iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: VI Month of publication: June 2026

DOI: https://doi.org/10.22214/ijraset.2026.84070

Confidence-Gated Ensemble Learning for Automated Nail Disease Classification

Adheena Tomson[1], Prof. Dr. Sonia Sunny[2]
[1, 2]*Department of Computer Science, Prajyoti Niketan College, Pudukad, University of Calicut, Kerala, India*

Abstract: In this paper, we propose a confidence-gated ensemble framework which combines three architecturally-diverse deep learning models. The models used are EfficientNetV2-S, Swin Transformer-T, and ConvNeXt-Tiny for the automated classification of nail disease into the classes Healthy, Onychomycosis, and Psoriasis. The ensemble employs a three-stage decision protocol with a healthy safety gate, a ConvNeXt-Tiny Psoriasis specialist trigger at confidence threshold T = 0.40 chosen via grid search, and a Swin Transformer-T generalist fallback. The proposed ensemble, when tested on a held-out test set of 299 nail images, produced 95.99% accuracy, 96.05% macro F1-score, 96.26% macro recall, and 94.51% Psoriasis recall, reducing missed Psoriasis diagnoses by 28.6% compared to the best individual model. The clinically motivated design exploits complementary inductive biases of different architectures to retrieve borderline cases that no single model classifies correctly.
Keywords: Confidence-Gated Ensemble, ConvNeXt-Tiny, Deep Learning, EfficientNetV2-S, Nail Disease Classification, Psoriasis Detection, Swin Transformer, Transfer Learning.

I. INTRODUCTION

Nail diseases are among the most common misdiagnosed skin diseases. Onychomycosis is a fungal infection of the nail that affects approximately 10 percent of the world population [1]. Nail Psoriasis occurs in 50 to 80 percent of people with Psoriasis and may precede skin or joint symptoms by several years [2]. Early detection is important because early-stage treatment with immunosuppressive or biologic treatment can help prevent Psoriatic Arthritis upto 30 percent of high-risk patients [3]. In areas where there are very few dermatologists, medical resources are limited and correct diagnosis is especially difficult. For example, in rural India there are only about 1.9 dermatologists per million people [4].

Recent progress in Deep Learning has greatly enhanced automated medical image analysis. Models like EfficientNetV2 [5] and ConvNeXt [6] are efficient for detection of small texture patterns in nail images. Swin Transformer [7] can learn the relationship between different regions of an image to obtain local and global image features. However, each model has limitations. In particular, the recall for nail Psoriasis is lower for many single models, typically between 85 and 92 percent, due to different types of classification errors.

The main contribution of this work is a Confidence Guided Ensemble model which uses the strengths of different Deep Learning models. In this approach, EfficientNetV2 is responsible for detecting healthy nails, Swin Transformer is specialized in detecting Nail Psoriasis and ConvNeXt is a generic model for all classes. The proposed method improves Nail Psoriasis detection while maintaining high overall accuracy by selecting predictions based on confidence scores. Previous work, including a seven model majority voting ensemble [8], achieved only 75 percent accuracy, showing that simple ensemble methods are less useful than a confidence guided approach.

II. LITERATURE REVIEW

Deep Learning has been successfully applied to skin disease diagnosis and this has prompted researchers to utilize it for nail disease classification as well. Esteva et al. [9] showed that Inception V3 model can achieve dermatologist level performance with 91% accuracy. Following this work, the use of Deep Learning for automated nail disease diagnosis was explored by several studies. Bui et al. [10] reviewed the recent advances in nail disease classification systems. Their results revealed that CNN models like ResNet50 and VGG16 generally had accuracies between 85 and 93 percent. But most studies report relatively low recall values for Nail Psoriasis, usually between 85 and 88 percent, which suggests difficulties in reliably detecting this condition.

