Al-Rafidain Journal of Computer Sciences and Mathematics (RJCSM)

www.csmj.uomosul.edu.iq

AI-Driven Non-Invasive Diagnosis of Diverse Medical Conditions Through Nail Image Analysis with High-Performance Ensemble Classifier

Abeer Mohamad Alshiha[1] and Wai Lok Woo[2]

[1]Remote Sensing Center, University of Mosul, Iraq
[2]Computer and Information Sciences Department, Northumbria University, United Kingdom
Email: abeer.allaf@uomosul.edu.iq[1] and wailok.woo@northumbria.ac.uk[2]

Article information

Article history:
Received 05 May ,2025
Revised 15 June ,2025
Accepted 22 June ,2025
Published 26 June ,2025

Keywords:
Human Diseases,
Nail Image,
Nail Color,
Transfer Learning,
Machine Learning.

Correspondence:
Abeer Mohamad Alshiha
Email:
abeer.allaf@uomosul.edu.iq

Abstract

In this research, the application of deep learning methods for the classification of human nail diseases using image analysis is investigated. The aim was to establish a non-invasive, automatic diagnosis tool for different nail conditions, utilizing deep convolutional neural networks (CNNs) for feature extraction. A total of 500 images of nails, divided into seven classes of diseases, were employed for training and testing. Feature extraction was performed using VGG16, ResNet50, and EfficientNetB0, and three machine learning classifiers, AdaBoost, LightGBM, and a Meta Classifier, were applied. The multi-classifier data classifier, the Meta Classifier, did better with 98.0% accuracy, 98.2% precision, 97.9% recall, and 98.0% F1 score when used in conjunction with EfficientNetB0. The study validates the efficacy of AI image diagnostics in non-invasive disease diagnosis, delivering a cost-effective and trustworthy method for early diagnosis, particularly in low-resource areas. The study verifies the accuracy of deep learning models, especially EfficientNetB0, for medical image examination, but extensive clinical validation and dataset acquisition are essential.

DOI: 10.33899/csmj.2025.160308.1189, ©Authors, 2025, College of Computer Science and Mathematics, University of Mosul, Iraq.

1. Introduction

Computer image processing techniques and algorithms have received a revolutionary usage in numerous various fields, particularly in the medical field, where they have exercised a definitive influence over the diagnosis of diseases as well as their treatment. These technologies are applied in the detection and diagnosis of numerous health diseases, with higher capability to detect diseases correctly being provided to physicians. It has been shown by research that medical imaging is extremely critical in disease diagnosis, offering perceptive information that aids clinicians in monitoring patients' health status and choosing appropriate treatments [1][2]. A disease is typically defined as any disturbance of a body part's structure or function, resulting in physiological changes that can largely be detected with the use of medical imaging modalities. These imaging procedures provide essential information to doctors, often before they prescribe treatment, by revealing subtle signs of illness that do not necessarily reveal themselves well through clinical inspection.

But while medical imaging is invaluable in diagnosis, on its own human eyesight is not without flaws and therefore can create inaccuracies in the diagnosis of specific conditions. For instance, doctors can miss subtle manifestations of diseases, especially at their early stages, and therefore lead to delayed diagnosis and treatment. Therefore, the evolution of modern diagnostic technologies, particularly digital image processing, became the necessity in overcoming these limitations. These technologies have proved more precise in disease detection, delivering a more reliable instrument of diagnosis and making early intervention possible in the health status of the patients [3].

Medical diagnostic science has advanced considerably, and various means of guiding medical practitioners in the diagnosis of a wide spectrum of diseases now abound. One of them is the science of diseases relating to nails, especially those manifested in the nail. Nails, primarily employed for the purpose of safeguarding fingers, also reflect on general health. Changes in the color, texture, and shape of nails can indicate underpinning diseases. Nail symptoms can be the initial presentation of systemic disease in organs beyond the skin, including chronic illnesses like liver disease, cancer, diabetes, and cardiovascular disease [4]. Nails have hence become an integral part of diagnostic procedures. Based on the examination of nail characteristics such as shape, color, texture, and the integrity of the nail plate, physicians can diagnose signs of diseases that are otherwise hidden, particularly if the disease lacks a clear clinical expression.

