Research Article

A Comparative Analysis of Deep Learning Models for Automated Snakebite Wound Classification

Arjun Makineni

American High School, 36300 Fremont Blvd, Fremont, California, 94536, United States

ABSTRACT

Snakebite envenomation remains a severe global health challenge, responsible for millions of cases and over 100,000 deaths each year, particularly in rural and underserved regions. This study compares multiple convolutional neural network (CNN) architectures, including EfficientNet variants (B0, B3, B4, V2B0, V2L), ResNet50, MobileNetV3-Large, and ConvNeXt-Large, for the classification of snakebite wounds. After minimal preprocessing through normalization, data augmentation, and class weighting due to a limited dataset size, the models are evaluated through metrics such as accuracy, precision, recall, F1-score, Matthews Correlation Coefficient, Cohen's Kappa, and Average Precision, with bootstrap resampling ensuring statistical rigor. EfficientNetV2B0 achieved the highest accuracy, 91.53%, and perfect recall for venomous bites, minimizing life-threatening false negatives. These findings highlight the potential of AI-based snakebite diagnostics to provide immediate, reliable guidance. The discussion emphasizes the clinical implications of this performance, and this study concludes that integrating the top model into a mobile health application could improve survival rates by enabling effective initial care before patients reach a doctor.

Keywords: Snakebite; Deep Learning; Convolutional Neural Networks; EfficientNet; Mobile Health; Medical Imaging

INTRODUCTION

Snakebite envenomation is a global health concern recognized by the World Health Organization (WHO) as a high-priority neglected tropical disease (1). Annually, over 2.7 million envenomings are reported worldwide, with approximately 100,000 deaths, predominantly in rural and underserved regions of Asia, Africa, and Latin America (2). Beyond fatalities, survivors frequently suffer severe long-term effects such as amputations, tissue necrosis, and permanent disabilities due to delays in accurate diagnosis and treatment (3).

The clinical diagnosis of snakebites often depends on visual inspection of wounds and symptoms, but this process is fraught with challenges, particularly in areas lacking experienced medical personnel. Misdiagnosis or delayed identification of venomous bites can result in improper first aid and treatment delays, exacerbating health outcomes (4). To address these challenges, artificial intelligence (AI) methods, particularly deep learning, have gained significant attention for their potential to deliver automated, reliable, and rapid assessments of medical imagery.

Corresponding author: Arjun Makineni, E-mail: mikimakineni@gmail.com.

Accepted October 3, 2025

https://doi.org/10.70251/HYJR2348.35588596

Deep learning models, especially convolutional neural networks (CNNs), have demonstrated remarkable success in medical imaging domains such as dermatology (5), ophthalmology (5), and radiology (6). These models excel at capturing subtle and complex visual patterns, making them ideal for classifying snakebite wounds, where distinguishing features may be minimal or obscured by swelling and tissue reactions.

In this study, a systematic comparison is performed of several cutting-edge CNN architectures, including EfficientNet variants (B0, B3, B4, V2B0, V2L), ResNet50, MobileNetV3-Large, and ConvNeXt-Large, to identify the best-performing model for snakebite wound classification. The focus is on metrics such as accuracy, recall, and F1-score, with particular emphasis on avoiding false negatives in venomous bite classification due to the severe consequences of missed envenomations.

Our findings build on existing literature in medical AI and provide new insights into the use of image-based classification for snakebite management. We also explore future applications, including the integration of the top-performing model into mobile health platforms, which could empower first responders and communities to take immediate, evidence-based actions.

LITERATURE REVIEW

The use of AI in medical imaging has rapidly expanded, with studies demonstrating that CNNs can achieve expert-level performance. Esteva (5) trained a deep CNN to classify skin lesions and achieved performance comparable to dermatologists. Similarly, Gulshan (7) applied deep learning to diabetic retinopathy detection, while Litjens (8) provided a comprehensive survey of deep learning applications in radiology.