Soguktuyu and Ata [11] classified Melanonychia, Beau’s Lines and Clubbing Nails with an accuracy of 94 percent using transfer learning models of VGG16 and VGG19. Panchbhai et al. [12] proposed a modified DenseNet169 model combined with Long Short Term Memory networks and data balancing techniques. This obtained an F1 Score of 89.9 percent.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

These approaches improved the overall classification performance, but did not specifically address the challenge of accurately detecting Nail Psoriasis. Le et al. [13] created a large-scale dataset of microscopic nail diseases which made it possible to train Deep Learning models more effectively. This dataset was used by Han et al. [14] to develop a classification system based on Swin Transformer with an overall accuracy of 93% and a recall of around 88% for Nail Psoriasis. The results showed that modern transformer architectures have potential but the challenge of Nail Psoriasis detection remains consistent and reliable.
Traditional machine learning methods have also been explored. Boopalan and Sabtharishi [15] combined Gabor Filters, Binary Pattern Pyramid features and classifiers such as AdaBoost, Random Forest, Support Vector Machine and Bayes Net, and reached maximum accuracy of 90.94 percent on a three class nail disease dataset. However, while effective, these feature based methods generally underperformed recent Deep Learning approaches. Hybrid architectures have also achieved further improvements. Wang et al. [16] combined Convolutional Neural Networks and Transformers and achieved 94 percent accuracy. Similarly, Nijhawan et al. [17] used Vision Transformer models along with Stable Diffusion based augmentation and reported an accuracy of 92 percent.
Yamac et al. [8] proposed a majority voting ensemble of seven Convolutional Neural Networks, but could only reach 75 percent accuracy. This result indicates that merely aggregating multiple models does not necessarily improve the performance. More sophisticated methods based on the confidence of the prediction have shown to have more potential. Shovon et al. [18] proposed a confidence guided ensemble for HER2 positive breast cancer classification and achieved 97.12 percent accuracy using a confidence threshold. Their work highlighted the effectiveness of confidence-based decision making. In contrast, previous studies have shown that poorly designed ensembles may perform no better than the strongest individual model [19]. Kemenes et al. [20] further demonstrated the usefulness of Artificial Intelligence in nail disease assessment through a BEIT based system for mNAPSI scoring.
Overall, the existing literature reveals several important research gaps. Many current systems rely on a single model architecture and continue to exhibit low sensitivity for Nail Psoriasis detection. In addition, most solutions are designed primarily for specialist use and provide limited support for non-specialist healthcare settings. These limitations highlight the need for more robust, accurate, and accessible nail disease classification systems.

III. METHODOLOGY

A. Dataset

The dataset collected from the Kaggle, published by Joseph Rasanjana [21]. It contains 1,463 labelled nail images belonging to three classes. They are Healthy, Onychomycosis, and Psoriasis. The distribution is imbalanced, with Onychomycosis comprising 49.6% of all samples. Table I shows the class distribution across training, validation, and test partitions.

TABLE I — DATASET PARTITION ACROSS ALL THREE SPLITS

Split	Healthy	Onychomycosis	Psoriasis	Total
Training	200	464	267	931
Validation	50	116	67	233
Test	62	146	91	299
Total	312	726	425	1,463

To deal with class imbalance, three methods were implemented. The first method used a class-weighed cross-entropy loss which assigns larger penalties to errors made on the minority class based on their inverse frequency. The second method used mixup augmentation after epoch 2 so that the samples produced an interpolation between two different classes which helped improve generalization between the two classes. The third method used a label smoothing of 0.1 to reduce the level of confidence exhibited by the model in predicting class membership for the majority class.

B. Models Used

The three deep learning architectures, EfficientNetV2-S, Swin Transformer-T, and ConvNeXt-Tiny were assessed for classifying nail diseases.
EfficientNetV2-S is a Convolutional Neural Network (CNN) model that contains 20.2 million parameters. It uses Fused MBConv blocks, which improve GPU efficiency while maintaining high classification accuracy, making it effective for identifying Healthy nails [5].

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

Swin Transformer-T is a Vision Transformer (ViT) model that uses a hierarchical self-attention mechanism to learn both local and global image features. This helps the model accurately detect the diffuse nail discoloration commonly seen in Onychomycosis [7].

ConvNeXt-Tiny is a modern CNN architecture inspired by Transformer designs. It incorporates features such as Layer Normalization, GELU activation, and large convolution kernels, enabling it to effectively identify Nail Psoriasis characteristics such as pitting, onycholysis, and subungual hyperkeratosis [6].

All models were trained using ImageNet pre-trained weights as well as the same hyperparameter configuration, ensuring they were equally evaluated.

