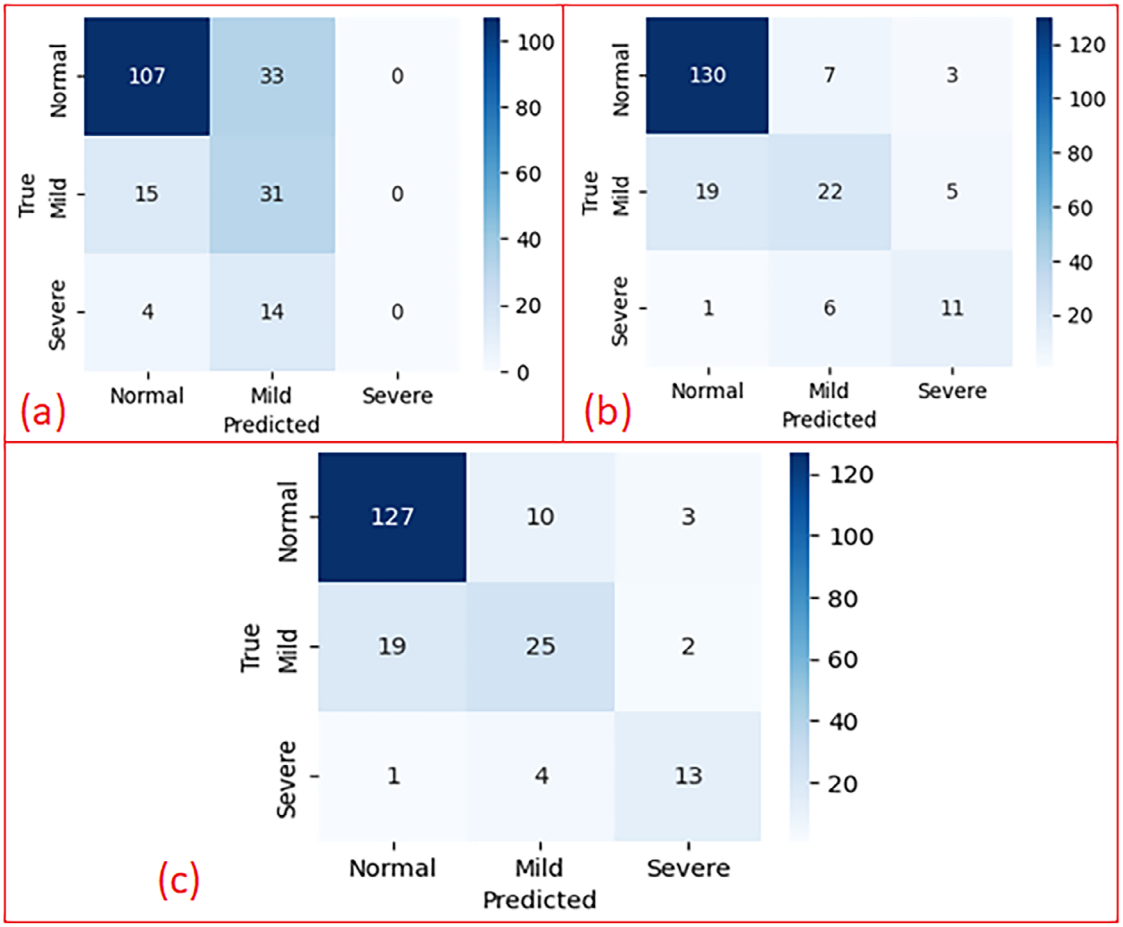

FIGURE 8 | Confusion matrices for ViT (a), ConvMixer (b) and ensemble model (c) on the imbalanced dataset. Each matrix shows the number of correct and incorrect predictions per class: normal, mild and severe lesions. Diagonal values represent correct classifications (true positives), whereas off-diagonal elements indicate misclassifications. The ensemble model achieves better balance across all categories, especially improving severe class detection.

performance in identifying severe cases compared to ViT and ConvMixer.

3.2 | Performance Evaluation of DL Models for a Balanced Dataset

The dataset underwent extensive preprocessing to improve the quality and variability of the input data. This included a number of data augmentation techniques designed to improve the robustness and generalization of the model. Preprocessing steps involved normalization to standardize pixel values, image scaling to ensure consistent input sizes, random horizontal flipping to increase variability, random rotations with a factor of 0.02 to account for changes in orientation and random zooming with height and width factors of 0.2 to simulate different scales. Originally, the dataset contained 697 images of normal skin lesions (labelled 0), 231 images of mild lumpy skin lesions (labelled 1) and 91 images of severe lumpy skin lesions (labelled 2). After applying the up-sampling method, the dataset was adjusted to include an equal number of images for each class (see Figure 11), resulting in 697 images for each category: normal (0), mild (1) and severe (2). This balancing ensures a fair representation of each class in the training process, facilitating a more accurate and unbiased model performance evaluation.

As illustrated in Figure 12, the ViT model achieved a training accuracy of 97.67% and a training loss of 0.06. The testing accuracy and loss for the ViT model were 95.94% and 0.14, respectively. The CPU runtime for executing the ViT model was 6 min and 7 s, with a total CPU wall time of 6 min and 12 s. For the ConvMixer model, the training accuracy reached 97.67%, and the training loss was 0.06. The validation performance showed a validation accuracy of 93.45% and a validation loss of 0.5920. The ensemble model demonstrated the highest performance, with a training accuracy of 98.07% and a training loss of 0.0531. Its validation results included a validation accuracy of 96.43% and a validation loss of 0.2574. The total CPU wall time for executing the ensemble model was 32 min and 47 s. These metrics underscore the superior performance of the ensemble model, particularly in achieving the highest accuracy and lowest loss among the tested models.

The ViT model yields the following confusion matrix results: for normal skin lesions, we obtained 131 TPs, 8 FNs, 5 FPs and 275 TNs. For mild lumpy skin lesions, the results are 131 TP, 9 FN, 5 FP and 274 TN. For severe lumpy skin lesions, the results are 140