Present studies focus on the importance of generating comprehensive datasets based exclusively on images of human nail diseases. Though there are many image-based medical diagnosis systems, few address the specific issues of disease diagnosis based on nail images, and this is one gap this research aims to fill. Previous studies have highlighted the importance of having nicely annotated large datasets that are diverse to train machine learning methods for the classification of nail diseases [5]. The current study seeks to bridge this literature gap by providing a general overview of the most common diseases that can be diagnosed through nail image analysis.

The most important aspect of this research is the creation of a core dataset using over 500 images acquired locally. These images are classified under seven classes that have been identified as different diseases or conditions of health that can be observed from the characteristics of nails. They are Healthy natural nails (disease-free), Psoriasis, Paronychia, Systemic diseases, Contact eczema, Malnutrition, Fungal infections.

These categories encompass a wide variety of nail conditions, with variation between nail features in shape, color, texture, and appearance of the nail plate. The dataset not only captures nail diseases but also captures systemic diseases affecting general health, making it an essential tool for training diagnostic models.

There are some fundamental objectives satisfied by the establishment of this dataset. To begin with, it fills the huge loophole concerning publicly available data that are specifically designed to diagnose nail disease hence it offers a priceless avenue to researchers aiming to structure machine learning classifiers to identify nail disease. Second, the data include a wide variety of diseases, and this factor gives a chance to researchers to investigate the relationship between some of the peculiarities of the nail and systematic illness. These diseases can be clustered by their appearance attributes; therefore, the dataset can be used to develop machine learning models that can help effectively diagnose different diseases using only nail pictures.

Being diverse and high quality is also a consideration made fundamental whilst the dataset was being created. The pictures were gathered in various groups of people such that diversity in the various groups could be offered with various intensities of each condition. This variety gives the data set strength and is not subject to bias in a way that may cause overfitting among the machine learning algorithms. That there exists a well-curated dataset with high-quality images means that the models trained on it can generalize across populations, and hence can be applied in practice in real-life clinical practice.

The value of the dataset does not end with the size of the dataset, but rather, it is an organized piece of research to be used by both researchers and doctors. By showing the photos of various ailments and disorders, it creates the opportunity to conduct research in the future regarding the prospect of developing the diagnostic ability further with the help of machine learning and AI. Specifically, an analysis of nail-image could enable massive benefits in the early detection and disease tracking, as an equivalent but cheaper and more readily available method of diagnosis, especially among poorer communities where medical equipment is limited [6].

Ongoing development and refinement of these datasets are crucial to propel diagnostic techniques in medicine forward. By using AI-driven image analysis, physicians can potentially streamline the process of diagnosis, enhance early detection, and monitor disease progression more effectively. The present study paves the way for the integration of nail image analysis into routine clinical practice as a non-invasive, cost-effective, and time-saving method for the detection of many diseases

2. NAIL ANATOMY AND DISEASE INDICATORS

Nails are accessory structures of the skin that are produced by epidermal cells overlying the fingertips. Healthy, normal nails tend to be translucent pink due to the underlying blood vessels, smooth and firm to the touch, shiny, and curved in shape. Uniform color, free from discoloration, lines, or pitting, is one of the most significant indications of healthy nails. While the nail is growing, it appears white as it is its natural growth phase. Approximately 10% of skin diseases are nail diseases, and nail diseases are also significant indicators of systemic health issues. The signs of disease can appear in various parts of the nail, including the nail plate, the nail fold, the lunula (nail crescent), and the nail matrix. The lunula is particularly important because it is the visible portion of the stroma, the tissue beneath the nail. While not every alteration in the appearance of the lunula is pathological, its absence or alteration can indicate some medical problems. For example, the absence of the lunula can be a sign of malnourishment, vitamin B12 deficiency, vitiligo, anemia (due to iron deficiency or blood loss), or renal failure [7]. In some cases, chronic diseases such as cirrhosis or heart and lung disease may cause color change or visibility of the lunula, redness or pallor of the crescents being common presentations. In addition, some toxic exposures, as in silver poisoning (argyria), can cause bluish discoloration of the lunula [8].