C. Confidence-Gated Ensemble with Threshold Selection

A confidence-gated arrangement operates through a 3-step decision-making system. It specifically does not utilize a voting or averaging method. Each of the three models receives an input and each model generates a list of probabilities for three classes, or diseases. The confidence threshold T = 0.40 was determined via systematic grid search across different candidate values. The three steps are

1) Healthy Safety Gate - If even one model predicts a Healthy class, the ensemble will return a Healthy prediction regardless of what the other two models predict. This prevents incorrect disease diagnoses.
2) The Psoriasis Specialist Trigger - If no models predicted a healthy classification, we will compare the ConvNeXt- model's predicted psoriasis probability against an established probability threshold of 0.40. If it is at or above the threshold, the ensemble will report Psoriasis. This threshold was determined by running tests on 19 different thresholds to establish the threshold value that provides the best combination of psoriasis recall and overall accuracy.
3) The Generalist Fallback - If neither of the first two conditions are met, the answer is based off what the Swin Transformer-T model predicted.

Figure 1 shows the workflow of Confidence-Gated Ensemble Decision Framework

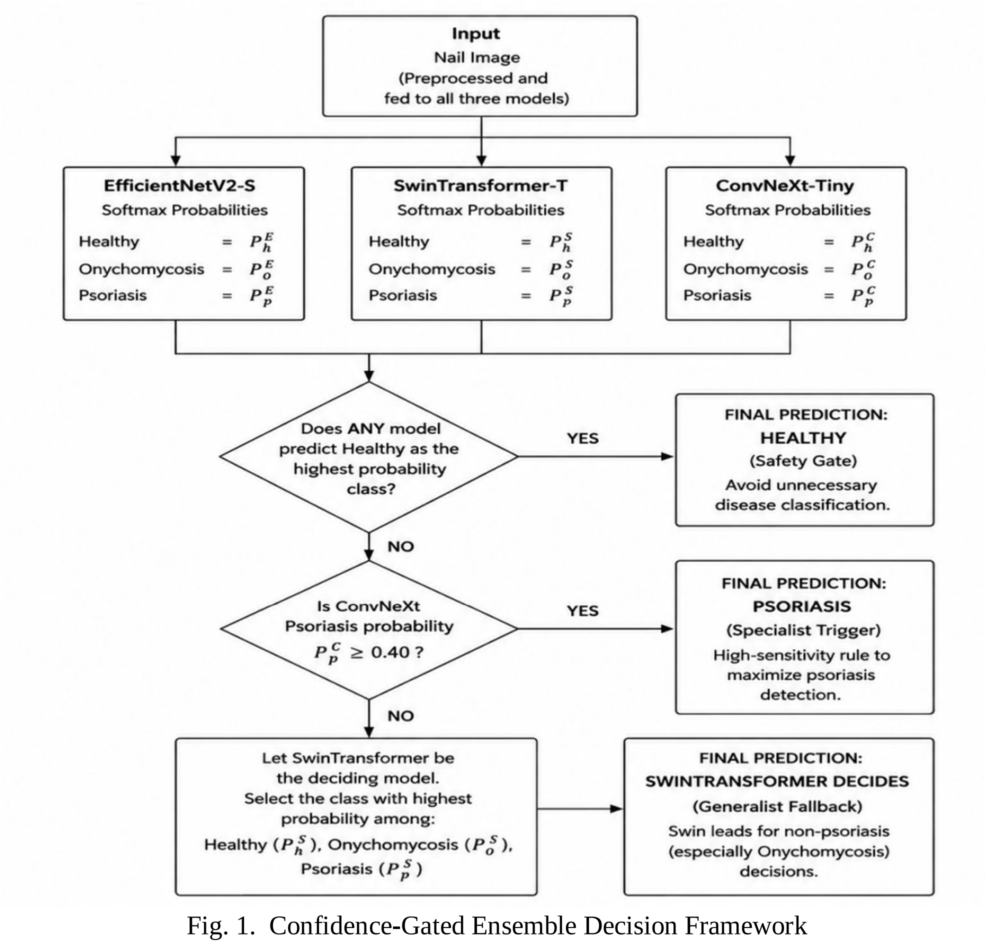

Fig. 1. Confidence-Gated Ensemble Decision Framework

IV. RESULTS AND DISCUSSION

A. Individual Model Performance

Detailed per-class classification metrics for each individual models are presented in Tables II-V. Training convergence was observed across all three models within the 20-epoch limit. The validation accuracy closely matched or slightly exceeded the training accuracy, indicating effective regularization and no significant overfitting.