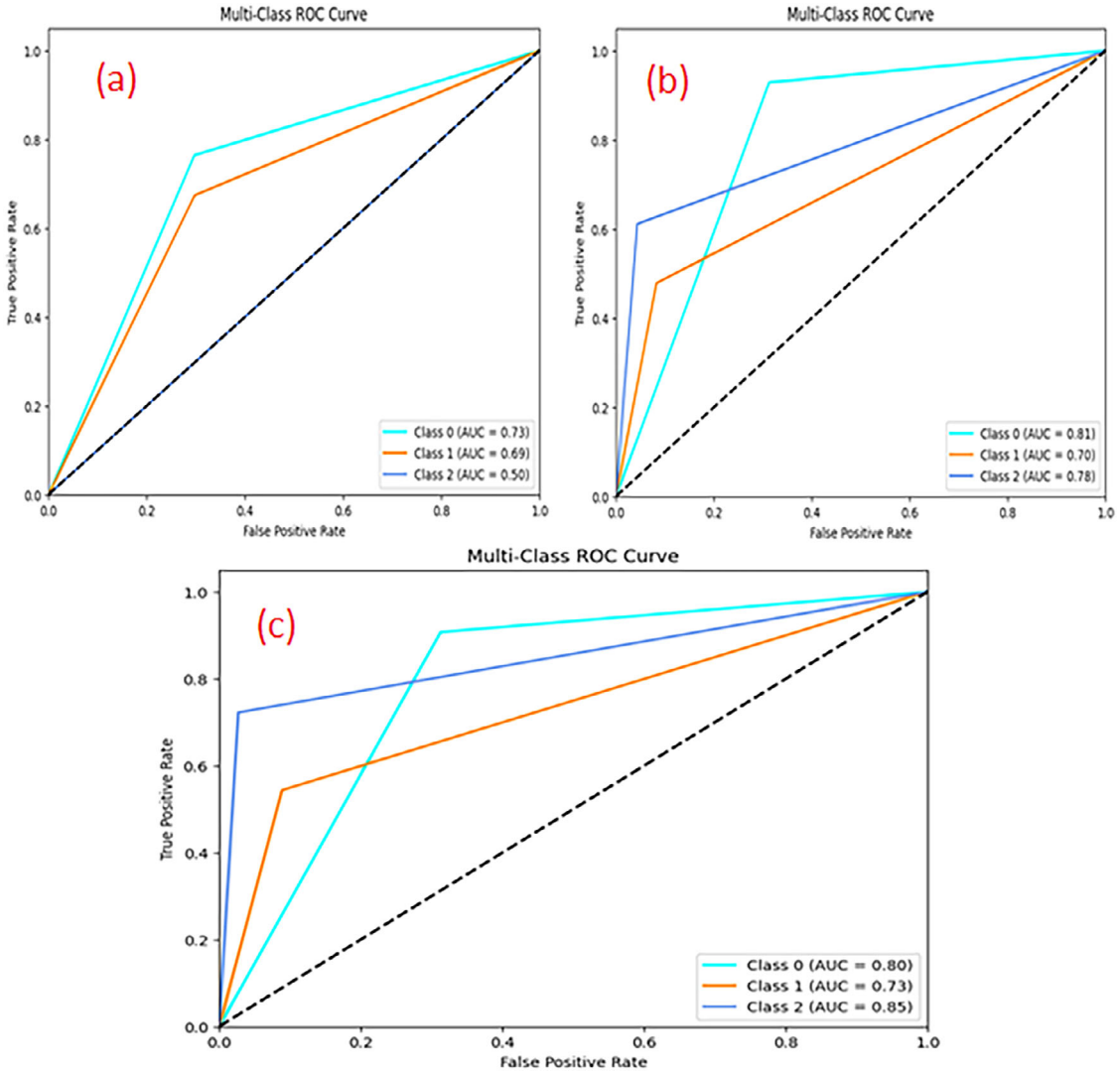

FIGURE 9 | AUC for each model and lesion class: (a) ViT shows limited discriminative ability for severe lesions; (b) ConvMixer demonstrates improved AUC values across all classes; (c) the ensemble model outperforms individual models, particularly in detecting severe cases (AUC = 0.85).

TP, 0 FN, 7 FP and 272 TN (see Figure 13a). Using the ConvMixer model, as shown in Figure 13b, we obtained 129 TP, 10 FN, 5 FP and 275 TN for normal skin lesions; 132 TP, 8 FN, 10 FP and 275 TN for mild lumpy skin lesions and 140 TP, 0 FN, 3 FP and 276 TN for severe lumpy skin lesions. Figure 13c illustrates the results for the ensemble model: 134 TP, 5 FN, 3 FP and 277 TN for normal skin lesions; 134 TP, 6 FN, 4 FP and 277 TN for mild lumpy skin lesions and 140 TP, 0 FN, 4 FP and 274 TN for severe lumpy skin lesions.

The diagnosis rates achieved using the ViT model are notably high (Figure 14a): 0.96 for normal skin lesions, 0.96 for mild lumpy skin lesions and 0.99 for severe lumpy skin lesions. For the ConvMixer model, shown in Figure 14b, the AUC values are equally impressive, with 0.96 for normal skin lesions, 0.95 for mild lumpy skin lesions and 0.99 for severe lumpy skin lesions. The ensemble model, depicted in Figure 14c, demonstrates superior performance with AUC values of 0.98 for normal skin lesions, 0.97 for mild lumpy skin lesions and 0.99 for severe lumpy skin lesions. These results highlight the effectiveness of the ensemble model in achieving higher diagnostic accuracy across all classes.

Figure 15 illustrates the precision–recall curves for three different DL models applied to skin lesion classification. For the ViT model (Figure 15a), the precision–recall values are 0.93 for normal lesions, 0.92 for mild lesions and 0.95 for severe lesions. The ConvMixer model (Figure 15b) shows precision–recall values of 0.92 for normal lesions, 0.90 for mild lesions and 0.98 for severe lesions. The Ensemble model (Figure 15c) achieves precision–recall values of 0.95, 0.94 and 0.97 for normal, mild and severe lesions, respectively. These curves reveal that the Ensemble model exhibits the most consistent performance across all classes, demonstrating high precision and recall. Although the ConvMixer model is particularly effective at detecting severe lesions, it shows slightly lower performance for mild lesions. Overall, these models exhibit robust classification capabilities, with the Ensemble model offering the most balanced and effective results across all categories. Details regarding the parameters used for constructing the ConvMixer model are provided in Table 3.