Change in color, shape, and texture of the nail plate is also an important diagnostic sign. For instance, white spots on the nail plate, or leukonychia, are a manifestation of systemic diseases such as lung, liver, or heart disease [8]. A white area of the nail can also be a sign of anemia, as shown by fluorescence. Similarly, rough, bumpy nail surfaces, or white lesions, often occur in psoriasis, while rheumatoid arthritis is characterized by nails that are thickened with white scales underneath. Dark brown or black lines around the nail plate may suggest melanoma or other forms of skin cancer. "Lover's lines," which are horizontal grooves in the nail plate, may be linked to nutritional deficiency, particularly deficiency in protein or iron. Flat nails can indicate hypothyroidism, vitamin B12 deficiency, anemia, or diabetes, while yellow nails indicate respiratory problems, including pulmonary disease like asthma . Furthermore, a physical change in nail shape and length, defined as Terry's nails, with approximately 80% of the nail plate being white and the tip red, is frequently noted in individuals with chronic liver disease [9].

In general, nails can provide valuable diagnostic clues for everything from nutritional deficiencies to systemic illnesses of the heart, liver, or lungs. By examining the shape, color, and texture of nails, clinicians can glean a lot about a patient's overall health, diagnosing many illnesses early in the process. That being said, it needs to be noted that while nail changes may point to health issues, they are never absolute and must be aligned with other clinical signs and investigations.

3. RELATED WORK

The following is a review of the various research works on diagnosing human diseases based on images of human nails to detect diseases:

Indi et al. [10] proposed a system for disease diagnosis at an early stage based on the color changes in human nails. They utilized the Weka tool for data preparation for the initial training phases, where the nail images of the patients were compared to an original image for forecasting probable diseases. The system, which has been termed the Early-Stage Disease Diagnosis (ESDD) system, employed a nail color detection and disease prediction algorithm using normal or mobile cameras. The ESDD system achieved a 65% match with the training data, whereas the discovery process was relatively slow.

Gandhat et al. . [11] suggested a Disease Detection System (DDS) to conserve time for doctors in the diagnosis of diseases from nail images. The system improves the precision of diagnosis with close-up images of nails processed by the MATLAB program for easier medical reading. This aids in improving the precision of disease detection from the examination of the color of the nail.

Kumuda et al. [12] proposed a procedure of subdividing of nail images under the microscope that can present different images of the areola of the nail clearly, thereby maintaining the quality of the image. The method is associated with enhanced visualization of the structures of the nail, including the free edge and longitudinal skip which are of extreme importance in making an accurate diagnosis.

Saranya and colleagues [13] suggest certain methods of image processing in the automatic segmentation within nail region and extraction of abnormalities. They introduced the use of segmentation by use of intermediate filters and transformation to gray-level of the image to be more effective in processing. They extracted abnormal area features by the use of threshold water shades and the abnormal area features were used in the process of diagnosis of the disease.

Hadiyoso et al. [14] recommended a deep learning solution to classification and detection of diseases based on nail pictures. They called their method DEEPLEAING and employed the ensemble to extract features of a freshly created dataset based on the Convolutional Neural Network (CNN) models. The system indicated a 84 percent accuracy rate on comparing the performance of Support Vector Machine (SVM), Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), and Random Forest (RF) algorithms.

Yani et al. [15] has suggested an early diagnosis of the Terry nails using a CNN algorithm standard transfer learning model and attained an accuracy of 94.24 percent. The procedure had good prospects in identifying particular problems in the nail that may be signs of internal illness.

Ariansyah et al. [16] proposed the concept of the deep learning-based model of health condition identification by the nail image. They have used a CNN-based method, and they have gathered 400 images of five kinds of nails (psoriasis, rheumatism, anemia, melanoma, and normal). Their model performed with 78 percent of average detection, which proved that their model is efficient in the diagnosis of different health conditions by analyzing the nail..

Abdulhadi et al. [1] proposed a transfer learning framework for the diagnosis of four diseases of nails: pigmentation, clubbing, fungus, and nail health. They applied CNN algorithms on a large dataset and utilized five different training models (VGG, AlexNet, ResNet, DenseNet, and GoogLeNet). The models were compared using six performance measures, and the results concluded that DenseNet and ResNet performed best with an accuracy rate of 96%. The dataset was passed through Softmax for faster display of results, and the models' accuracy was calculated using MATLAB, with 60% of the data being used for training and 40% for testing.

