



## Ensemble Deep Learning Approach for Nail Disease Classification Using EfficientNet-V2 and ConvNeXt-V2

Nirupama Dora<sup>1</sup>, Sachikanta Dash<sup>2</sup>

<sup>1</sup>Department of CSE, GIET University, Gunupur, India

<sup>2</sup>Department of CSE, MITS University, Madanapalle, India

nirupamadora2000@gmail.com, dash.sachikanta@gmail.com

**Abstract:** A broad range of human nail disorders affects the health, texture, and appearance of toenails and fingernails. Various reasons such as injury, infection, systemic diseases, and nutrition deficiency can lead to these nail disorder conditions. Additionally, the nails might be impacted by certain skin dis-orders like psoriasis and eczema. Since nails are associated with good grooming habits and hygiene, the quality of nails plays an important role in creating an initial impression. Hence, providing a timely and accurate diagnosis provides satisfactory results for the patient. Many traditional diagnosis methods of nail disorders, such as nail scraping or clipping for lab experiments, dermatoscopy, and clinical observations, are very time-consuming, often subjective to bias, involve many steps, and require the assistance of a specialist. In recent years, technology including digital photography and computer vision have opened up numerous opportunities to increase efficiency by way of automation in this area through the use of digitization and Deep Learning algorithms, enabling more objective and timely assessments of nails in regard to abnormalities. Research in deep learning has shown positive performance in medical image classification by using Convolutional Neural Networks (CNNs). This study proposes an ensemble technique using EfficientNet-V2 and ConvNeXt-V2 as the CNN architectures for enhancing the performance of nail disease classification. ConvNeXt-V2 scored 98.51% on the test, compared to 95.12% for EfficientNet-V2. The suggested fusion ensemble performs noticeably better than individual backbone models, according to experimental findings. The feature-level fusion ensemble proposed in this study further enhanced the results, giving a test accuracy of 99.04% with a heightened degree of robustness. By studying the performance between these two state-of-the-art architectures, we expect to see improvements in both the performance and generalization of performance in actual clinical settings.

**Keywords:** Nail Disease Classification, EfficientNetV2, Con-vNeXtV2, Ensemble Learning, Feature Fusion, Deep Learning, Medical Image, Classification.

## 1. Introduction

Human nail illnesses are a broad category of conditions that can impact the health, texture, and appearance of fingernails and toenails. These disorders may arise from various factors including infections, injuries, systematic diseases, and nutritional deficiencies. In addition, certain skin conditions such as psoriasis and eczema can affect the nails. Since nails are linked to proper grooming and hygiene, their condition enhances initial impressions. Therefore, an early and timely diagnosis provides patients with suitable and accurate outcomes.

Traditional diagnostic methods, such as nail scraping or clipping for laboratory experiments, dermatoscopy, and clinical observations, are time-consuming, subjective, and require specialists. In recent years, technological developments in digital photography and computer vision have opened up exciting possibilities for automating the diagnostic procedure, enabling quicker and more impartial evaluations of nail anomalies. However, visually diagnosing these illnesses can be challenging and requires specialized knowledge.

Convolutional Neural Networks (CNNs) have illustrated encouraging outcomes in medical image classification as deep learning techniques have advanced. This study proposes an ensemble method to improve diagnostic accuracy for nail disease classification by combining EfficientNet-V2 with ConvNeXt-V2, two state-of-the-art convolutional neural network architectures, for high-accuracy nail disease classification. In real-world clinical circumstances, we aim to investigate whether combining models with different architectural elements can improve performance and generalization.

### 1.1. Research Objectives

The objectives of this research are as follows:

- To construct an ensemble scheme suitable for the auto-mated diagnosis of nail disorders, enhancing dermatological accuracy and efficiency of the assessment [16].
- To ensure the framework can be adapted for a real-time clinical environment or can be deployed on platforms such as mobiles, tablets, etc.
- To test and compare the proposed ensemble model with other CNN models such as EfficientNetV2, ResNet, VGG, and ConvNeXt-V2, using accuracy, precision, re-call, F1-score, and AUC as evaluation measures.

### 1.2. Contributions of the Paper

The key contributions of this work can be summarised as follows:

- First feature-level fusion of EfficientNetV2-RW-S and ConvNeXtV2 for six-class nail disease classification compared to previous decision-level ensemble approaches [2], [13].
- Both networks are fine-tuned in full using transfer learning rather than as frozen feature extractors, allowing both backbones to be adapted to the task at hand.
- Outperforms the individual EfficientNetV2 and ConvNeXtV2 models in terms of accuracy, AUC and F1-score.
- Achieves 99.04% testing accuracy, exceeding the individual EfficientNetV2 (95.12%) and ConvNeXtV2 (98.51%) baselines.

The paper is organized as follows. Section 2 Literature Survey. The section 3 contains the proposed methodology and Section 4 tabled the result and discussions. The section 5 discussed the results and section 6 delivers the conclusion and future direction.

## 2. Literature Survey

Numerous studies have approached nail disease classification from different architectural and methodological angles, ranging from single-backbone CNNs to hybrid and ensemble frameworks; this section reviews these efforts and situates the present work relative to them.

Regin et al. [2] took an ensemble route, training several CNN classifiers on Weka-preprocessed images and averaging their predictions; the resulting ensemble reached 95% accuracy, and the authors attributed this gain in stability and generalization to the combination of multiple learners rather than any single model.

A graph-based alternative was explored by Indriani (2025) [3] at Widyatama University, where a Graph Attention Net-work (GAT) modeled inter-class relationships through cosine similarity on top of ResNet-50 features. Trained on 3,835 Kaggle images spanning six classes, the model achieved 88% accuracy; the authors identified balanced datasets, augmentation, and more expressive GNN designs as directions for future improvement.