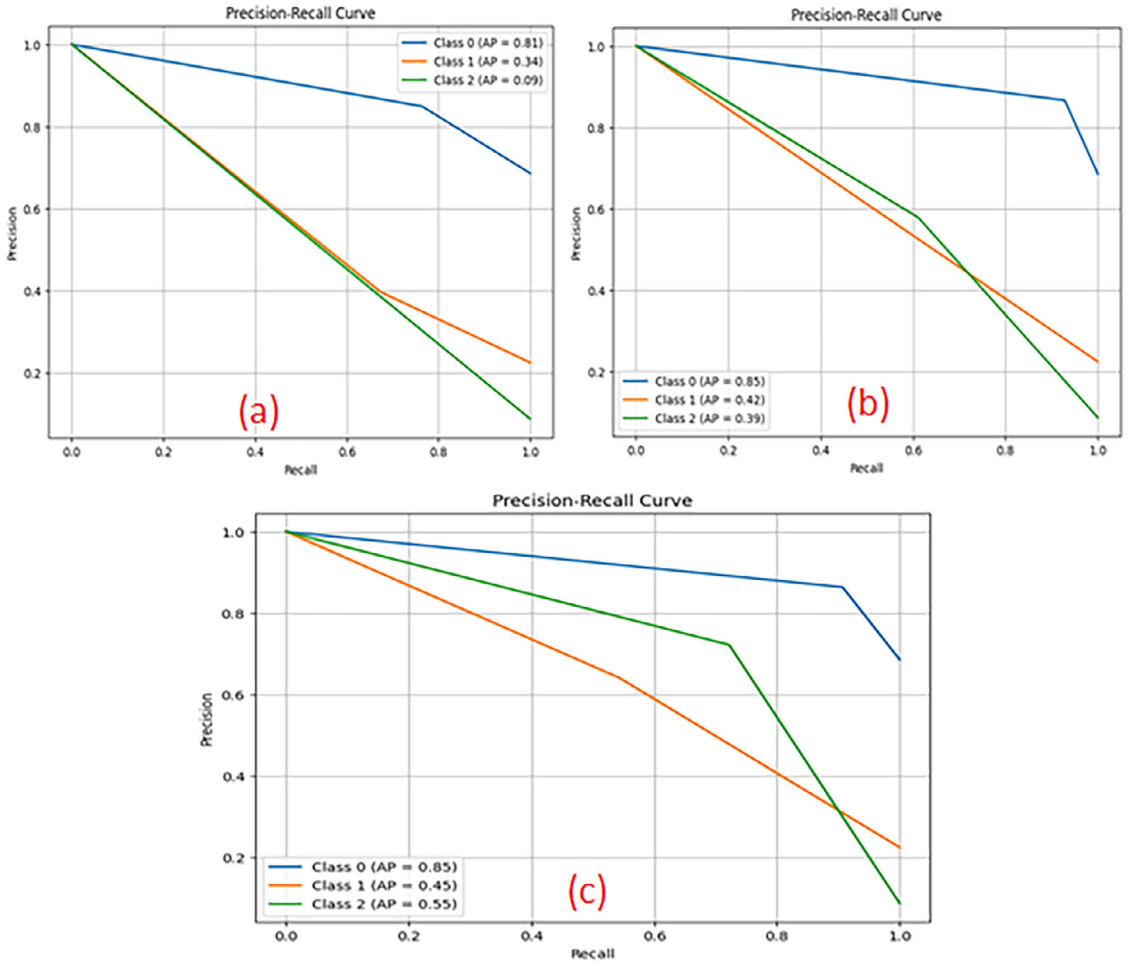

FIGURE 10 | Precision–recall curves for the three models on the imbalanced dataset. These curves assess model performance under class imbalance: (a) ViT shows poor precision–recall tradeoff for severe lesions; (b) ConvMixer improves on mild and severe categories; (c) the ensemble model balances precision and recall across all three classes, especially for under-represented lesions.

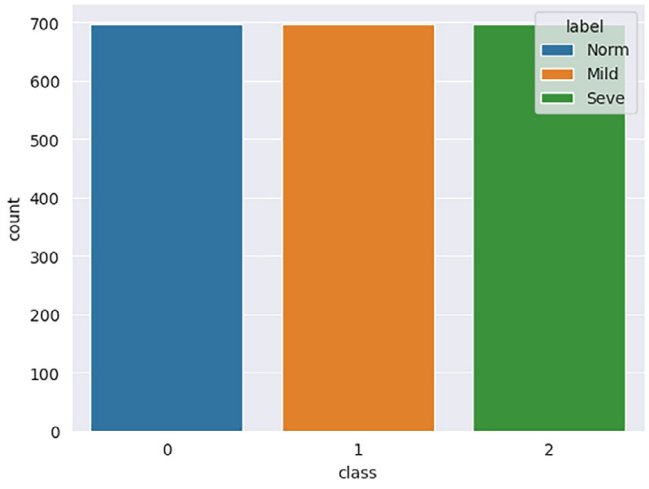

FIGURE 11 | The number of animals affected with lumpy skin lesions using the balance dataset.

3.3 | Comprehensive Comparison of DL Models

Table 4 provides a comprehensive comparison of DL models for detecting and classifying LSD on both balanced and imbalanced datasets. The ensemble model consistently outperforms the ViT and ConvMixer models across all performance metrics, including accuracy, precision, recall, *F*1-score and AUC, on both types of datasets. Specifically, the ensemble model achieves the highest accuracy, precision, recall and *F*1-score values, with notable AUC scores across normal, mild and severe skin lesions. This indicates its superior ability to handle the complexity of the LSD classification task. The balanced dataset shows significantly better performance across all metrics compared to the imbalanced dataset. For instance, the ensemble model on the balanced dataset reaches an accuracy of 97.37%, precision of 97.38%, recall of 97.37% and an *F*1-score of 97.37%. In contrast, the performance metrics for the imbalanced dataset are notably lower, highlighting the challenges posed by class imbalance. We have added 95% confidence intervals for the main performance metrics derived via bootstrapping to provide a more complete and robust assessment of model performance.

Table 5 provides a detailed comparison of the DL models according to their execution time. For the balanced dataset, the ensemble model achieves the fastest execution time with 18.1 ms, followed by ConvMixer with 18.6 ms. The ViT model, on the other hand, takes much longer, with an execution time of 6 min and 12 s. For the unbalanced dataset, the ensemble