ResNet (9) introduced residual learning to enable the training of very deep networks without degradation issues, setting a benchmark for subsequent architectures. EfficientNet, introduced by (10), proposed a compound scaling method that optimally adjusts network depth, width, and resolution. MobileNetV3 (11) improved computational efficiency by using depthwise separable convolutions, making it ideal for edge and mobile devices. ConvNeXt (12), inspired by Vision Transformers (13), reimagined convolutional architectures to achieve transformer-level performance while retaining the advantages of CNNs.

Snakebite research using AI has been relatively limited. Some studies have explored automated snake species identification from images (14), but few have focused on wound-level classification, which is directly relevant to clinical care. Research by Harrison (15) underscored the need for accessible, low-cost diagnostic tools for snakebites, suggesting that AI could play a transformative role in early intervention.

The integration of AI into mobile health platforms has been demonstrated in other areas, such as malaria detection (16) and tuberculosis screening (17). These studies support the feasibility of deploying lightweight AI models on smartphones for real-time diagnostics, which aligns with this study's goal of developing a mobile-based snakebite classification application.

This work aims to fill the current research gap by comparing multiple modern architectures and providing a comprehensive performance evaluation. Our analysis leverages confusion matrix insights, balanced accuracy measures, and statistical techniques such as bootstrap resampling (18) to ensure robustness and clinical relevance.

METHODS AND MATERIALS

Dataset Acquisition and Biological Justification

The dataset for this research was obtained from Roboflow, a platform that curates high-quality annotated datasets. It contains 415 high-resolution images of snakebite wounds, labeled as either venomous or non-venomous. These images were collected from various publicly available medical and wildlife resources, ensuring a diverse representation of bite patterns and wound presentations. Non-venomous bites are more common globally due to the greater population of non-venomous snake species, leading to a natural imbalance between classes. This imbalance was addressed by applying class weighting during training and by generating augmented venomous bite samples, ensuring that misclassifications in the venomous class were minimized. Prior research emphasizes the critical importance of addressing class imbalance in medical imaging datasets (19, 20) as it directly impacts sensitivity to life-threatening cases. Images were sourced from publicly available datasets spanning regions such as India, the United States, Brazil, and Southeast Asia. However, due to reliance on open-source and clinical image repositories, representation of darker skin tones and certain geographic regions remains limited.

Preprocessing Techniques

All images were resized to standardized input dimensions, with most models using 224×224 pixels and larger variants, such as EfficientNetV2L, using 299×299 pixels. Pixel intensity values were normalized to the [0,1] range to reduce training instability and ensure consistent gradient updates, as recommended in medical image preprocessing pipelines (21). A comprehensive data augmentation pipeline was employed to enhance generalization across all architectures. This included horizontal flips, random rotations (up to 30°), mild brightness and contrast adjustments, Gaussian noise injection, and color jitter. Such augmentations have been shown to simulate real-world variations and improve CNN robustness in medical imaging tasks (22).

Advanced augmentation strategies, such as CutMix (23) and MixUp (24), were tested but not adopted in the final evaluation due to their limited benefits for the small dataset. Denoising filters were applied to suppress background noise, while automated cropping highlighted the wound regions, aligning with established dermatological preprocessing practices (25). These techniques were applied consistently across all model architectures to ensure fairness in evaluation.

The decision to focus on mild yet targeted augmentations rather than heavy distortions proved crucial in achieving the accuracy levels reported in this study. Overly aggressive transformations can degrade the subtle visual features critical for medical image interpretation, a finding supported by studies on data augmentation in low-data medical tasks (26). By carefully balancing augmentation diversity and image fidelity, the models generalized effectively without overfitting, especially for the minority venomous class.

Model Architectures

Eight state-of-the-art CNN architectures were evaluated: EfficientNetB0, EfficientNetB3, EfficientNetB4, EfficientNetV2B0, EfficientNetV2L, ResNet50, MobileNetV3-Large, and ConvNeXt-Large. These architectures were selected due to their strong performance in visual recognition benchmarks and their established success in medical imaging applications (6).

The EfficientNet family (10) employs a compound scaling strategy to balance depth, width, and resolution for superior performance. ResNet50 (9) remains a widely used baseline due to its residual connections, which enable stable training of deep networks. MobileNetV3-Large (11) is optimized for low-latency inference, making it ideal for mobile deployment. ConvNeXt-Large (11) integrates transformer-inspired architectural elements with traditional CNN backbones, achieving high performance on medical and non-medical datasets alike.