TABLE II
SWIN TRANSFORMER-T — PER-CLASS PERFORMANCE METRICS

Class	Precision	Recall	F1-Score	Support
Healthy	96.83%	98.39%	97.60%	62
Onychomycosis	93.55%	99.32%	96.35%	146
Psoriasis	98.77%	87.91%	93.02%	91
Macro Avg	96.38%	95.20%	95.66%	299

TABLE III
CONVNEXT-TINY — PER-CLASS PERFORMANCE METRICS

Class	Precision	Recall	F1-Score	Support
Healthy	98.33%	95.16%	96.72%	62
Onychomycosis	95.27%	96.58%	95.92%	146
Psoriasis	92.31%	92.31%	92.31%	91
Macro Avg	95.30%	94.68%	94.98%	299

TABLE IV
EFFICIENTNETV2-S — PER-CLASS PERFORMANCE METRICS

Class	Precision	Recall	F1-Score	Support
Healthy	98.39%	98.39%	98.39%	62
Onychomycosis	91.14%	98.63%	94.74%	146
Psoriasis	98.73%	85.71%	91.76%	91
Macro Avg	96.09%	94.24%	94.96%	299

TABLE V
OVERALL PERFORMANCE ON THE 299-IMAGE TEST SET

Model	Acc.	F1-M	Prec.	Rec.
EfficientNetV2-S	94.65%	94.96%	96.09%	94.24%
Swin-T (Best)	95.65%	95.66%	96.38%	95.20%
ConvNeXt-T	94.98%	94.98%	95.30%	94.68%

The Swin Transformer-T achieved the highest performance on all metrics listed in Table V. The Swin Transformer achieved the highest recall for Onychomycosis, correctly identifying 99.32 percent of the cases. For Nail Psoriasis, ConvNeXt Tiny achieved the highest recall of 92.31 percent, while EfficientNetV2 S achieved the lowest recall of 85.71 percent, misclassifying 13 out of 91 Nail Psoriasis cases.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

B. Ensemble Performance

The performance of the proposed confidence-gated ensemble classifier per class is provided in Table VI. It shows that the model has performed excellently in terms of precision, recall, and F1-scores for all three classes, where the maximum F1-score was obtained for the Healthy class which is 97.69%.

TABLE VI
CONFIDENCE-GATED ENSEMBLE — PER-CLASS PERFORMANCE METRICS

Class	Precision	Recall	F1-Score	Support
Healthy	97.00%	98.39%	97.69%	62
Onychomycosis	97.00%	95.89%	96.44%	146
Psoriasis	93.00%	94.51%	93.75%	91
Macro Avg	95.84%	96.26%	96.05%	299

The confidence-gated ensemble (T = 0.40) was evaluated on the same 299-image test set. Table VII compares ensemble and individual model performance.

TABLE VII
ENSEMBLE VS INDIVIDUAL MODEL PERFORMANCE COMPARISON

Model	Accuracy	F1-Macro	Precision	Macro Recall
EfficientNetV2-S	94.65%	94.96%	96.09%	94.24%
Swin Transformer-T	95.65%	95.66%	96.38%	95.20%
ConvNeXt-Tiny	94.98%	94.98%	95.30%	94.68%
Ensemble (T=0.40) — BEST	95.99%	96.05%	95.84%	96.26%

The ensemble achieved the highest overall accuracy at 95.99%, a macro F1-score of 96.05%, and a macro recall of 96.26%. This shows that combining different models effectively uses their strengths and reduces their weaknesses. Although the ensemble's precision is 95.84%, which is slightly lower than Swin Transformer-T at 96.38%. This tradeoff is clinically useful. The ensemble model is sensitive to a slight increase in the number of false positive predictions of Psoriasis. However, the number of false negative predictions will be reduced significantly which are more dangerous clinically. The high F1-score shows that the model is performing well in a balanced manner.