To deal with the dataset, we have used SOFTMAX to show the results faster through several Parameter. Also, accuracy metric

- TP/(TP+FP) is adopted where 60% of the data was used for training and 40% of the data for testing.
- MATLAB was used to calculate the accuracy of each of the five models of the CNN algorithm and it was found that DESNET and RESNET are the best with a percentage of 96% [1].

Table (1) below includes a summary of previous studies and success rates for a number of researchers:

Table 1. Related works Summary

Author(s)	Year	Methodology	Key Findings	Accuracy/Performance
Indi et al.	2011	Early-stage Disease Diagnosis system (ESDD) using nail color analysis with Weka tool	Used nail images to predict diseases based on color differences, with mobile or regular camera	65% match with training dataset, discovery slow
Gandhat et al.	2016	Disease Detection System (DDS) for analyzing nail color using MATLAB	Increased diagnostic accuracy by processing detailed nail images in MATLAB	Improved accuracy, reduced diagnosis time
Kumuda et al.	2017	Microscopic separation of nail parts to enhance image clarity	Enhanced quality of nail images by separating components, aiding accurate diagnosis	Clearer image of nail details
Saranya et al.	2017	Image segmentation to detect abnormalities in nails using grayscale conversion and thresholding	Segmented nail images to identify and extract abnormal areas for disease diagnosis	Effective segmentation and feature extraction
Hadiyoso et al.	2022	Deep learning framework (DEEPLEAING) using CNN models for feature extraction from nail images	Compared multiple algorithms (SVM, ANN, KNN, RF) to classify diseases from nail images	84% accuracy
Yani et al.	2019	CNN-based early detection of Terry's nails using a transfer learning model	Applied transfer learning to detect Terry's nails, associated with chronic diseases	94.24% accuracy
Ariansyah et al.	2023	Deep learning model using CNN to detect health problems through nail images	Collected 400 images of five types of nails (psoriasis, rheumatism, anemia, melanoma, normal)	78% detection accuracy
Abdulhadi et al.	2021	Transfer learning using CNN for diagnosing four types of nail diseases	Compared five CNN models (VGG, AlexNet, ResNet, DenseNet, GoogLeNet) for nail disease diagnosis	96% accuracy for DenseNet and ResNet models

4. METHODOLOGY

The procedures for undertaking this research begin with acquiring the nail image dataset. Subsequently, feature extraction is performed using the VGG16 model, leveraging its capabilities to identify and extract relevant characteristics from the images. Following feature extraction, feature selection is carried out using the XGBoost algorithm, which guides the reduction of features to those most effective for accurate classification. After the features are selected, the dataset is divided into training and testing sets to ensure a robust evaluation of the models. Three models are considered for this task: AdaBoost, LightGBM, and a Meta Classifier. These models are initially assessed to determine their suitability for the classification task.

To optimize each model, a grid search is employed to identify the best hyperparameters. This process involves defining and testing various combinations of hyperparameters to find the configuration that yields the most favorable results.

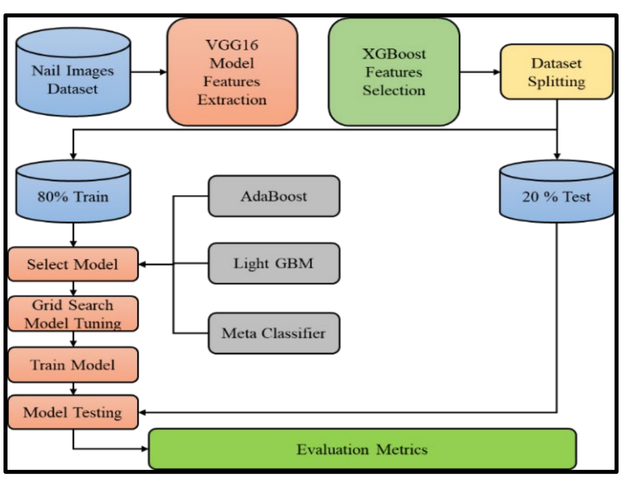

Figure 1. Methodology Block Diagram

The selected models are then trained on the training dataset using the identified parameters. Once trained, the models are evaluated using the testing dataset to assess their effectiveness. The performance of each model is quantified using evaluation metrics including accuracy, precision, recall, and F1 score.