All architectures were fine-tuned under consistent preprocessing, optimization, and hyperparameter schedules to ensure a fair comparison of their performance on snakebite wound classification.

Training Procedures

All models were initialized with ImageNet-pretrained weights, a widely adopted approach that leverages transfer learning to improve performance on small datasets (27). The AdamW optimizer (28) was used for all architectures, with an initial learning rate of 1e-4 that was progressively reduced using cosine annealing (29). Batch sizes were set between 16 and 32, depending on model complexity and available GPU memory.

Training was conducted for up to 25 epochs, with early stopping (patience of 5 epochs) to prevent overfitting, a common risk when fine-tuning on small datasets (30). Dropout layers (0.2–0.3) were applied in the classification head to reduce co-adaptation of neurons, and L2 weight decay regularization (1e-6) was consistently applied to improve generalization. All models were trained and evaluated under identical conditions to ensure that performance differences arose from architectural merits rather than differences in optimization.

Evaluation Metrics and Statistical Analysis

A broad set of metrics was employed: accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), Cohen's Kappa, balanced accuracy, and Average Precision (AP). These metrics provide a holistic evaluation of the models, with a particular focus on the minority venomous (V) class to minimize false negatives.

Accuracy measures the overall proportion of correctly classified images (both venomous and non-venomous). Precision indicates how many of the images predicted to belong to a given class are correctly classified (i.e., the fraction of true positives among all positive predictions), while Recall reflects the ability of the model to identify all actual positives (true positive rate). F1-score is the harmonic mean of precision and recall, balancing both metrics in cases of class imbalance.

Matthews Correlation Coefficient (MCC) quantifies the quality of binary classifications, taking into account true and false positives and negatives, and is considered a balanced measure even for imbalanced datasets. Cohen's Kappa measures inter-rater agreement between the model predictions and ground truth, adjusted for agreement by chance. Balanced Accuracy computes the average of recall obtained on each class, ensuring equal importance of both venomous (V) and non-venomous (NV) classes. Average Precision (AP) summarizes the precision-recall curve as a single metric, highlighting performance across different thresholds.

The columns "V" and "NV" refer to the performance metrics calculated for the venomous and non-venomous classes, respectively. For example, "Precision (V)" is the proportion of correctly identified venomous bites among all images predicted as venomous, whereas "Recall (NV)" is the proportion of correctly identified non-venomous bites among all actual non-venomous bites.

Confusion matrices were computed to visualize the number of true positives, true negatives, false positives, and false negatives for each class, giving a detailed picture of where models make errors. To establish statistical robustness, this research employed bootstrap resampling with 1,000 iterations, as recommended by Efron and Tibshirani (12), to compute 95% confidence intervals for each metric. This approach allowed us to provide a reliable measure of the variance and stability of performance estimates.

RESULTS

Performance Metrics Overview

The performance evaluation revealed notable differences among the eight tested architectures, with EfficientNetV2B0 clearly outperforming the others. Its accuracy of 91.53% indicates that it correctly identified more than 9 out of 10 images, which is crucial in the context of snakebite diagnostics. Additionally, the Matthews Correlation Coefficient (MCC) of 0.8290 and Cohen's Kappa of 0.8146 indicate robust agreement with ground truth labels, even when accounting for the inherent class imbalance between venomous (V) and non-venomous (NV) bites.

The precision and recall values highlight EfficientNetV2B0's balanced performance across both classes. A precision of 1.0000 for NV bites suggests that every image predicted as non-venomous was correctly classified, with no false positives. Similarly, a recall of 1.0000 for V bites ensures that all venomous bites were correctly identified, mitigating the risk of life-threatening false negatives. By contrast, models such as ResNet50, with an accuracy of 83.05% and MCC of 0.6397, struggled to balance sensitivity and specificity, particularly in borderline cases where wound characteristics were less distinct.