Working on a broader skin-disease dataset, Shams (2025) [15] benchmarked ResNet-50, MobileNet, and EfficientNet-B0 across six dermatological conditions. With data augmentation and transfer learning applied, EfficientNet-B0 outperformed the other two, reaching 99% accuracy despite the limited size of the training data — evidence that lightweight EfficientNet variants can generalize well even from sparse samples.

Shandilya et al. (2024) [5] targeted six nail disorders Blue Finger, Clubbing, Pitting, Onychogryphosis, Acral Lentiginous Melanoma, and healthy nails — using a hybrid capsule CNN architecture. Their results varied by model and task. VGG16 reached 92–96% accuracy on a three-class subset, ResNet101v2 achieved recall and precision above 87%, and a combined CNN-random forest classifier exceeded 96% accuracy on four dermatological categories; vision transformers pushed this further to 97.7% for psoriasis-type conditions. The authors note, however, that small and imbalanced datasets continue to limit how well these gains generalize.

Jalal et al. (2025) [6] focused on nail fold capillary abnormalities, applying EfficientNet-B0 with a cascade transfer-learning scheme to a small clinical dataset of 288 images. Cascading the training stages lifted performance from 67% accuracy (single-stage baseline) to a perfect 1.00 across accuracy, precision, recall, and ROC-AUC — a striking result that the authors themselves caution needs validation on larger datasets before it can be considered generalizable. For nail-specific classification, Sogukkucu et al. (2023) [7] fine-tuned VGGNet variants on 723 images spanning clubbing, Beau's lines, and melanonychia, reporting 92.7% (VGG16) and 92% (VGG19) test accuracy; a custom CNN they also trained fell short at 77% validation accuracy. The authors suggested extending this line of work with medical records, animal-nail datasets, and mobile-based CAD tools.

Nijhawan et al. [8] took a segmentation-first approach, splitting each nail image into upper, lower, and full-section regions before extracting and fusing CNN features from each; on a modest dataset of 300–400 images, a Random Forest classifier applied to these fused features achieved the best result at 84.58% accuracy, with the authors recommending the inclusion of patient metadata to further strengthen robustness.

Sharma et al. (2024) [9] explored fingernail imagery as a proxy for broader health screening, distinguishing Kidney Disorder, Anemia, Melanoma, and Healthy classes with a hybrid CNN-Random Forest model. The approach reached 97.2% accuracy with matching precision (97.02%) and F1-score (97.01%), and the authors pointed to dataset expansion and smartphone deployment as next steps for adaptability.

An earlier and more classical approach came from Indi and Patil (2019) [10], who relied on conventional image-processing techniques — extracting color and texture attributes from nail images and feeding them into simple machine-learning classifiers. Although this confirmed that nail disease could be detected from imagery without deep architectures, the method's dependence on hand-engineered features and sensitivity to lighting conditions limited its real-world robustness. Maniyan et al. [11] similarly combined handcrafted image features with an ensemble of K-Nearest Neighbors classifiers for early disease diagnosis; while classifier combination improved reliability somewhat,

the reliance on manual feature engineering left the framework vulnerable whenever nail diseases exhibited high intra-class visual similarity.

Rather than proposing a new model, Gaurav et al. (2025) [12] surveyed the broader landscape of AI applications in nail disorder diagnosis, cataloguing recurring obstacles across the field: a shortage of annotated data, class imbalance, the absence of standardized benchmarks, and inconsistent validation practices — all of which they argue must be addressed for AI-based nail diagnostics to mature into generalizable tools.

Closer to the ensemble strategy adopted here, Yamac, et al. (2023) [13] combined multiple classifiers through majority voting for nail disease detection, showing that this decision-level fusion improved both stability and accuracy relative to any single model. Because the fusion happened only after each model had already committed to a prediction, however, the approach could not exploit the complementary information present in the models' intermediate feature representations — a limitation the feature-level fusion proposed in this paper is designed to address.

Taken together, this body of work shows steady progress in AI-driven nail disease classification, with hybrid architectures, ensembling, and increasingly sophisticated backbones (including GNNs and transformers) each contributing incremental gains. What remains consistently under-addressed, however, is the combination of imbalanced or small datasets, narrow disease-class coverage, and — with few exceptions — validation that stops short of real clinical deployment.

### 2.1. Limitations of the Existing Systems

Some challenges may be observed after reviewing numerous research studies on nail disease detection and its applications in different fields:

- Some approaches, like hybrid capsule CNNs and transformer-based architectures, showed high accuracy but required more computational power, limiting practical deployment in low-resource environments.
- Most current deep learning approaches for nail disease focused on a limited subset of nail disorders, which restricts the applicability of these models for comprehensive nail disease screening.
- Several studies were conducted on relatively small and imbalanced datasets, often containing fewer images. Such limited datasets restrict the ability of models to generalize effectively across diverse populations and different clinical settings.

## 3. Proposed Methodology

This section outlines the advanced deep learning-based framework designed for classification of nail disorders into six different categories. The research process begins with dataset collection, preprocessing, model training, and ensemble pre-diction. Figure 1 illustrates the overall flow of the proposed methodology.

### 3.1. Dataset Collection

This research utilized a publicly available Kaggle dataset [14], consisting of 3,835 labeled nail images, partitioned into 3,066 training images and 769 testing images. The dataset covers six distinct classes: Acral Lentiginous Melanoma, Healthy Nail, Onychogryphosis, Blue Finger, Clubbing, and Pitting. Each image is associated with a ground-truth class label, as shown in Figure 2. The training image set was further split into training and validation subsets in a 90:10 ratio to ensure reliable performance evaluation.