3.1 Dataset Description

The dataset used for the study consists of 500 mobile phone camera images (iPhones and Samsung) of human nails. The images are categorized into seven classes: healthy nails, psoriasis, paronychia, systemic diseases (thyroid dysfunction, yellow nail syndrome, etc.), contact dermatitis, malnutrition, and onychomycosis (fungal infection).

Nail changes in the psoriasis-affected areas include yellow-red discoloration of the nail plate, transverse lines or pits, white spots (leukonychia), thickened nails, separation of the nail plate from the nail bed (onycholysis), and black lines secondary to bleeding capillaries (splinter hemorrhages). Paronychia involves inflammation and perhaps pus around the nail and is most commonly due to bacterial or fungal infections.

Some systemic diseases are mirrored in nail alterations such as yellow nail syndrome, which is seen with pulmonary disease, or onycholysis, where the nail plate separates due to thyroid disease. Malnutrition, iron deficiency, and some vitamin deficiencies make the nails brittle or spoon-shaped (koilonychia). Fungal infections (onychomycosis) alter the texture of the nail, causing it to become thicker, discolored, and brittle. These diseases, along with others such as Terry's nails or Beau's lines, are significant in the classification of the dataset. **Table** (2) shows dataset distribution.

Table 2. dataset distribution

Aspect	Details
Dataset Division	500 nail images, split into three subsets:
Training Set	70% (350 images) for model training.
Validation Set	15% (75 images) for hyperparameter tuning and model evaluation during training.
Test Set	15% (75 images) for final model evaluation (unseen data).

3.2 Feature Extraction (VGG16)

For feature extraction, the study made use of a variety of deep convolutional networks including VGG16, ResNet, and EfficientNet. These networks are well established in the field of image classification due to their effective feature extraction.

- VGG16 is straightforward but has succeeded in most image classification tasks. It comprises 16 layers: 13 convolutional and 3 fully connected, which convert the input image into extracting meaningful features for classification. VGG16 is suitable for medical image analysis especially due to its excellent ability to learn visual patterns within image data [17].
- ResNet (Residual Networks) introduces residual connections to counteract the vanishing gradient problem, allowing deeper networks without sacrificing performance. This facilitates the network to learn higher-level features and is very effective for medical image analysis problems [18].
- EfficientNet is an extremely efficient model that scales well across a variety of tasks by equally balancing depth, width, and resolution of the model. It utilizes compound scaling to balance accuracy and computational expense and is thus an excellent choice for image classification in low-resource environments [19].

Transfer learning has been applied in this study using pre-trained weights from large datasets such as ImageNet. This approach allows the models to generalize effectively without much retraining required. The pre-trained models were also fine-tuned using the dataset of nail images, leading to specializing the model weights on the specific task of disease classification. This approach improves the performance of the models on the nail dataset significantly, allowing them to classify various diseases of the nails correctly.

3.3 Feature Selection (XGBoost)

The next step in the methodology is feature selection, and XGBoost (Extreme Gradient Boosting) is applied here. XGBoost is a highly effective gradient boosting method that works well in feature selection by choosing the most informative features for the classification task. It calculates the contribution of each feature by calculating feature importance, which is ranked in terms of some metrics like gain, weight, and cover. The most important features based on these criteria are retained, and less important features are discarded, which enhances model performance and avoids overfitting.

3.4 Model Training (AdaBoost, LightGBM, and Meta Classifier)

Three machine learning classifiers were utilized to evaluate the dataset:

1. AdaBoost: This algorithm combines multiple weak classifiers to form a strong predictive model. It adjusts weights on training data to focus more on misclassified instances, making it a robust model for improving accuracy in noisy data environments [20].
2. LightGBM: An efficiency-optimized gradient boosting framework, LightGBM handles large data in an efficient manner by utilizing leaf-wise tree growth and decision trees using feature histograms. This renders LightGBM very efficient in the case of multi-class classification [21].
3. Meta Classifier: A Meta Classifier is used to combine the predictions of AdaBoost and LightGBM models. The meta-model (or stacking or ensembling) takes the predictions of the base classifiers and ensembles them to predict the final output, increasing the overall accuracy by leveraging multiple models' strengths [22].