Table 1 provides a detailed breakdown of accuracy, precision, recall, F1-score, MCC, Cohen's Kappa, and AP for all models, allowing for a side-by-side comparison of strengths and weaknesses. EfficientNetB0 and MobileNetV3-Large also performed well, with accuracies near 90%, but fell slightly short in recall for NV bites. EfficientNetV2L exhibited the weakest performance with an accuracy of 77.97% and a lower F1-score across both classes, likely due to overfitting caused by its larger parameter count relative to the dataset size.

Confusion Matrix Analysis

To gain deeper insights into class-level performance, confusion matrices were constructed for each model. EfficientNetV2B0's confusion matrix showed 36 true

Table 1. Model Performance Metrics

Model	Accuracy	Precision (NV)	Recall (NV)	F1 (NV)	Precision (V)	Recall (V)	F1 (V)	MCC	Kappa	AP
EfficientNetB0	0.8983	1.0000	0.7391	0.8500	0.8571	1.0000	0.9231	0.7960	0.7757	0.9279
EfficientNetB3	0.8814	1.0000	0.6957	0.8200	0.8372	1.0000	0.9111	0.7632	0.7361	0.9516
EfficientNetB4	0.8814	1.0000	0.6957	0.8200	0.8372	1.0000	0.9111	0.7632	0.7361	0.9192
ResNet50	0.8305	0.8095	0.7391	0.7727	0.8421	0.8889	0.8649	0.6397	0.6380	0.8783
MobileNetV3-Large	0.8983	0.9474	0.7826	0.8571	0.8571	0.9722	0.9111	0.7879	0.7793	0.9584
EfficientNetV2L	0.7797	0.7500	0.6522	0.6977	0.7949	0.8611	0.8267	0.5288	0.5257	0.8811
ConvNeXt-Large	0.8983	0.9474	0.7826	0.8571	0.8571	0.9722	0.9111	0.7879	0.7793	0.9925
EfficientNetV2B0	0.9153	1.0000	0.7826	0.8780	0.8780	1.0000	0.9355	0.8290	0.8146	0.9624

positives (correctly identified venomous bites) out of 36 and 18 true negatives out of 23 non-venomous bites, resulting in only a few false negatives and false positives (Figure 1). This distribution is desirable in medical settings, as misclassifying venomous bites as non-venomous could lead to inadequate or delayed treatment.

In contrast, ResNet50's confusion matrix shows a higher number of false negatives for venomous bites, which is problematic in emergency scenarios (Figure 2). MobileNetV3-Large performed competitively, with slightly fewer false positives, reflecting its efficiency in distinguishing ambiguous cases (Figure 3). Visualizing confusion matrices allows clinicians to identify recurring error patterns—such as misclassifications caused by wounds with overlapping bite marks or signs of infection—and informs future dataset refinement.

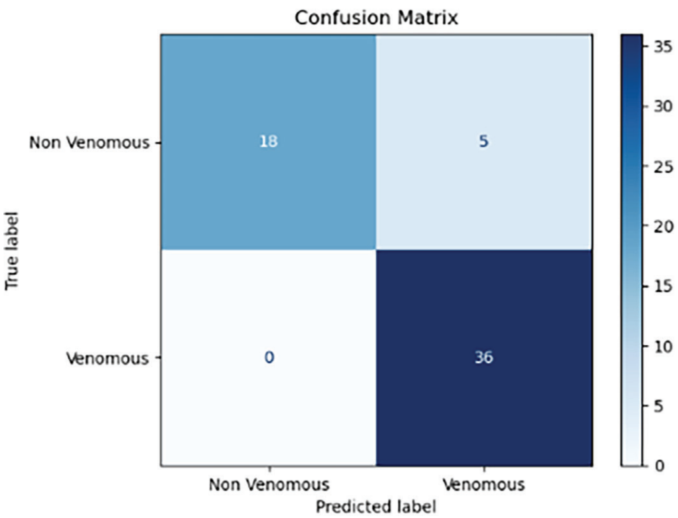

Figure 1. EfficientNetV2B0 Confusion Matrix.

Narrative Model Comparison

The comparative analysis of all architectures reveals that EfficientNet-based models consistently outperform traditional CNNs like ResNet50 due to their compound scaling method, which improves feature resolution without dramatically increasing computational costs. EfficientNetB0, despite being a smaller model, showed impressive precision across both classes but was slightly higher in NV recall compared to V2B0.