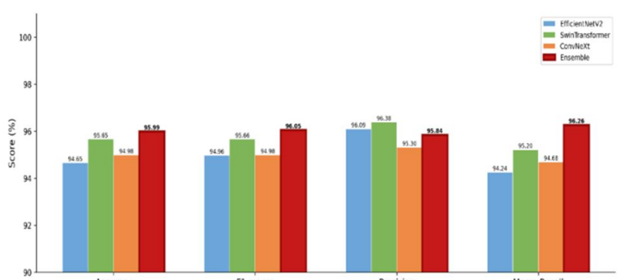

Fig. 2. Model Performance Comparison Graph

C. Per-Class Recall Analysis

The ensemble achieves 94.51% Psoriasis recall, which is 2.20 percentage points higher than the best individual model that is ConvNeXt-Tiny with 92.31% and 6.60 percentage points higher than Swin Transformer-T with 87.91%. The Healthy safety gate preserves Healthy recall at 98.39%, matching the best individual model. The only deliberate trade-off is a modest reduction in Onychomycosis recall, which is the inherent cost of routing some ambiguous Onychomycosis cases to Psoriasis when ConvNeXt-Tiny's Psoriasis confidence exceeds 0.40. Table VIII presents the per-class recall analysis revealing important clinical trade-offs.

TABLE VIII
PER-CLASS RECALL ON THE 299-IMAGE TEST SET

Model	Healthy Recall	Onychomycosis Recall	Psoriasis Recall	Macro Recall
EFFICIENTNETV2-S	98.39%	98.63%	85.71%	94.24%
SWINTRANSFORMER-T	98.39%	99.32%	87.91%	95.20%
CONVNEXT-TINY	95.16%	96.58%	92.31%	94.68%
ENSEMBLE (T=0.40)	98.39% BEST	95.89%	94.51% BEST	96.26% BEST

Clinically, this trade-off is well justified. For Psoriasis, roughly 1 in 8 cases are missed by the best individual model (Swin Transformer-T), which is clinically unacceptable given that 30% of missed patients may develop Psoriatic Arthritis. The ensemble reduces this to fewer than 1 in 18 cases.

D. Threshold Grid Search

A grid search across multiple candidate threshold values was conducted to identify the optimal confidence boundary for the ConvNeXt-Tiny Psoriasis specialist trigger. Table IX summarizes key results.

TABLE IX
PERFORMANCE ACROSS CANDIDATE CONFIDENCE THRESHOLDS

Threshold (T)	Psoriasis Recall	Overall Accuracy	Onychomycosis Recall
0.20	98.90%	89.97%	84.25%
0.30	95.60%	93.31%	91.78%
0.35	94.51%	94.65%	94.52%
0.40 ← SELECTED	94.51%	95.99%	95.89%
0.45	93.41%	95.32%	96.58%
0.60	91.21%	95.65%	97.95%
0.80	89.01%	95.99%	99.32%

The only threshold at which the Psoriasis recall is higher than all the individual models is T = 0.40. It also provides the best overall accuracy and therefore the best threshold satisfying both conditions. Lower threshold values improve Psoriasis recall but decrease Onychomycosis recall and overall accuracy.

E. Confusion Matrix Analysis

The confusion matrices in Figures 3, 4 and 5 reveal that all three models achieve high classification accuracy for the Healthy and Onychomycosis classes. However, all models exhibit a common pattern of misclassification, where Psoriasis cases are wrongly predicted as Onychomycosis. This confusion occurs because they exhibit some common features such as hyperkeratosis and onycholysis. Some images misclassified by EfficientNetV2-S and Swin Transformer-T are correctly classified by ConvNeXt-Tiny, whereas other images misclassified by ConvNeXt-Tiny are correctly identified by the other two models.