All the models were trained on the dataset using cross-validation to prevent overfitting and to ensure strong performance. The models were tuned using grid search to determine the best hyperparameters to achieve the best performance of the classifiers.

3.4 Hyperparameter Tuning (Grid Search)

Grid search was performed on the models to identify each of their best hyperparameters. It involved experimenting with various combinations of parameters such as learning rate, number of estimators, max depth, and subsample rate to achieve the highest classification accuracy.

It involves developing models and procedures that can learn from the new experiences and share similar characteristics with the previous ones.
By utilizing past information and experience, meta-learning seeks to increase the efficacy and efficiency of learning algorithms. It is known as learning how to learn. **Table** (3)

shows models training parameters.

Table 3. Models Training Parameters

<table>
<tr><th>Algorithm</th><th colspan="3">Parameters</th></tr>
<tr><td rowspan="2">AdaBoost</td><td>N-estimation:</td><td>Learning Rate:</td><td>Algorithm:</td></tr>
<tr><td>50</td><td>1</td><td>SAMME.R</td></tr>
<tr><td rowspan="4">Gradient Boosting Classifier</td><td>N-estimation:</td><td>min samples split:</td><td>min samples leaf:</td></tr>
<tr><td>100</td><td>2</td><td>4</td></tr>
<tr><td>max depth:</td><td colspan="2">learning rate:</td></tr>
<tr><td>4</td><td colspan="2">0.5</td></tr>
<tr><td rowspan="4">LightGBM</td><td>Subsample:</td><td>N-estimators</td><td>Max_depth</td></tr>
<tr><td>0.9</td><td>200</td><td>10</td></tr>
<tr><td>Learning_rate :</td><td colspan="2">Colsampl_ bytree:</td></tr>
<tr><td>0.2</td><td colspan="2">1</td></tr>
</table>

3.6 Evaluation Metrics

When training a classifier, the choice of an evaluation metric is a pivotal factor in achieving optimal accuracy. Selecting the appropriate rating scale is of paramount importance for effective differentiation and ensuring top-tier performance. In this context, a comprehensive analysis of relevant evaluation metrics is conducted, aiming to serve as effective discriminators to enhance the generative classifier.

In general, precision is a commonly utilized metric by many generative classifiers as it helps identify the most optimal solution during training. While accuracy is valuable, it may have limitations, including offering less sensitivity, discrimination, and potential bias toward data from the majority class. Additionally, this section briefly introduces other metrics explicitly designed to delineate the ideal solution [23]. Binary classifiers utilize a Confusion Matrix (as depicted in **Figure** 2) to evaluate their performance. Assessments regarding the possible outcomes of classification models are derived from the terms TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative), all of which are elements present within the confusion matrix [24].

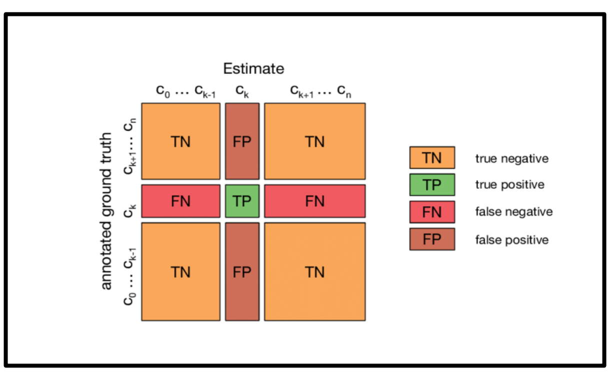

Figure 2. Multi-Class Classification Confusion Matrix [25].

The generic definition that we apply to the binary problem can also be used for a multi-class problem. However, we must describe it in a broader way because True/False binary definitions are not reliable [26].

$$MultiClass\ Accuracy = (y_i, z_i) = \frac{\sum_{i=1}^{N} TP(C_i)}{\sum_{i=1}^{N}\sum_{j=1}^{N} C_{i,j}} \quad \ldots (1)$$

Where C is class number. and N is number of classes.