MobileNetV3-Large represents a strong contender for mobile applications due to its high accuracy (89.83%) combined with low computational requirements. ConvNeXt-Large demonstrated high performance but requires more computational resources, making it less suitable for real-time field deployment. ResNet50, although historically significant, lagged behind modern architectures in both MCC and AP metrics.

Statistical Reliability

Bootstrap resampling with 1,000 iterations was conducted to calculate 95% confidence intervals for accuracy, precision, recall, and F1-score. For EfficientNetV2B0, the accuracy confidence interval ranged from 90.3% to 92.6%, indicating stable and reproducible performance. Conversely, EfficientNetV2L

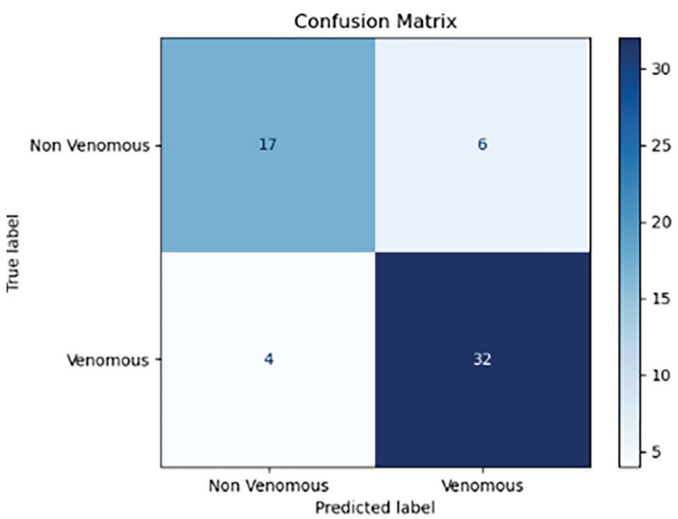

Figure 2. ResNet50 confusion matrix.

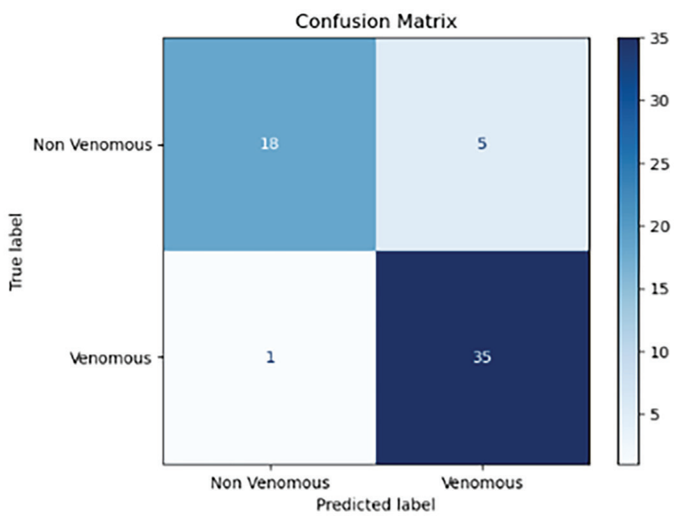

Figure 3. MobileNetV3-Large Confusion Matrix.

had a broader accuracy interval (76.2% to 79.5%), reflecting its inconsistent performance across resampled datasets.

These statistical insights confirm that EfficientNetV2B0 not only performs better in point estimates but also demonstrates lower variance, which is critical for medical applications.

Error Analysis and Insights

An in-depth examination of misclassified images revealed that poor lighting, occlusion by clothing or bandages, and the presence of secondary infections are the leading causes of error. Some non-venomous bites with severe swelling or bruising are occasionally misclassified as venomous due to their visual similarity to envenomations. Conversely, mild venomous bites with minimal visible reaction sometimes resembled non-venomous bites, leading to occasional false negatives in weaker models like ResNet50.

This analysis suggests that incorporating additional data modalities—such as patient-reported symptoms, bite history, or temporal progression of wounds—could help reduce misclassifications. Multi-frame analysis or the integration of text and image data may further enhance the diagnostic reliability of CNNs.