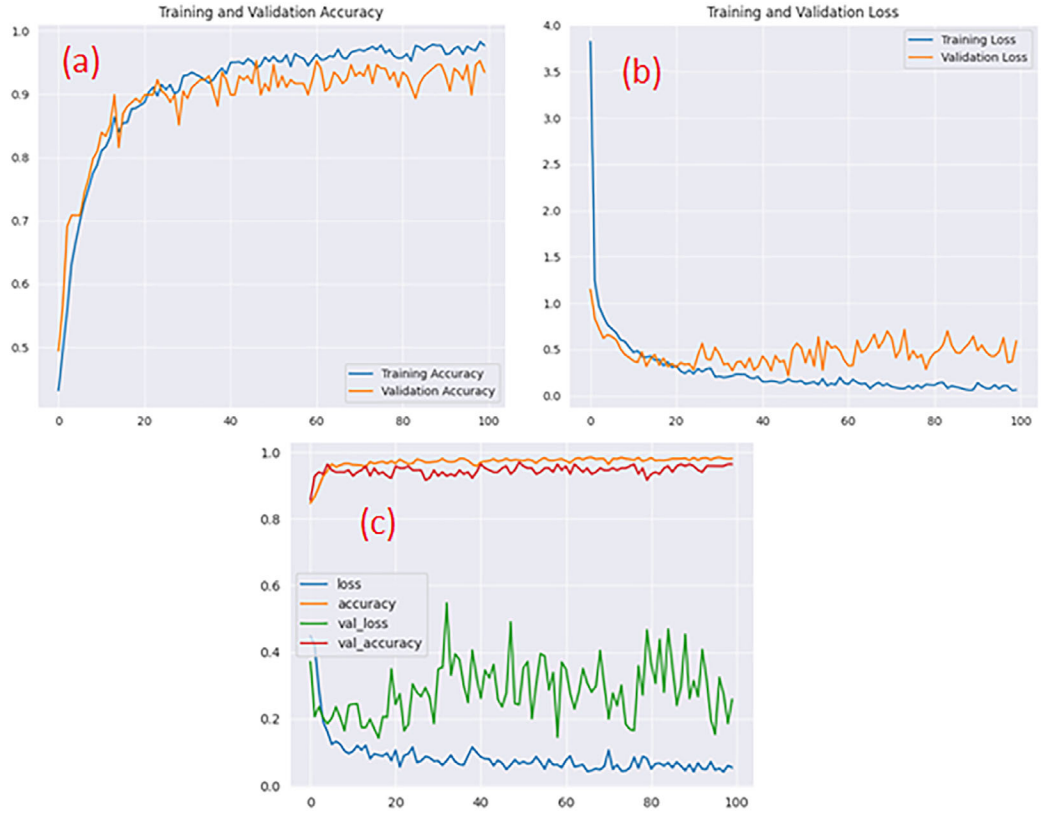

FIGURE 12 | Accuracy and loss curves for the three models: (a) ViT achieves 97.67% training accuracy and 95.94% test accuracy with low loss; (b) ConvMixer shows similar training accuracy but lower validation performance (93.45%) and higher loss; (c) the ensemble model outperforms both, with 98.07% training and 96.43% validation accuracy, demonstrating superior generalization and stability.

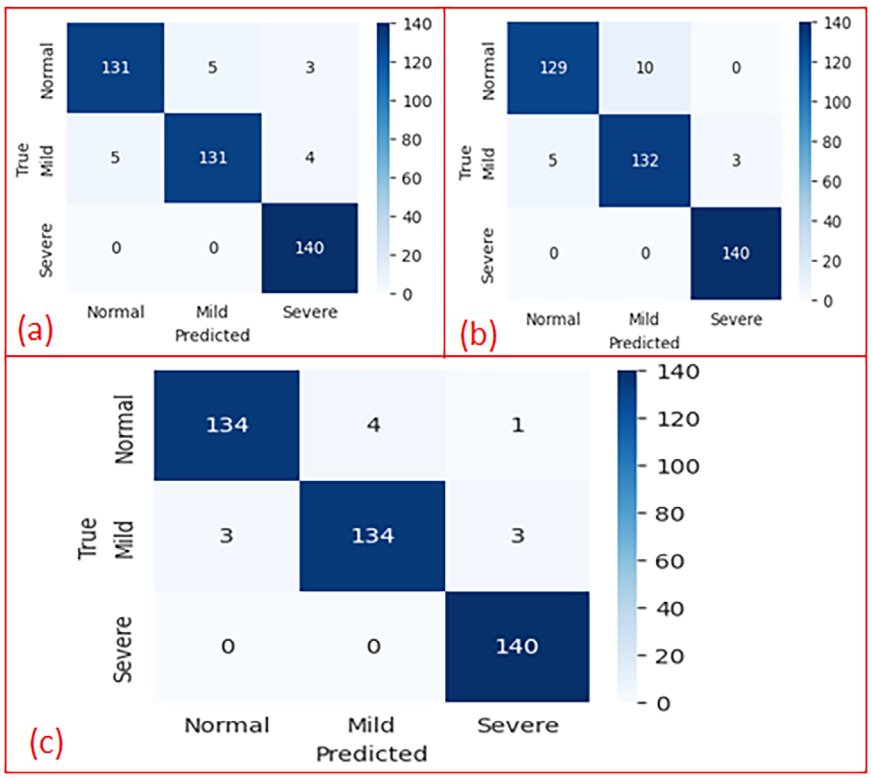

FIGURE 13 | Confusion matrices for the three models. (a) ViT performs well on severe lesions but shows more false negatives for normal and mild cases. (b) ConvMixer improves precision on severe lesions with fewer false positives. (c) The ensemble model achieves the best overall balance, reducing errors across all categories, especially for normal and mild lesions.

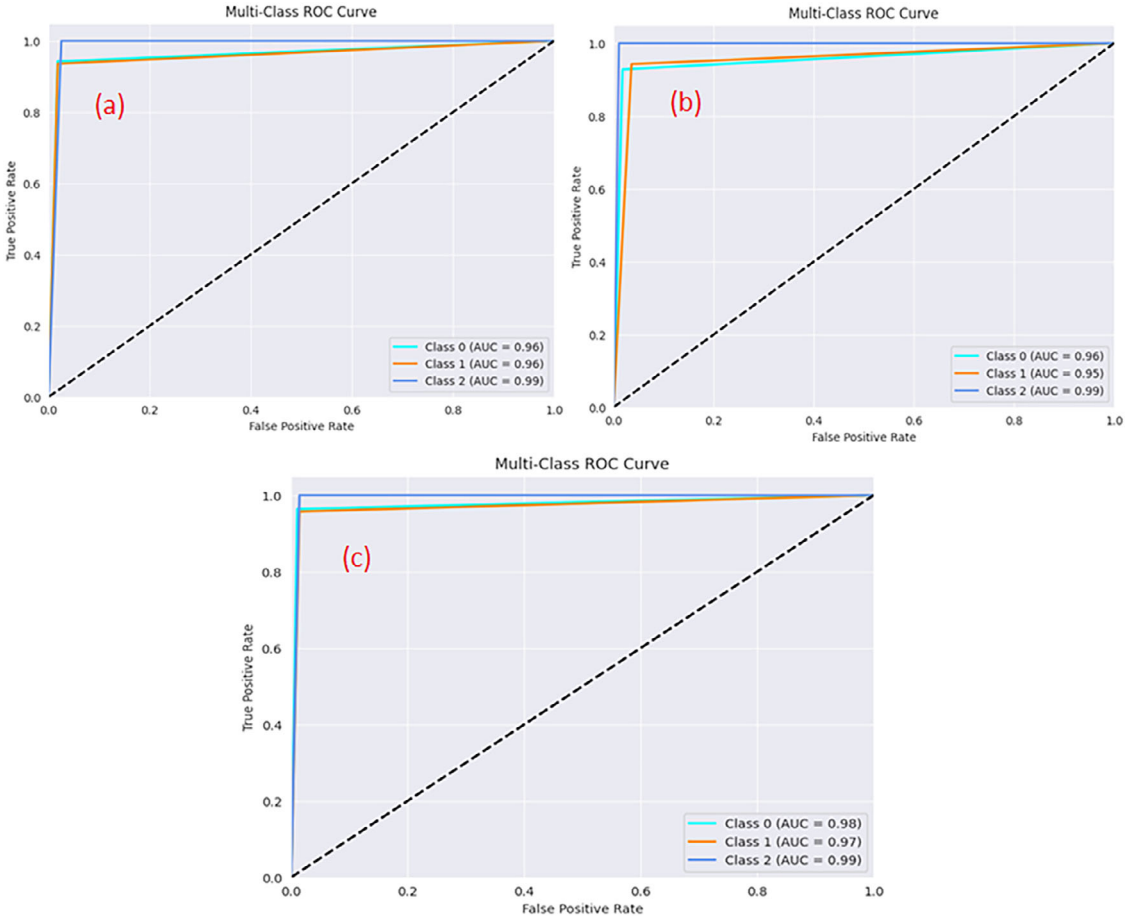

FIGURE 14 | AUC values for the three models: (a) ViT and (b) ConvMixer show strong diagnostic performance, especially on severe lesions; (c) the ensemble model achieves the highest AUC across all classes, indicating superior overall classification accuracy.

model again shows superior efficiency, with an execution time of 18.3 ms. ConvMixer's execution time is 29.6 ms, whereas ViT remains the slowest at 4 min and 32 s. This comparison emphasizes the efficiency of the ensemble model on different types of data, making it the most time-efficient option for skin lesion classification.