### 3.2. Data Preprocessing

In this study, the input images from the nail disease dataset were preprocessed using a sequence of transformations to ensure uniformity and to improve the robustness of the model. The preprocessing pipeline is described below:

- 1) **Resize:** To maintain a fixed input size required by the EfficientNetV2 deep learning architecture, all images were resized to  $224 \times 224$  pixels.

- 2) **Random Horizontal Flip:** For each image, a random horizontal flipping transformation was applied with a probability of 0.5. This augmentation provides mirrored variations of the images, which helps in preventing overfitting.
- 3) **Random Rotation:** The images were randomly rotated within a range of  $\pm 10^\circ$  to introduce rotational invariance into the dataset. This rotation helps to enhance model generalization and reduce overfitting.
- 4) **Tensor Conversion:** The preprocessed images were converted from PIL/NumPy format into PyTorch tensors with pixel values normalized to the range [0, 1].
- 5) **Normalization:** Finally, all images were normalized using the ImageNet mean and standard deviation values for each RGB channel as in equation 1.

$$\mu = (0.485, 0.459, 0.406) \quad \sigma = (0.229, 0.224, 0.225) \quad (1)$$

This normalization ensures that the data distribution matches the pretrained ImageNet weights used for EfficientNetV2. The above preprocessing steps not only standardized the input data but also enhanced variability through augmentation, enabling better generalization of the model during training.

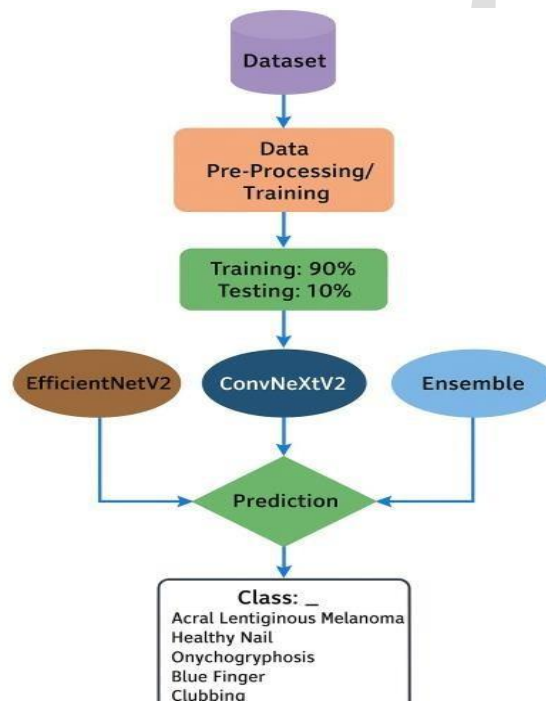


Fig. 1. Proposed Flow of the Architecture



Fig. 2. Dataset Samples across the Six Nail Disease Classes.

### 3.3. ConvNeXt-V2 Model

ConvNeXt-V2 is a convolutional neural network architecture combining efficient CNN design with fully convolutional masked auto encoder pre training and a global response normalization layer. The model was trained for 30 epochs with the Adam W optimizer ( $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ), an initial learning rate of  $1 \times 10^{-4}$ , and a learning rate decay factor of 0.05 per epoch. A batch size of 32 was maintained. The convolutional backbone employs depth wise separable convolutions as in equation 2.

$$Y_{i,j,c} = \sum_{u=-k/2}^{k/2} \sum_{v=-k/2}^{k/2} X_{i+u,j+v,c} \cdot K_{u,v,c} \quad (2)$$

Here  $X$  is the input feature map,  $K$  is the depthwise kernel, and  $Y$  is the output feature map. After depth wise convolution, a pointwise convolution integrates channel information using in equation 3 where  $W \in R^{c \times m}$  projects  $C$  channels to  $M$  output channels.

$$Z_{i,j,m} = \sum_{c=1}^C Y_{i,j,c} \cdot W_{c,m} \quad (3)$$

### 3.4. EfficientNetV2-RW-S Model

EfficientNetV2-RW-S, pre-trained on ImageNet, was fully fine-tuned for nail disease classification. The model leverages compound scaling for depth, width and resolution, as well as faster training convergence. Training was performed for 30 epochs using the AdamW optimizer with One Cycle LR scheduler, which decreased the learning rate every 5 epochs to ensure smooth convergence. Augmentation strategies, including random cropping and color jittering, were applied. A batch size of 32 was used throughout.

#### 3.4.1 Evaluation

The trained weights were loaded, and the model's performance was evaluated on the test image set using accuracy as the primary metric shown in equation 4.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \% \quad (4)$$

Where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. In our implementation, the model achieved a test accuracy of 95.12%, demonstrating the efficiency of EfficientNetV2 in nail disease classification tasks

### 3.5. Ensemble Model: EfficientNetV2 + ConvNeXtV2

To leverage the strengths of different architectures, an ensemble model was constructed by combining feature representations from EfficientNetV2 and ConvNeXtV2. Both networks were pre trained on ImageNet and fine-tuned on the nail disease dataset. The classification heads of both models were removed, and their feature maps were concatenated before passing through a joint classifier. Ensemble methods are particularly effective in medical image analysis as they reduce model bias and variance while improving generalization. In this work, both prediction-level ensemble and feature-level fusion ensemble strategies were investigated.