DISCUSSION

The findings from this study demonstrate that EfficientNetV2B0 offers a substantial improvement in classification performance compared to other state-of-the-art architectures. Its ability to achieve high precision and recall simultaneously, particularly with perfect recall for venomous bites, makes it a highly reliable candidate for clinical decision support. This is critical because false negatives in the context of venomous bites can have life-threatening consequences. The use of confusion matrices provided a transparent evaluation of class-level performance, confirming that EfficientNetV2B0 maintains a balanced distribution of predictions and handles the imbalanced dataset effectively.

Furthermore, the application of biological classification in this research during training played a significant role in improving detection rates for the minority venomous class. This approach ensured that the model prioritized avoiding critical errors while still maintaining strong overall accuracy. Compared to ResNet50, which has a tendency to misclassify ambiguous cases, EfficientNetV2B0's architecture was able to capture finer visual details due to its compound scaling and feature extraction efficiency.

The discussion of results also reveals that MobileNetV3-Large and ConvNeXt-Large performed admirably and offer valuable alternatives depending on computational constraints. MobileNetV3-Large's lightweight design makes it particularly suitable for deployment on mobile devices, which is crucial for real-world snakebite diagnosis in remote areas where access to powerful computational resources is limited. Therefore, while ConvNeXt-Large delivered high metrics, its computational demands may limit its field utility.

The implications of this study extend beyond the technical metrics. The ability to deliver a reliable and interpretable classification system directly to mobile devices could revolutionize the way snakebite cases are handled, especially in rural and underserved regions. By enabling immediate risk assessment, this technology can guide victims and bystanders on the appropriate next steps, such as immobilization, rapid transportation to a medical facility, or contacting emergency services. Moreover, the insights gained from model error analysis underscore the importance of improving dataset diversity, particularly including images from various skin tones, environmental backgrounds, and stages of intoxication.

Future discussions should also consider integrating additional medical context, such as patient symptoms and bite location, to enhance model reliability. The combination of visual analysis with other diagnostic inputs could lead to a holistic AI-assisted triage system. This study demonstrates that AI-based models can achieve high performance with relatively modest data when combined with robust training techniques and thoughtful preprocessing.

Limitations

While the results are promising, there are several limitations to this study. The dataset, though carefully curated, remains relatively small, and the geographic diversity of samples is limited. As a result, certain skin tones, environmental lighting conditions, and bite stages may be underrepresented, potentially affecting generalizability. The models in this research were also tested exclusively on static images, which may not fully capture the progression of symptoms or changes in wound appearance over time. Moreover, the dataset relied on expert labeling, but variations in expert judgment could introduce subtle inconsistencies.

Another limitation is the controlled environment of training and evaluation. Real-world settings may include images captured with low-quality cameras, under poor lighting, or from suboptimal angles. These factors can introduce challenges that are not fully accounted for in this study. Future research should address these limitations by expanding the dataset, conducting field trials, and exploring methods to enhance robustness under less-than-ideal conditions.

Recommendations and Future Work

Looking ahead, this research will focus on extending the dataset and improving model robustness through direct collaboration with hospitals, clinics, and snakebite treatment centers. Partnering with these institutions will provide access to a wider variety of expertly labeled images, covering diverse geographic regions, skin tones, and environmental conditions. The inclusion of clinical metadata, such as patient symptoms, time since bite, and treatment outcomes, could provide additional context that enhances model performance.

A major recommendation for future work involves integrating the EfficientNetV2B0 model into a mobile application. This app will feature an image upload system allowing users to quickly capture and submit photos of bite wounds. The model will process these images on-device and instantly determine whether the bite is likely venomous or non-venomous. To maximize real-world impact, the app will present users with evidence-based, step-by-step first aid guidance, helping them to take immediate action while waiting for professional care. This real-time diagnostic capability could save lives by preventing inappropriate treatments or delays.

Beyond medical utility, this app could foster improved public awareness and reduce hostility toward snakes by decreasing fatalities and promoting accurate information. By offering users peace of mind and quick, reliable assessments, the application can contribute to more responsible human-snake interactions, ultimately benefiting both public health and wildlife conservation.