3.4 | Comparison With Previous Studies

Few studies have been reported on DL applications in veterinary diagnostics, particularly in the classification of dermatological conditions in cattle. However, these works have been limited to binary classification tasks—distinguishing merely between healthy and affected animals—and have predominantly relied on conventional CNN architectures with limited lesion severity resolution and constrained generalizability.

For instance, Senthilkumar et al. (2024), in Veterinary Sciences, evaluated pre-trained CNN models (EfficientNetB7, MobileNetV2, ResNet-50, VGG16 among others) for the early detection of LSD, achieving a maximum accuracy of approximately 96%, but still within a binary classification framework. Similarly, Saqib et al. (2024) used a MobileNetV2 model optimized with RMSprop, reporting an accuracy of 95%, again focused solely on distinguishing healthy from LSD-affected cattle. Recently, Saha (2024) has also classified LSD in dairy cows by using four CNN models (DenseNet, Xception, MobileNetV2 and InceptionResNetV2) and achieved 96% classification accuracy with 98% AUC score for MobileNetV2.

In contrast, our study introduces several important innovations that go beyond accuracy benchmarks. First, we propose a multi-class classification system that differentiates between normal, mild and severe lesions—providing more granular diagnostic support aligned with clinical progression of LSD, which is notably absent in previous work. Second, we integrate two advanced DL architectures—ViT and ConvMixer—into a novel ensemble model. Although both models are state-of-the-art in the broader AI domain, they remain underutilized in veterinary imaging. Their combined use leverages ViT's strength in capturing global contextual features and ConvMixer's efficiency in extracting local spatial patterns, offering improved robustness across all severity levels.

Our ensemble model achieved strong performance (accuracy: 97.37%, *F*1-score: 97.37% and AUC: 0.99) on a balanced dataset, outperforming conventional CNN-based approaches. In addition to improved performance metrics, our approach introduces several conceptual and methodological innovations that distinguish it from prior work. Beyond numerical accuracy, the novelty lies in our integration of clinically grounded, multi-class annotation

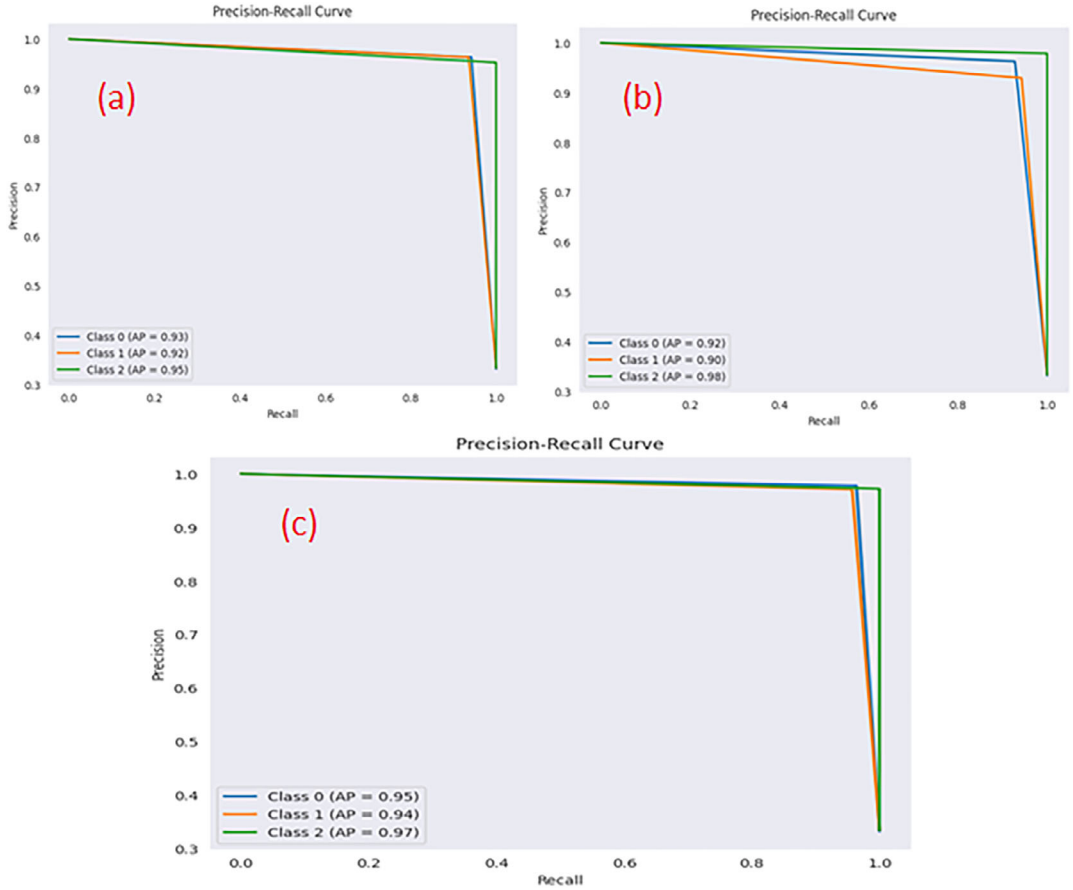

FIGURE 15 | Precision–recall curves for the three models: (a) ViT and (b) ConvMixer show strong performance, especially on severe lesions; (c) the ensemble model achieves the most balanced precision and recall across all classes.

TABLE 3 | The structure of ConvMixer model.

Types of layer	Shape of the output layer	Parameter
Input layer	128 × 128 × 3	0
Conv2D	128 × 128 × 256	1024
Activation	128 × 128 × 256	0
Depthwise conv2D	128 × 128 × 256	2560
Activation	128 × 128 × 256	0
Conv2D	128 × 128 × 256	65,792
Activation	128 × 128 × 256	0
Depthwise conv2D	128 × 128 × 256	2560
Activation	128 × 128 × 256	0
Conv2D	128 × 128 × 256	65,792
Activation	128 × 128 × 256	0
Global average pooling	256	0
Dense	512	131,584
Dense	3	1539
Total trainable parameter	—	270,851

strategies—normal, mild and severe—developed in collaboration with veterinary experts. Briefly, the novelty lies not only in model selection but also in the broader pipeline design: (i) a carefully curated and augmented dataset of over 2000 images, sourced from both online and field photography; (ii) expert-driven annotation based on lesion severity, reviewed through inter-annotator agreement; (iii) targeted data augmentation and up-sampling strategies to address class imbalance and (iv) open access to all code and data, promoting reproducibility—still a rarity in veterinary DL research.