#### 3.5.1. Mathematical Formulation

Let  $f_{Eff}(X)$  and  $f_{Conv}(X)$  denote the feature representations extracted from EfficientNetV2 and ConvNeXtV2 respectively, given an input image  $x$ . Each network produces a high-dimensional feature vector as shown in equation 5.

$$f_{Eff}(X) \in R^{d_1}, \quad f_{Conv}(X) \in R^{d_2} \quad (5)$$

The fusion step concatenates these features into a joint embedding using formula in equation 6.

$$f_{Fusion}(X) = [f_{Eff}(X) \parallel f_{Conv}(X)] \in R^{d_1+d_2} \quad (6)$$

Where  $\parallel$  denotes vector concatenation. This fused feature vector is then passed through a fully connected classification network using Softmax() in equation 7.

$$\hat{Y} = \text{Softmax}(W_2 \sigma(W_1 f_{Eff}(X) + b_1) b_2) \quad (7)$$

Here,  $W_1 \in R^{512 \times (d_1+d_2)}$  and  $W_2 \in R^{C \times 512}$  are trainable weight matrices,  $b_1$  and  $b_2$  are bias terms,  $\sigma(\cdot)$  is the ReLU activation function, and  $C = 6$  is the number of output classes.

### 3.5.2. Training Strategy

In the first ensemble approach, a soft voting mechanism was implemented to aggregate the outputs of EfficientNetV2-RW-S and ConvNeXtV2-Base models. For a given input image, each model outputs an array of logits for different classes. These logits are averaged to obtain the final ensemble prediction using equation 7.

$$\hat{Y} = \frac{1}{2} (f_{Eff}(X) + f_{Conv}(X)) \quad (8)$$

Where  $f_{Eff}(X)$  and  $f_{Conv}(X)$  denote the output logits of EfficientNetV2 and ConvNeXtV2, respectively.

### 3.5.3. Evaluation

The trained ensemble was evaluated using classification metrics such as Accuracy, Precision, Recall, F1-score, and AUC. The ensemble model consistently out-performed the individual EfficientNetV2 and ConvNeXtV2 models, demonstrating that feature-level fusion enhances the discriminative power for nail disease classification.

Advantages of the Proposed Framework:

The proposed deep learning framework offers several advantages:

- Integration of complementary representations from EfficientNetV2 and ConvNeXtV2.
- Improved robustness through ensemble learning.
- Reduced overfitting via feature fusion and backbone freezing.
- Enhanced discrimination of subtle visual differences among nail disease classes.

The final predicted output class is obtained by applying the softmax function followed by an argmax operation. This soft voting strategy harnessed the complementary decision boundaries learned by the two architectures, resulting in enhanced classification accuracy compared to individual models.

To unlock deeper complementary representations, a feature-level fusion ensemble was proposed. Instead of combining final predictions, high-level feature representations extracted from EfficientNetV2 and ConvNeXtV2 were fused to form a unified feature vector.

Given an input image  $x$ , feature maps were obtained using the backbone feature extraction layers of both networks. A global average pooling layer was applied to each feature map to obtain a set of compact feature vectors using formula in equation 9.

$$f_e = \text{GAP}(\phi_{\text{Eff}}(x)), \quad f_c = \text{GAP}(\phi_{\text{Conv}}(x)) \quad (9)$$

Where  $\phi_{\text{Eff}}$  and  $\phi_{\text{Conv}}$  represent the feature extraction functions of EfficientNetV2 and ConvNeXtV2, respectively.

The pooled feature vectors were concatenated to form a fused representation as shown in equation 10.

$$f_{\text{fusion}} = [f_e \parallel f_c] \quad (10)$$

This merged feature vector is able to capture both fine-grained spatial patterns and high-level semantic information,

enabling more discriminative learning for complex nail disorder patterns.

The fused feature vector was then fed to a fully connected classification head consisting of a dense layer with ReLU activation, followed by dropout regularization to mitigate overfitting. The final layer maps the fused representation to six output classes.

During training, both EfficientNetV2-RW-S and Con-vNeXtV2 were fully fine-tuned together with the fusion classifier. Fine-tuning all network parameters enabled the pre-trained models to adapt their feature representations to the nail disease dataset, resulting in improved feature extraction and better discrimination among visually similar nail disease classes. The categorical cross-entropy loss function in equation 11 was used to optimize the complete network:

$$L = \sum_{i=1}^c Y_i \log(\hat{Y}_i) \quad (11)$$

Where  $Y_i$  and  $\hat{Y}_i$  denote the ground-truth label and predicted probability for class  $i$  respectively. Optimization was performed using the Adam optimizer with an initial learning rate of  $1 \times 10^{-3}$  for the fusion classifier.

## 4. Implementations and Results

Our proposed model, based on deep learning and specifically Convolutional Neural Networks (CNNs), was implemented using a Python environment. We employed torch vision for preprocessing, data augmentation, and access to pre-trained CNN architectures, while the optim library provided access to the optimization algorithms used to train the neural networks. Experiments were conducted on a Kaggle dataset of nail diseases, consisting of 3,066 training images and 769 testing images across six nail disease classes. Training and testing of image sets were performed in a GPU-enabled environment to accelerate computation. The results highlight the model's ability to accurately identify and classify nail disorders.

### 4.1. Result Analysis

The comparative performance analysis of EfficientNetV2, ConvNeXtV2, and their Fusion Ensemble is presented in Table I. Several evaluation measures were considered, including training accuracy, validation accuracy, testing accuracy, F1-score, precision, recall, and the area under the ROC curve (AUC).

On the held-out test, EfficientNetV2 correctly classified 95.12% of samples, following a training run that reached 96.80% accuracy with 95.67% on validation. Precision, recall and F1-score all achieved 95%, and the model returned an AUC of 97.50% indicate a solid result overall, though it still confused visually similar classes such as Blue Finger and Clubbing more often than desired. ConvNextV2, by contrast, pushed testing accuracy to 98.51% (against 99.12% training and 98.73% validation accuracy), with precision, recall and f1-score averaging 98.5% and an AUC of 99.20%, reflecting a stronger and more consistent fit across all three data splits.