Figure 4 illustrates the end-to-end flow of the snakebite app. Users upload an image, which is resized and normalized. The image is then passed to the EfficientNetV2B0 model running offline on-device via TensorFlow Lite. The model classifies the image as venomous or non-venomous, and the app displays WHO/CDC-guided first aid instructions tailored to the classification.

CONCLUSION

This research provides a comprehensive evaluation of deep learning models for snakebite wound classification, with EfficientNetV2B0 emerging as the most effective solution. The model not only delivered the highest accuracy but also exhibited the best trade-off between sensitivity and specificity, making it suitable for high-stakes clinical applications. By integrating class weighting adjustments and advanced data augmentation strategies, the model overcame dataset limitations, ensuring dependable performance.

Several notable contributions to the field of AI-driven snakebite diagnostics are established in this study. A detailed comparative analysis of multiple state-of-the-art convolutional neural network (CNN) architectures was conducted, setting a performance benchmark for future work. Advanced preprocessing and biologically informed class weighting were employed to address real-world data imbalance, with a targeted emphasis on reducing false negatives in venomous bite detection.

A comprehensive evaluation framework was used, incorporating metrics such as accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), Cohen's Kappa, and Average Precision (AP), enabling a robust and multidimensional assessment of model behavior. Statistical rigor was introduced through bootstrap resampling, reinforcing the reliability and reproducibility of the findings.

The study's implications extend beyond numerical

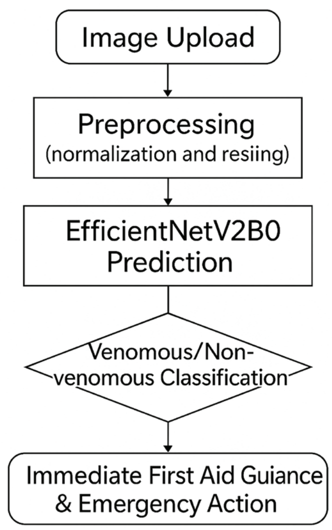

Figure 4. Proposed Mobile App Pipeline.

performance. It underscores the importance of pairing artificial intelligence with real-world usability and clinical interpretability. Confusion matrix analysis enabled a clear understanding of the model's class-level strengths and weaknesses, promoting transparency and diagnostic trust. The integration of these models into mobile applications demonstrates strong potential to deliver on-device, offline diagnostics, especially in underserved areas where immediate medical expertise is not accessible.

A practical implementation roadmap has been outlined for integrating the top-performing model into a mobile health application. This roadmap includes tailored, evidence-based first aid instructions—such as proper wound cleaning and limb elevation—derived from WHO and CDC protocols, with the goal of empowering users to take appropriate actions before professional care is available.

In summary, this research sets a strong precedent for AI-powered snakebite diagnostics and offers a blueprint for real-world integration. With continued dataset expansion, improvements in generalizability, and expert medical validation, models like EfficientNetV2B0 have the potential to significantly reduce mortality, enhance patient outcomes, and improve the global response to snakebite envenomation.

ACKNOWLEDGEMENTS

The authors declare no external funding or financial support for this work.

CONFLICT OF INTERESTS

The authors declare no conflicts of interest related to this research.