This fine-grained labelling schema offers greater clinical applicability compared to binary classifications. Furthermore, our use of real-world field data, combined with open-access resources, enhances ecological validity and ensures the model is trained on diverse, authentic conditions. The ensemble of ViT and ConvMixer represents another significant contribution, as these models are rarely applied in veterinary imaging and offer complementary strengths in global and local feature extraction. By emphasizing data transparency, ethical sourcing and reproducibility—including public release of code and dataset—we advance not only the technical but also the scientific rigour of AI applications in animal health. These elements collectively contribute to a more deployable and trustworthy AI system for real-world veterinary diagnostics.

TABLE 4 | Performance comparison of vision transformer (ViT), ConvMixer and ensemble models on balanced and imbalanced datasets.

Model	Infection type	Accuracy (%)	Precision (95% CI)	Recall (95% CI)	*F*1-score (95% CI)	AUC
Comparison of performance parameters for DL models on balanced dataset						
ViT	Normal	95.94	0.96 ± 0.03	0.94 ± 0.03	0.95 ± 0.03	0.96
	Mild	95.94	0.96 ± 0.03	0.94 ± 0.03	0.95 ± 0.03	0.96
	Severe	95.94	0.95 ± 0.03	1.00 ± 0.03	0.98 ± 0.03	0.99
ConvMixer	Normal	96.12	0.96 ± 0.03	0.94 ± 0.03	0.95 ± 0.03	0.96
	Mild	96.12	0.96 ± 0.03	0.94 ± 0.03	0.95 ± 0.03	0.97
	Severe	96.12	0.98 ± 0.03	1.00 ± 0.03	0.99 ± 0.03	0.99
Ensemble	Normal	97.37	0.98 ± 0.03	0.96 ± 0.03	0.97 ± 0.03	0.98
	Mild	97.37	0.97 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.97
	Severe	97.37	0.97 ± 0.03	1.00 ± 0.03	0.99 ± 0.03	0.99
Comparison of performance parameters for DL models on imbalanced dataset						
ViT	Normal	67.64	0.85 ± 0.03	0.76 ± 0.03	0.80 ± 0.03	0.73
	Mild	67.64	0.40 ± 0.07	0.67 ± 0.07	0.50 ± 0.07	0.69
	Severe	67.64	0.40 ± 0.20	0.62 ± 0.20	0.40 ± 0.20	0.50
ConvMixer	Normal	79.90	0.87 ± 0.03	0.93 ± 0.03	0.90 ± 0.03	0.81
	Mild	79.90	0.63 ± 0.07	0.48 ± 0.07	0.54 ± 0.07	0.70
	Severe	79.90	0.58 ± 0.20	0.61 ± 0.20	0.59 ± 0.20	0.78
Ensemble	Normal	80.88	0.86 ± 0.03	0.91 ± 0.03	0.89 ± 0.03	0.80
	Mild	80.88	0.64 ± 0.07	0.54 ± 0.07	0.59 ± 0.07	0.73
	Severe	80.88	0.72 ± 0.20	0.72 ± 0.20	0.72 ± 0.20	0.85

Note: The ensemble consistently outperforms individual models across all metrics (accuracy, precision, recall, *F*1-score and AUC). Results highlight the effect of class balancing and model integration in improving diagnostic accuracy, especially for severe cases.

Abbreviation: DL, deep learning.

TABLE 5 | Execution time comparison of models on balanced and imbalanced datasets.

Type of dataset	Type of DL model	Run time
Imbalanced	ViT	4 min 32 s
	ConvMixer	29.6 ms
	Ensemble	18.3 ms
Balanced	ViT	6 min 12 s
	ConvMixer	18.6 ms
	Ensemble	18.1 ms

Note: Run time measured in milliseconds (ms) for ConvMixer and ensemble; in minutes and seconds for ViT. The ensemble model is both the most accurate and the fastest, showing high computational efficiency.
Abbreviations: DL, deep learning; ViT, vision transformer.

3.5 | Limitations and Future Directions

A limitation of our current approach is that the classification task was restricted to LSD-specific lesions and healthy skin, without including visually confounding conditions such as ringworm, bacterial infections or parasitic dermatoses. By contrast, studies like Hwang et al. (2022) have employed DL techniques for broader differential diagnosis across multiple skin diseases in animals using more heterogeneous datasets. Further, the used imbalanced dataset contains less number of 'mild' and 'severe' LSD images, so future work is recommended to increase dataset diversity by collecting more 'mild' and 'severe' cases.

The future work should focus on several key areas to further advance this research. Expanding the dataset to include a more diverse range of images and conditions is key to improving the generalization and robustness of the model. This includes collecting images of different races, ages and stages of disease progression, as well as incorporating environmental variations to more accurately simulate real-world conditions.

In addition, it is essential to refine the architectures of the models to optimize their performance. This could involve experimenting with different hyperparameters, exploring new architectural modifications and incorporating techniques such as transfer learning and self-supervised learning to improve model training and efficiency. Exploring hybrid models that combine the strengths of multiple DL approaches could also yield further improvements.

The development of user-friendly interfaces and the integration of the diagnostic tool with mobile and edge computing devices can facilitate its use by veterinarians and livestock keepers, enabling on-the-spot and immediate disease detection and management. In addition, the integration of other advanced DL models and techniques could further improve the robustness and applicability of the system. Techniques, such as generative adversarial networks (GANs) for data augmentation, attention mechanisms to improve the model's focus on relevant features and ensemble learning to increase confidence in the model, are promising for future research. As DL-based diagnostic tools advance, we can better protect livestock health, ensure food security and support sustainable agricultural practices. The implications of this research extend beyond LSD detection, providing a framework for addressing diverse veterinary and agricultural challenges through the innovative application of DL technologies.

4 | Conclusion

This study presents a significant advance in the application of DL techniques to veterinary diagnostics, with particular reference to the detection and classification of LSD in cattle. Leveraging state-of-the-art models such as ViT and ConvMixer, we developed a robust system capable of accurately identifying LSD lesions from a set of cattle images. Data creation and data augmention were crucial steps to ensure the ability of the models to generalize well across different conditions, thus improving their practical applicability in real-world scenarios. The integration of the ViT and ConvMixer models has proven to be very effective for the detection and classification of LSD in cattle. This study shows the significant advantages of combining multiple DL architectures to improve diagnostic performance, achieving higher accuracy, recall, precision and $F1$ scores than individual models. The successful application of these advanced DL models underscores their potential in revolutionizing veterinary diagnostics by providing early and accurate detection of LSD, which is critical for effective management and control of the disease.