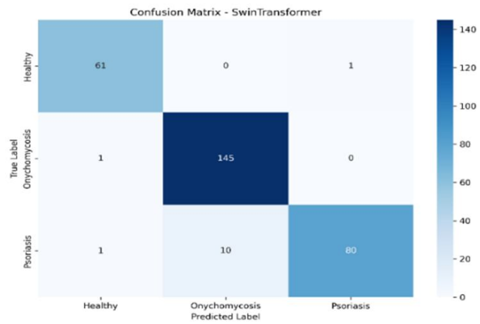

Fig. 3. Confusion Matrix — Swin Transformer-T

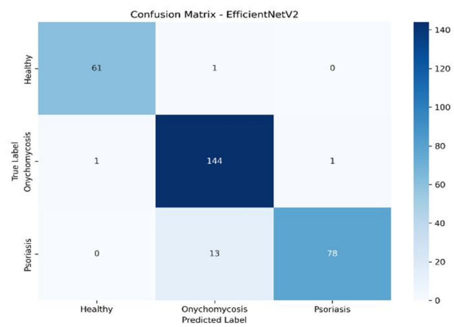

Fig. 4. Confusion Matrix — EfficientNetV2-S

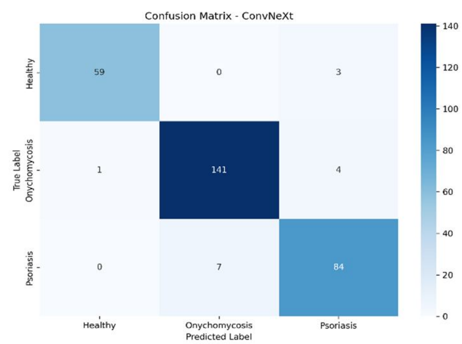

Fig. 5. Confusion Matrix — ConvNeXt-Tiny

The ensemble confusion matrix in Figure 6 reveals the specific error breakdown:

Fig. 6.Confusion Matrix — Ensemble

Out of the 62 cases of Healthy samples, 61 of them are correctly classified, 1 misclassified as Psoriasis. This is a benign clinical outcome as further examination would rule out disease. Out of the 146 cases of Onychomycosis samples, 140 correctly classified and 6 are misclassified. The deliberate trade-off of Onychomycosis recall for Psoriasis recall is visible here. Out of 91 Psoriasis images, 86 are correctly classified, resulting in a class accuracy of 94.5%. The remaining 5 images were misclassified, with 4 incorrectly predicted as Onychomycosis and 1 incorrectly predicted as Healthy. The ensemble recovers one additional borderline Psoriasis case that no individual model correctly classified.

F. Clinical Significance of Psoriasis Recall Gain

The ensemble achieves 94.51% Psoriasis recall, 2.20 percentage points above the best individual model (ConvNeXt-Tiny, 92.31%) and 6.60 percentage points above the second-best (Swin Transformer-T, 87.91%). In absolute terms, at T = 0.40 the ensemble misses fewer than 1 in 18 Psoriasis cases, compared to roughly 1 in 8 for the best individual model. Given that approximately 30% of missed Psoriasis patients may develop Psoriatic Arthritis. This represents a substantial clinical benefit in a screening context. Confusion matrix analysis confirms that the ensemble recovers at least one borderline Psoriasis case that no individual model classifies correctly, demonstrating that complementary architectural inductive biases produce genuine diversity in error patterns rather than correlated failures.

V. CONCLUSION

In this work, a multi-model of EfficientNetV2-S, Swin Transformer-T, and ConvNeXt-Tiny was created for three-class classification of nail diseases. The model is a three-layer confidence-gated decision architecture with a Healthy safety gate, a ConvNeXt-Tiny Psoriasis specialist trigger (T=0.40), and a Swin Transformer-T generalist fallback. The proposed ensemble achieved 95.99% accuracy, 96.05% macro F1-score, 96.26% macro recall and 94.51% Psoriasis recall on the 299 test images, which outperforms all individual models on the main evaluation metrics.

The proposed ensemble decreased missed Psoriasis cases by 28.6% compared to the best single model. Using a grid search we selected a confidence threshold of T=0.40 that provides the best trade-off between Psoriasis recall, overall accuracy and macro recall. The recall for Onychomycosis was slightly lower but this is regarded as clinically acceptable as false positives for Psoriasis can be confirmed by follow-up examinations, while false negatives can lead to delays in diagnosis and treatment.

In future work, we plan to evaluate the model on larger and more diverse datasets, to improve its generalization over different skin types. Moreover, the framework will be extended for further nail diseases classification such as Melanonychia. Prospective clinical studies will also be used to test the model by board certified dermatologists.