Compared with EfficientNetV2, the proposed fusion ensemble improved testing accuracy from 95.12% to 99.04%, an increase of 3.92 percentage points, along with a corresponding AUC improvement from 97.50% to 99.60% (2.10 points) and an F1-score improvement from 95.00% to 99.00% (4.00 points). Compared with ConvNeXtV2, testing accuracy improved by 0.53 percentage points (98.51% to 99.04%), AUC improved by 0.40 points (99.20% to 99.60%), and F1-score improved by 0.50 points (98.50% to 99.00%), demonstrating the added benefit of feature-level fusion even over the stronger individual backbone. The fusion ensemble model exhibited resilience and strong generalization across all nail disease classes by integrating complementary features from both networks.

The PR curve shows precision above 0.8 for most recall levels, indicating reliable detection confidence.

**Table I. Performance Metrics of Models on Nail Disease Dataset**

| Model              | Train<br>Acc (%) | Val<br>Acc (%) | Test<br>Acc (%) | F1<br>(%) | Prec.<br>(%) | Recall<br>(%) | AUC<br>(%) |
|--------------------|------------------|----------------|-----------------|-----------|--------------|---------------|------------|
| EfficientNetV2     | 96.80            | 95.67          | 95.12           | 95.00     | 95.00        | 95.00         | 97.50      |
| ConvNeXtV2         | 99.12            | 98.73          | 98.51           | 98.50     | 98.60        | 98.50         | 99.20      |
| Fusion<br>Ensemble | 99.76            | 99.30          | 99.04           | 99.00     | 99.10        | 99.00         | 99.60      |

#### 4.2. Evaluation Metrics

To extensively assess the performance of EfficientNetV2, ConvNeXtV2, and the proposed Fusion Ensemble model, several widely used evaluation metrics were employed. These metrics provide comprehension into both classification accuracy and the model's capability to generalize across multiple nail disease classes.

- 1) Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC): The ROC curve illustrates the trade-off between the true positive rate (TPR) and false positive rate (FPR) at different threshold values. The formula are given in equation 12 and 13 respectively.

$$TPR = \frac{TP}{TP+FN} \times 100 \% \quad (12)$$

$$FPR = \frac{FP}{FP+TN} \times 100 \% \quad (13)$$

The Area Under the Curve (AUC) quantifies the overall classification capability of the model. An AUC value closer to 1 indicates excellent discriminative performance.

- 2) Accuracy: Accuracy measures the overall correctness of the model by computing the ratio of correctly classified observations to the total number of observations as mention in equation number 4.
- 3) Macro-Averaging for Multi-Class Classification: Since nail disease classification is a multi-class problem, macro-averaged precision, recall, and F1-score were used. These metrics are computed using formula in equation 14 by averaging the scores of all individual classes:

$$Macro\ Metric = \frac{1}{N} - \sum_{i=1}^N M_i \quad (14)$$

- 4) Precision: Precision in equation 15 indicates the reliability of the positive predictions made by the model. It answers the question, "Of all the items the model labeled as positive, how many were actually positive?" It reflects how many predicted positive instances are actually correct.

$$Precision = \frac{TP}{TP+FP} \times 100 \% \quad (15)$$

A higher precision value indicates fewer false positive predictions, which is crucial in medical diagnosis applications.

5) Recall: Recall, also known as sensitivity, measures the model's ability to find all the positive samples as shown in equation 16. It answers the question, "Of all the actual positives, how many did the model correctly identify?"

High recall ensures that fewer disease cases are missed during classification.

$$Recall = \frac{TP}{TP+FN} \times 100 \% \quad (16)$$

6) F1-Score: The F1-score in equation 17 balances precision and recall into a single value by providing a harmonic mean of precision and recall, offering a balanced evaluation when class distributions are uneven or the dataset is imbalanced:

$$F_1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100 \% \quad (17)$$

#### 4.3. Comparative Discussion

The performance of the fusion ensemble over the individual models has been represented in Table I and Figure 3. Despite its efficiency, EfficientNetV2 had trouble recognizing rare or visually similar/overlapping classes, whereas there was a notable decrease in the number of incorrect predictions in ConvNeXtV2. Nevertheless, the Fusion Ensemble successfully combined the advantages of both models, achieving clinical grade accuracy exceeding 99%. Class-specific variations emphasize the benefit of the ensemble approach even more. For example, a recall of 0.99 was obtained for Blue Finger when ConvNeXtV2 as well as the ensemble model was used, while EfficientNetV2 tended to mistakenly classify it as Clubbing. Correspondingly, a higher precision for Clubbing was obtained under the proposed ensemble model as compared to individual models.