Our results highlight the importance of leveraging state-of-the-art DL technologies to address critical animal health challenges. The development of a robust DL-based diagnostic tool for LSD can lead to significant improvements in early detection of the disease, thereby reducing the spread of LSD and minimizing its economic impact on the livestock sector. In addition, this study sets a precedent for future research on the application of advanced DL techniques to other veterinary and agricultural challenges. In particular, the ability of the ensemble model to leverage the strengths of ViT and ConvMixer demonstrates the importance of model diversity in achieving high diagnostic performance. ViT's ability to capture long-range dependencies in images complements ConvMixer's efficiency in handling spatial hierarchies and local features, resulting in a synergistic effect that improves the robustness and accuracy of the ensemble model. This maiden attempt paves the way to adopt and employ digital platforms and state-of-art IT tools like AI and ML for diagnosing and classifying many other neglected but critical bovine ailments crippling the productivity and agrarian socio-economy.

Author Contributions

Madhumita Pal: investigation, methodology, writing – original draft, formal analysis. **Soudamini Behera**: formal analysis, writing – original draft, validation. **Ranjan K. Mohapatra**: conceptualization, writing – original draft, supervision. **Francesco Branda**: software, formal analysis, writing – original draft, validation. **Gaber A. M. Mersal**: validation, writing – original draft, resources. **Snehasish Mishra**: writing – review and editing, project administration. **Salah M. El-Bahy**: writing – original draft, validation, resources. **Gurudutta Pattnaik**: writing – original draft, validation. **Sovan Pattanaik**: writing – original draft, validation. **Md Sajid Ali**: validation, writing – original draft. **Amiyakanta Mishra**: writing – original draft, validation. **Lawrence Sena Tuglo**: validation, software, writing – review and editing. **Mona Youssef**: writing – original draft, validation.

Acknowledgements

The authors extend their appreciation to Taif University, Saudi Arabia for supporting this work through project number (TU-DSPP-2024-21).

Ethics Statement

The nature of this article does not require any ethical approval.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The datasets generated during and/or analysed during the current study are available from the link: https://zenodo.org/records/11003967.

Peer Review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/vms3.70664.

References

Alkhamis, M. A., and K. VanderWaal. 2016. "Spatial and Temporal Epidemiology of Lumpy Skin Disease in the Middle East, 2012–2015." *Frontiers in Veterinary Science* 3: 19. https://doi.org/10.3389/fvets.2016.00019.

Amin, D. M., G. Shehab, R. Emran, et al. 2021. "Diagnosis of Naturally Occurring Lumpy Skin Disease Virus Infection in Cattle Using Virological, Molecular, and Immunohistopathological Assays." *Veterinary World* 14, no. 8: 2230–2237. https://doi.org/10.14202/vetworld.2021.2230-2237.

Anwar, A., K. Na-Lampang, N. Preyavichyapugdee, and V. Punyapornwithaya. 2022. "Lumpy Skin Disease Outbreaks in Africa, Europe, and Asia (2005–2022): Multiple Change Point Analysis and Time Series Forecast." *Viruses* 14: 2203. https://doi.org/10.3390/v14102203.

Awad, W. S., A. K. Ibrahim, K. Mahran, et al. 2010. "Evaluation of Different Diagnostic Methods for Diagnosis of Lumpy Skin Disease in Cows." *Tropical Animal Health and Production* 42, no. 4: 777–783. https://doi.org/10.1007/s11250-009-9486-5.

Azeem, S., B. Sharma, S. Shabir, et al. 2022. "Lumpy Skin Disease Is Expanding Its Geographic Range: A Challenge for Asian Livestock Management and Food Security." *Veterinary Journal* 279: 105785. https://doi.org/10.1016/j.tvjl.2021.105785.

Bayyappa, M. R. G., M. K. G. Pyatla, S. M. Pabbineedi, et al. 2025. "Exploring the Evolutionary Journey of the Lumpy Skin Disease Virus Through the Phylogenetic and Phylo-Geo Network Analysis." *Frontiers in Cellular and Infection Microbiology* 15: 1575538. https://doi.org/10.3389/fcimb.2025.1575538.

BBC. 2022. "Lumpy Skin Disease: Viral Cattle Disease Sends Rumours Flying in India," BBC. Published October 20. https://www.bbc.com/news/world-asia-india-63262411.

Bhatt, L., R. C. Bhoyar, B. Jolly, et al. 2022. "The Genome Sequence of the Lumpy Skin Disease Virus From the Outbreak in India Suggests a Distinct Lineage of the Virus." Preprint, Biorxiv, September 16. https://doi.org/10.1101/2022.09.15.508131.

Cavalera, S., G. Pezzoni, S. Grazioli, et al. 2022. "Investigation of the "Antigen Hook Effect" in Lateral Flow Sandwich Immunoassay: The Case of Lumpy Skin Disease Virus Detection." *Biosensors (Basel)* 12, no. 9: 739. https://doi.org/10.3390/bios12090739.

Chen, J., Y. Lu, Q. Yu, et al. 2021. "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation." Preprint, Arxiv, February 8. https://doi.org/10.48550/arXiv.2102.04306.

Datten, B., A. A. Chaudhary, S. Sharma, et al. 2023. "An Extensive Examination of the Warning Signs, Symptoms, Diagnosis, Available Therapies, and Prognosis for Lumpy Skin Disease." *Viruses* 15, no. 3: 604. https://doi.org/10.3390/v15030604.

Davies, F. G. 1991. "Lumpy Skin Disease, an African Capripox Virus Disease of Cattle." *British Veterinary Journal* 147, no. 6: 489–503. https://doi.org/10.1016/0007-1935(91)90019-J.

Dosovitskiy, A., L. Beyer, A. Kolesnikov, et al. 2021. "An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale." Preprint, Arxiv, June 3. https://arxiv.org/pdf/2010.11929/1000.

Farag, T. K., H. A. A. Abou-Zeina, S. Abdel-Shafy, et al. 2025. "Progress in Diagnostic Methods and Vaccines for Lumpy Skin Disease Virus: A Path Towards Understanding the Disease." *Veterinary Research Communications* 49, no. 3: 134. https://doi.org/10.1007/s11259-025-10667-2.

Food and Agriculture Organization of the United Nations. 2017. "Lumpy Skin Disease—A Manual for Veterinarians." Food and Agriculture Organization. https://openknowledge.fao.org/items/ec5cdedd-5c8c-4026-ab26-52c42b925d13.

Gongal, G., H. Rahman, K. C. Thakuri, and K. Vijayalakshmy. 2022. "An Overview of Transboundary Animal Diseases of Viral Origin in South Asia: What Needs to Be Done?" *Veterinary Sciences* 9, no. 11: 586. https://doi.org/10.3390/vetsci9110586.