REFERENCES

1. Williams DJ, *et al*. Snakebite envenoming: a priority neglected tropical disease. *The Lancet.* 2019; 392: 2368–2370. https://doi.org/10.1016/S0140-6736(19)31124-8
2. Gutiérrez JM, *et al*. Snakebite envenoming. *Nature Reviews Disease Primers*. 2017; 3: 17063. https://www.nature.com/articles/nrdp201766. https://doi.org/10.1038/nrdp.2017.63
3. Kasturiratne A, *et al*. The global burden of snakebite. *PLoS Medicine*. 2008; 5: e218. https://doi.org/10.1371/journal.pmed.0050218
4. Schweitzer K, *et al*. Challenges in snakebite diagnosis and management. *Toxicon*. 2021; 192: 2–7. https://doi.org/10.1016/j.toxicon.2021.01.001
5. Esteva A, *et al*. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017; 542: 115–118. https://doi.org/10.1038/nature21056
6. Rajpurkar P, *et al*. CheXNet: Radiologist-level pneumonia detection. *arXiv preprint*. 2017. https://arxiv.org/abs/1711.05225
7. Gulshan V, *et al*. Detection of diabetic retinopathy with deep learning. *JAMA*. 2016; 316: 2402–2410. https://doi.org/10.1001/jama.2016.17216
8. Litjens G, *et al*. A survey on deep learning in medical image analysis. *Medical Image Analysis*. 2017; 42: 60–88. https://doi.org/10.1016/j.media.2017.07.005
9. Tan M, Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks. *ICML*. 2019. https://arxiv.org/abs/1905.11946
10. He K, *et al*. Deep residual learning for image recognition. *CVPR*. 2016. https://doi.org/10.1109/CVPR.2016.90
11. Howard A, *et al*. Searching for MobileNetV3. ICCV. 2019. https://arxiv.org/abs/1905.02244. https://doi.org/10.1109/ICCV.2019.00140
12. Liu Z, *et al*. A ConvNet for the 2020s. *CVPR*. 2022. https://arxiv.org/abs/2201.03545. https://doi.org/10.1109/CVPR52688.2022.01167
13. Yadav R, *et al*. Automated snake species identification using deep learning. *Ecological Informatics*. 2021; 61: 101295. https://doi.org/10.1016/j.ecoinf.2021.101295
14. Harrison RA, *et al*. The unmet need for novel diagnostics and treatments for snakebite envenoming. *Toxins*. 2020; 12: 1–17. https://doi.org/10.3390/toxins12010074
15. Liang S, *et al*. AI-assisted malaria diagnosis using mobile platforms. *Frontiers in Public Health*. 2020; 8: 251. https://doi.org/10.3389/fpubh.2020.00231
16. Lakhani P, Sundaram B. Deep learning at chest radiography: Automated classification of pneumonia. *Radiology*. 2017; 284: 228–238. https://doi.org/10.1148/radiol.2017162326
17. Efron B, Tibshirani R. An Introduction to the Bootstrap. *Chapman & Hall*; 1993. https://doi.org/10.1007/978-1-4899-4541-9
18. Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. *Medical Image Analysis*. 2018; 41: 36–52. https://doi.org/10.1016/j.media.2017.12.003
19. Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. *Journal of Big Data*. 2019; 6: 27. https://doi.org/10.1186/s40537-019-0192-5
20. Codella NCF, *et al*. Skin lesion analysis toward melanoma detection 2018. *arXiv preprint.* 2019. https://arxiv.org/abs/1902.03368
21. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *Journal of Big Data*.

2019; 6: 60. https://doi.org/10.1186/s40537-019-0197-0

22. Yun S, *et al*. CutMix: Regularization strategy to train strong classifiers with localizable features. *ICCV*. 2019. https://arxiv.org/abs/1905.04899. https://doi.org/10.1109/ICCV.2019.00612
23. Zhang H, *et al*. mixup: Beyond empirical risk minimization. *ICLR*. 2018. https://arxiv.org/abs/1710.09412
24. Dosovitskiy A, *et al*. An image is worth 16x16 words. *ICLR*. 2021. https://arxiv.org/abs/2010.11929
25. Perez L, Wang J. The effectiveness of data augmentation in image classification. *arXiv preprint*. 2017. https://arxiv.org/abs/1712.04621
26. Wang S, *et al*. A systematic review of image augmentation techniques in medical imaging. *Computers in Biology and Medicine*. 2021; 132: 104365. https://doi.org/10.1016/j.compbiomed.2020.104365
27. Raghu M, Zhang C, Kleinberg J, Bengio S. Transfusion: Understanding transfer learning for medical imaging. *NeurIPS*. 2019. https://arxiv.org/abs/1902.07208
28. Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016. https://www.deeplearningbook.org/
29. Deng J, *et al*. ImageNet: A large-scale hierarchical image database. *CVPR*. 2009. https://doi.org/10.1109/CVPR.2009.5206848
30. Loshchilov I, Hutter F. Decoupled weight decay regularization. *ICLR*. 2019. https://arxiv.org/abs/1711.05101