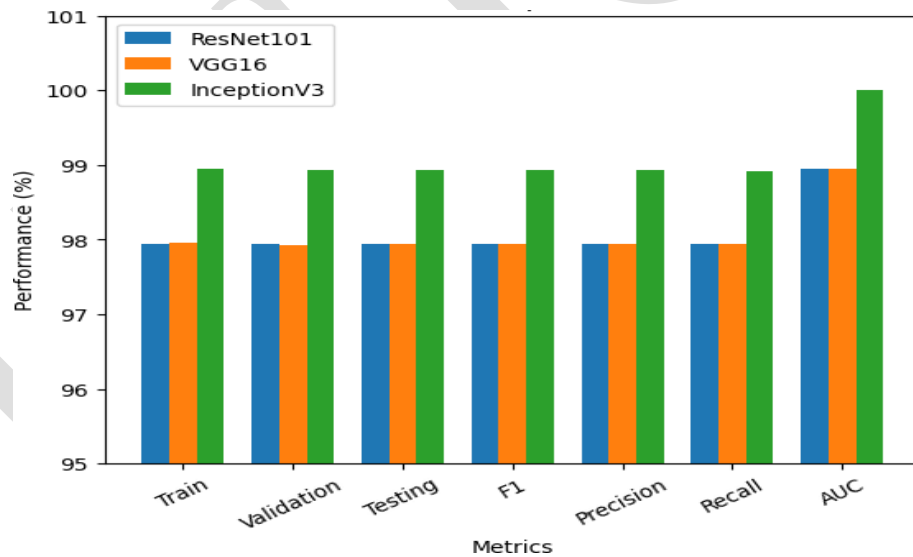


**Fig. 3. Comparison of Model Accuracies across Training, Validation, and Testing Phases**

Figure 3 displays the accuracy trends for the three models during training and testing, while Figure 4 presents a comparative visualization of performance metrics obtained across all these models. The AUC values showed that the ensemble approach had greater discriminative power and more steady convergence.

Several prior studies have reported strong results using single backbone architectures. For instance, Sogukkucu et al. [7] obtained testing accuracies of 92.7% (VGG16) and 92% (VGG19) on three classes of nail disease, while Nijhawan et al. [8] obtained a maximum of 84.58% overall testing accuracy by employing the fusion of CNN features with a Random Forest classifier. In contrast, the present work achieves higher testing accuracies across the classes. More

recent works utilizing EfficientNet-based architectures have shown promising improvements. Shams et al. [15] demonstrated an accuracy rate of 99% for skin disease classification using EfficientNet-B0 with extensive data augmentation. Similarly, Jalal et al. [6] achieved optimal performance metrics using EfficientNet-B0 with cascade transfer learning, but on a small dataset of only 288 images, which raises concerns regarding generalizability. In comparison, the proposed fusion model maintains clinical-grade performance exceeding 99% on a larger and more diverse nail disease dataset, indicating stronger robustness. Ensemble learning has been explored in previous works with the goal of improving the stability of classification results; Regin et al. [2] showed that accuracy can be increased to 95% using an ensemble of CNNs, emphasizing the advantage of combining multiple learners. Building upon this idea, the proposed research further advances ensemble learning with an emphasis on feature-level fusion instead of decision-level voting, facilitating the use of complementary features derived from EfficientNetV2 and ConvNeXtV2 models, which in turn helps reduce the misclassification rate of visually similar nail diseases like Blue Finger and Clubbing.

5. DISCUSSION

This study integrates the EfficientNet-V2 and ConvNeXt-V2 architectures into an ensemble-based deep learning framework for the diagnostic analysis of nail images. The dataset used in the experiments consisted of 3,835 images representing different aspects of nail disease classes. Of the total, 3,066 images were utilized for training, while 769 images were reserved for testing. This data distribution improved the model’s capacity to generalize over a range of nail conditions, ensuring that the model was exposed to substantial variety and quality during training.

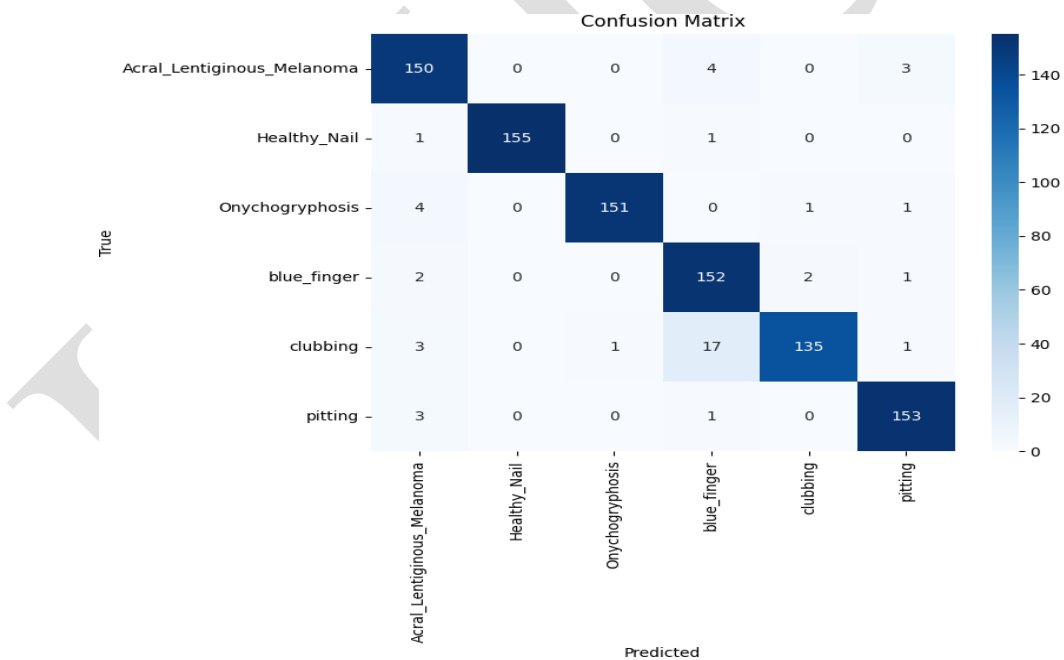


Fig. 4. Confusion Matrix of the Efficientnetv2 Model.

Insightful tendencies were found when comparing individual models with the suggested ensemble, as shown in Figures 4, 5, and 6. ConvNeXt-V2 demonstrated better feature extraction because of its updated convolutional design, while EfficientNet-V2 performed admirably with robust generalization capabilities. However, the ensemble framework outperformed both models alone by leveraging the complementary strengths of the individual networks, improving classification accuracy and robustness.