Gupta, T., V. Patial, D. Bali, et al. 2020. "A Review: Lumpy Skin Disease and Its Emergence in India." *Veterinary Research Communications* 44: 111–118.

Han, K., Y. Wang, and H. Chen, et al. 2023. "A Survey on Vision Transformer." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, no. 1: 87–110. https://doi.org/10.1109/TPAMI.2022.3152247.

Hatamizadeh, A., Y. Tang, V. Nath, et al. 2022. "UNETR: Transformers for 3D Medical Image Segmentation." In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE. https://doi.org/10.1109/WACV51458.2022.00181.

Hwang, S., H. K. Shin, J. M. Park, et al. 2022. "Classification of Dog Skin Diseases Using Deep Learning With Images Captured From Multispectral Imaging Device." *Molecular & Cellular Toxicology* 18: 299–309.

Kameswari, C. S., J. Kavitha, T. S. Reddy, et al. 2023. "An Overview of Vision Transformers for Image Processing: A Survey." *International Journal of Advanced Computer Science and Applications* 14, no. 8: 273–289. https://doi.org/10.14569/IJACSA.2023.0140830.

Mehta, S., and M. Rastegari. 2022. "MobileViT: Light-Weight, General-Purpose, and Mobile-Friendly Vision Transformer." Preprint, Arxiv, March 4. https://doi.org/10.48550/arXiv.2110.02178.

Mohapatra, R. K., A. Mahal, P. K. Mohapatra, et al. 2024. "Structure-Based Discovery of *F. religiosa* Phytochemicals as Potential Inhibitors Against Monkeypox (mpox) Viral Protein." *Journal of Biosafety and Biosecurity* 6, no. 3: 157–169.

Mohapatra, R. K., S. Mishra, P. Satapathy, et al. 2024. "Surging Oropouche Virus (OROV) Cases in the Americas: A Public Health Challenge." *New Microbes and New Infections* 59: 101243.

Mohapatra, R. K., S. Verma, V. Kandi, et al. 2023. "The SARS-CoV-2 Omicron Variant and Its Multiple Sub-Lineages: Transmissibility, Vaccine Development, Antiviral Drugs, Monoclonal Antibodies, and Strategies for Infection Control–a Review." *ChemistrySelect* 8, no. 9: e202201380.

Pal, M., A. Mahal, R. K. Mohapatra, et al. 2023. "Deep and Transfer Learning Approaches for Automated Early Detection of Monkeypox (Mpox) Alongside Other Similar Skin Lesions and Their Classification." *ACS Omega* 8, no. 35: 31747–31757.

Pal, M., F. Branda, A. Q. Alkhedaide, et al. 2025. "Early Detection of Human Mpox: A Comparative Study by Using Machine Learning and Deep Learning Models With Ensemble Approach." *Digital Health* 11: 20552076251344135.

Pal, M., R. K. Mohapatra, A. K. Sarangi, et al. 2025. "A Comparative Analysis of the Binary and Multiclass Classified Chest X-Ray Images of Pneumonia and COVID-19 With ML and DL Models." *Open Medicine* 20, no. 1: 20241110.

Pal, M., R. Tiwari, K. Dhama, et al. 2022. "Machine Learning Algorithms and COVID-19: A Step for Predicting Future Pandemics With a Systematic Overview." In *Machine Learning and Deep Learning in Medical Data Analytics and Healthcare Applications*. CRC Press.

Pal, M., S. Parija, G. Panda, et al. 2022. "Risk Prediction of Cardiovascular Disease Using Machine Learning Classifiers." *Open Medicine* 17, no. 1: 1100–1113.

Pal, M., S. Parija, G. Panda, et al. 2023. "COVID-19 Prognosis From Chest X-Ray Images by Using Deep Learning Approaches: A Next Generation Diagnostic Tool." *Journal of Pure & Applied Microbiology* 17, no. 2: 919–930.

Pal, M., S. Parija, R. K. Mohapatra, et al. 2022. "Symptom-Based COVID-19 Prognosis Through AI-Based IoT: A Bioinformatics Approach." *BioMed Research International* 2022, no. 1: 3113119.

Paslarua, A. I., N. O. Verhulst, L. M. Maurer, et al. 2021. "Potential Mechanical Transmission of Lumpy Skin Disease Virus (LSDV) by the Stable Fly (*Stomoxys calcitrans*) Through Regurgitation and Defecation." *Current Research in Insect Science* 1: 100007.

Saha, D. K. 2024. "An Extensive Investigation of Convolutional Neural Network Designs for the Diagnosis of Lumpy Skin Disease in Dairy Cows." *Heliyon* 10, no. 14: e34242.

Sanz-Bernardo, B., I. R. Haga, N. Wijesiriwardana, et al. 2020. "Lumpy Skin Disease Is Characterized by Severe Multifocal Dermatitis With Necrotizing Fibrinoid Vasculitis Following Experimental Infection." *Veterinary Pathology* 57, no. 3: 388–396. https://doi.org/10.1177/0300985820913268.

Saqib, S. M., M. Iqbal, M. T. B. Othman, et al. 2024. "Lumpy Skin Disease Diagnosis in Cattle: A Deep Learning Approach Optimized With RMSProp and MobileNetV2." *PLoS ONE* 19, no. 8: e0302862.

Senthilkumar, C. C. S., G. Vadivu, and S. Neethirajan. 2024. "Early Detection of Lumpy Skin Disease in Cattle Using Deep Learning—A Comparative Analysis of Pretrained Models." *Veterinary Sciences* 11, no. 10: 510.

Sudhakar, S. B., N. Mishra, S. Kalaiyarasu, et al. 2022. "Genetic and Phylogenetic Analysis of Lumpy Skin Disease Viruses (LSDV) Isolated From the First and Subsequent Field Outbreaks in India During 2019 Reveals Close Proximity With Unique Signatures of Historical Kenyan NI-2490/Kenya/KSGP-Like Field Strains." *Transboundary and Emerging Diseases* 69, no. 4: e451–e462. https://doi.org/10.1111/tbed.14322.

The Indian Express. 2022. "Lumpy Skin Disease: Genome Sequence From Current Outbreak Suggests Distinct Lineage of Virus, Say Researchers." Indian Express. Published September 20. https://indianexpress.com/article/cities/pune/lumpy-skin-disease-genome-sequence-outbreak-distinct-lineage-virus-researchers-8160501/.

Trockman, A., and J. Z. Kolter. 2022. "Patches Are All You Need?" Preprint, Arxiv, January 24. https://doi.org/10.48550/arXiv.2201.09792.

Zeedan, G. S. G., A. M. Abdalhamed, A. M. Allam, and S. Abdel-Shafy. 2024. "Molecular Detection of Lumpy Skin Disease Virus in Naturally Infected Cattle and Buffaloes: Unveiling the Role of Tick Vectors in Disease Spread." *Veterinary Research Communications* 48, no. 6: 3921–3939. https://doi.org/10.1007/s11259-024-10541-7.