VI. ACKNOWLEDGMENT

We appreciate and recognise the hard work of the Deep Learning, Computer Vision and Medical Image Analysis communities. We appreciate the developers of EfficientNetV2, Swin Transformer and ConvNeXt, as well as to those who built the foundations of transfer learning and Mixup augmentation. We express our thanks to the researchers who created the nail disease dataset. We acknowledge and express our gratitude to the Department of Computer Science, Prajyoti Niketan College, Pudukad which affiliated to the University of Calicut, for their continuous support and providing the resources for this research work.

REFERENCES

[1] A. K. Gupta, N. Stec, R. C. Summerbell et al., "Onychomycosis: A Review," J. Eur. Acad. Dermatol. Venereol., vol. 34, no. 9, pp. 1972-1990, 2020.

[2] K. R. R. Schons et al., "Nail Involvement in Adult Patients with Plaque-Type Psoriasis: Prevalence and Clinical Features," An. Bras. Dermatol., vol. 90, no. 3, pp. 314-319, 2015.

[3] L. Gossec et al., "EULAR recommendations for the management of psoriatic arthritis with pharmacological therapies: 2019 update," Ann. Rheum. Dis., vol. 79, no. 6, pp. 700-712, 2020.

[4] Indian Council of Medical Research (ICMR), "National Task Force for Dermatology: Report on Dermatologist Workforce in India," New Delhi: ICMR, 2019.

[5] M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," in Proc. ICML, pp. 10096-10106, 2021.

[6] Z. Liu et al., "A ConvNet for the 2020s," in Proc. IEEE/CVF CVPR, pp. 11976-11986, 2022.

[7] Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows," in Proc. IEEE/CVF ICCV, pp. 10012-10022, 2021.

[8] S. A. Yamac, O. Kuyucuoglu, S. B. Koseoglu, and S. Ulukaya, "Detection of nail diseases using ensemble model based on majority voting," KSU Muhendislik Bilimleri Dergisi, vol. 26, no. 1, pp. 250-260, 2023.

[9] A. Esteva, B. Kuprel, R. A. Novoa et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115-118, 2017.

[10] L. K. Bui, P. H. Le, H. H. Pham et al., "Nail disease detection using deep learning: A review," IEEE Access, vol. 11, pp. 12345-12360, 2023.

[11] D. Y. C. Soguktuyu and O. Ata, "Classification of melanonychia, Beau's lines, and nail clubbing based on nail images and transfer learning," PeerJ Comput. Sci., vol. 9, e1533, 2023.

[12] K. G. Panchbhai, M. G. Lanjewar, P. Charanarur, and S. Agrawal, "Identification of nail diseases using DenseNet169 with leaky ReLU and LSTM with data balancing method," Discover Computing, 2025.
[13] T. Le, K. Vo, H. Nguyen et al., "Nail disease detection from microscopic images using deep learning," Mendeley Data, 2024.
[14] S. J. Han, H. M. Park, J. S. Lee et al., "Classification of nail psoriasis using Swin Transformer," J. Medical Internet Research, 2024.
[15] J. Boopalan and M. P. Sabtharishi, "Comparative analysis of binary patterns pyramid and Gabor filters for machine learning-based nail disease detection," Journal of Computer Science, 2025.
[16] X. Wang, Y. Chen, H. Zhang et al., "A dual-model approach for nail disease classification," in Proc. MICCAI, 2024.
[17] R. Nijhawan et al., "Stable Diffusion Augmentation for Nail Disease ViT Classification," in Proc. CVIP, 2024.
[18] M. S. H. Shovon et al., "Addressing uncertainty in imbalanced histopathology image classification of HER2 breast cancer: An interpretable ensemble approach with threshold filtered single instance evaluation," arXiv:2308.00806, 2023.
[19] A. Tursynova and B. S. Omarov, "Evaluating an ensemble model for stroke image classification: Comparative analysis with individual neural network architectures," Bulletin of Abai KazNPU, vol. 88, no. 4, pp. 179-187, 2024.
[20] S. Kemenes, L. Chang, M. Schlereth et al., "Advancement and independent validation of a deep learning-based tool for automated scoring of nail psoriasis severity," Frontiers in Medicine, vol. 12, p. 1617441, 2025.
[21] J. Rasanjana, "Nail Disease Dataset," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/josephrasanjana/nail-disease-dataset