The ensemble model achieved an average of 99.76% for training, 99.30% for validation, and 99.04% for testing.

Additionally, the system successfully reduced false positives and false negatives, as evidenced by precision, recall, and F1-score all consistently exceeding 98%. Significantly, the ensemble model demonstrated exceptional class separation performance, with an AUC score of 99.60%.

These findings demonstrate the significance of not only improving the accuracy of predictions through ensembling machine learning models – including convolutional and recurrent neural networks – but also improving their reliability when operating under practical conditions found in clinical practice.

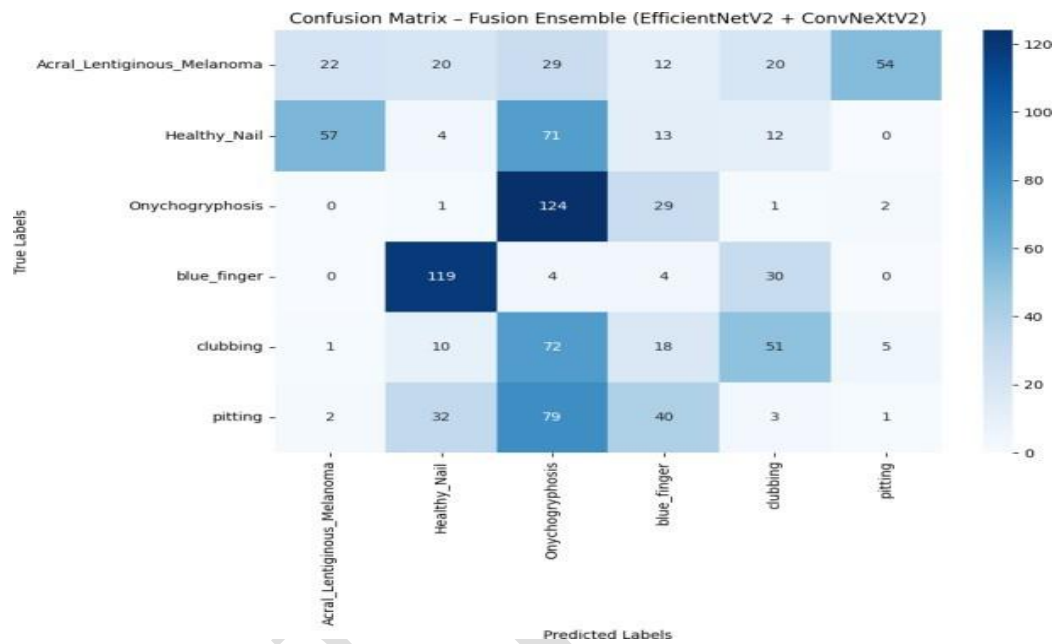


Fig. 5. Confusion Matrix of the Convnextv2 Model.

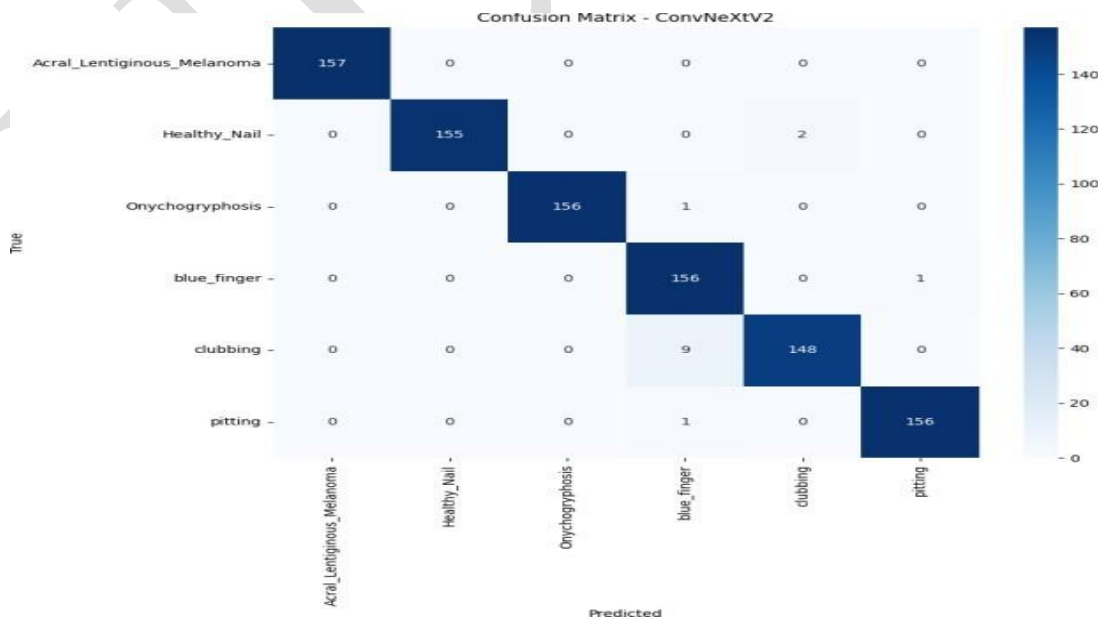


Fig. 6. Confusion Matrix of the Efficientnetv2–Convnextv2 Fusion Ensemble.

## 6. Conclusions and Future Directions

For automated nail disease identification, this paper proposes an ensemble learning framework that combines features of the EfficientNet-V2 and ConvNeXt-V2 models. The proposed model is capable of efficiently diagnosing nail abnormalities using medical image databases, outperforming the conventional diagnostic process, which is time-consuming, highly prone to error, and subjective. From the experimental results, the ensemble model performs better than individual models in terms of accuracy, F1-score, and area under the curve.

The results confirm that ensemble-based deep learning techniques can function as effective and trustworthy diagnostic instruments for dermatological applications. Specifically, it is observed that the model has a great ability to identify conditions with a clear visual difference, such as pitting, clubbing and blue finger etc.

Future work can focus on increasing the size of the dataset through data augmentation and collaboration with multiple clinical centers. Additionally, incorporation of explainable AI (XAI) methodology would provide better explainability by pointing to important visual cues that impact predictions. Deployment of a lightweight ensemble architecture on mobile devices would provide enormous potential for real-time applications in point-of-care diagnostics. Although the proposed model achieved excellent performance on the benchmark dataset, evaluation using k-fold cross-validation, external clinical datasets, and statistical significance analysis (e.g., McNemar's test) will be investigated in future work.

Overall, this study verifies that deep learning-based ensemble techniques can greatly help enhance nail disease diagnosis automation and can potentially open up a range of applications in telemedicine and dermatology.

## References

1. P. S. Roy, V. Kukreja, S. N. Chandran, and A. Choudhary, "Empirical analysis of nail diseases using hybrid cnn-lstm algorithms," IEEE IC2PCT Proceedings, pp. 54–59, 2024.
2. R. Regin, G. Reddy, S. K. Ch, and J. Cvn, "Nail disease detection and classification using deep learning," Central Asian Journal of Medical and Natural Science, vol. 3, no. 3, pp. 574–594, 2022.
3. I. Indriani, "Nail disease classification using graph attention network (gat) and resnet," Ranah Research: Journal of Multidisciplinary Re-search and Development, vol. 7, no. 4, pp. 2522–2529, 2025.
4. M. Alruwaili and M. Mohamed, "An integrated deep learning model with efficientnet and resnet for accurate multi-class skin disease classification," Diagnostics, vol. 15, no. 5, p. 551, 2025.
5. G. Shandilya, S. Gupta, S. Bharany, A. U. Rehman, U. Kaur, H. M. Som, and S. Hussien, "Autonomous detection of nail disorders using a hybrid capsule cnn," BMC Medical Informatics and Decision Making, vol. 24, no. 1, p. 414, 2024.
6. M. E. Jalal, F. Torres, and C. Alvarez, "Detection of nailfold capillary abnormalities using efficientnet-b0 with cascade transfer learning," Computers in Biology and Medicine, vol. 173, p. 108921, 2025.
7. M. Sogukkuyu, A. Demir, and B. Kaya, "Transfer learning-based classification of nail diseases using vggnet architectures," Biomedical Signal Processing and Control, vol. 88, p. 105493, 2023.

8. R. Nijhawan, R. Verma, Ayushi, S. Bhushan, R. Dua, and A. Mittal, "An integrated deep learning framework approach for nail disease identification," in 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), pp. 197–202, 2017.
9. P. Chaudhary, A. Verma, V. Kukreja, and R. Sharma, "Integrating deep learning and ensemble methods for robust tomato disease detection: A hybrid cnn-rf model analysis," in 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 1–4, 2024.
10. T. S. Indi and D. D. Patil, "Nail feature analysis and classification techniques for disease detection," International Journal of Computer Science and Engineering, vol. 7, no. 5, pp. 1376–1383, 2019.
11. P. Maniyan and B. L. Shivakumar, "Early disease detection through nail image processing based on ensemble of knn classifier and image features," IOSR Journal of Computer Engineering (IOSR-JCE), vol. 20, no. 3, 2018.
12. V. Gaurav, C. Grover, M. Tyagi, and S. Saurabh, "Artificial intelligence in diagnosis and management of nail disorders: A narrative review," Indian Dermatology Online Journal, vol. 16, no. 1, pp. 40–49, 2025.
13. S. A. Yamac, O. Kuyucuog˘lu, S. . B. Ko˘seog˘lu, and S. Ulukaya, "Detec-tion of nail diseases using ensemble model based on majority voting," Kahramanmaras, Su˘tcu˘ I˘mam U˘niversitesi Mu˘hendislik Bilimleri Dergisi, vol. 26, no. 1, pp. 250–260, 2023.
14. N. Gurav, "Nail disease detection dataset." <https://www.kaggle.com/datasets/nikhilgurav21/nail-disease-detection-dataset>, 2024. Accessed: Dec. 27, 2025.
15. M. Y. Shams, E. Hassan, S. Gamil, A. Ibrahim, E. Gabr, S. Gamal, E. Ibrahim, F. Abbas, A. Mohammed, A. Khamis, M. Hamed, M. Mokhtar, and R. Bhatnagar, "Skin disease classification: A comparison of resnet50, mobilenet, and efficient-b0," Journal of Current Multidis-ciplinary Research, 2025.
16. N. Dora, S. Dash, A. Baboo, and S. Ranjan Mishra, "Efficient Nail Disease Diagnosis Using Deep Neural Networks for Predicting Abnor-malities," in 2025 International Conference on Next Generation of Green Information and Emerging Technologies (GIET), Gunupur, India, 2025, pp. 1–5, doi: 10.1109/GIET65294.2025.11234860.
17. S. Dash, S. K. Behera, P. K. Sahu, and K. Panda, "Deep Neu-ral Networks for Accurate Detection of Nail Disorders," in 2025 International Conference on Information, Implementation, and Inno-vation in Technology (I2ITCON), Pune, India, 2025, pp. 1–6, doi: 10.1109/I2ITCON65200.2025.11210454.
18. A. Baboo, S. R. Mishra, and S. Dash, "An improvised diabetes prediction system using hybrid ensemble approach," in 2024 IEEE 11th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Lucknow, India, 2024, pp. 1–6, doi: 10.1109/UPCON62832.2024.10983076.