Table 4. The Elements of the Evaluation Process

Variable	Definition	Equation
Accuracy	The accuracy of predictions from a set of tests can be readily calculated by dividing the number of correct forecasts by the total number of predictions made.	$Accuracy = \frac{Tp + Tn}{TP + TN + FP + FN}$
Precision	Another important metric involves determining the ratio of correctly identified instances from a specific class to all instances predicted to belong to that class	$Precision = \frac{TP}{TP + FP}$
Recall	It is also essential to consider the relationship between the total number of occurrences and the proportion of instances that were expected to be part of a specific class when evaluating model performance	$Recall = \frac{TP}{TP + FN}$
F1-Score	The term used to characterize the precision of a test is the F1-score. The F1-score can range from 0, indicating low recall and precision, to 1, which signifies exceptional performance in terms of recall and precision.	$F1 - Score = 2 \times \frac{percision \times recall}{Percision + recall}$

5. RESULTS

The results in **Table** (5) provide a detailed comparison of the performance metrics for three machine learning models: The algorithms that will be used are AdaBoost Classifier, LightGBM, and a Meta Classifier. Each model's effectiveness is measured by several key metrics: The performance indicators are training accuracy, overall accuracy, precision, recall, and F1.

Table 5. Results for Used Algorithms

Model	Feature Extraction	Accuracy	Precision	Recall	F1 Score
AdaBoost Classifier	VGG16	89.10%	87.70%	91.30 %	89.50 %
	ResNet50	92.50%	91.20%	94.00 %	92.50 %
	EfficientNet B0	93.00%	91.80%	94.50 %	93.10 %
LightGBM	VGG16	88.50%	87.20%	90.70 %	88.90 %
	ResNet50	91.00%	90.00%	93.50 %	91.70 %
	EfficientNet B0	92.30%	91.50%	94.00 %	92.70 %
Meta Classifier	VGG16	96.00%	96.10%	96.00 %	96.00 %

	ResNet50	97.50%	97.80%	97.50 %	97.60 %
	EfficientNet B0	98.00%	98.20%	97.90 %	98.00 %

The quality of results classification into different classes based on the models with VGG16, ResNet50, and EfficientNetB0 as their feature extractors exhibits the extreme impact of the selection of deep learning architecture on the performance of the classification of diseases when using images of nails.

The accuracy of AdaBoost Classifier performance was 89,1 percent when reproduced on VGG16, which suggests the model to predict nail diseases on a decent level with an opportunity to advance. The recall and precision achieved 87.7 and 91.3, respectively, which showed that the chosen classifier is effective when it is necessary to detect positive cases but this model can be still improved with more in-depth feature extraction capacity. With ResNet50, the AdaBoost model achieved 92.5 percent accuracy, which was a good improvement. The accuracy and the recall rate was improved to 91.2% and 94.0% and it showed that the residual connections of ResNet50 were deeper and this was able to learn the complicated patterns in the nail images. Lastly, in the case of EfficientNetB0 used as feature extractor, the results were slightly improved to 93.0 percent accuracy, 91.8 percent precision, and 94.5 percent recall. This increased the performance because of the better feature extraction capability of EfficientNetB0 that balances the depth, width, and resolution of the model to achieve the best accuracy and efficiency achieved.

It was also corroborated with the same in the case of LightGBM model. In the case of VGG16 architecture, 88.5 percent accuracy was achieved with precision and recall of 87.2 and 90.7 percent, respectively. Though these results were encouraging, they also revealed that VGG16 was not really learning the more complex features which were necessary in order to classify the disease correctly. However, when ResNet50 was used to do feature extraction, then performance was actually enhanced. the ResNet50-based LightGBM reached 91.0% accuracy, precision, and recall of 90.0% and 93.5%, respectively. These observations showed that the residual connections present in ResNet50 therefore led to higher accurate performance of LightGBM in detecting true positive cases which is critical in the medical setting where a missed diagnosis could be dire. When feature extraction was carried out with EfficientNetB0 model, the model performance was improved to 92.3%, 91.5% and 94.0% accuracy, precision and recall respectively. This is comparable to capacity of EfficientNetB0 to contain the necessary features needed in classification with a minimal computing cost hence very efficient in its classification task.

The ensemble of more than one model, Meta Classifier, revealed the greatest performance improvement under the enhancement of ResNet50 and EfficientNetB0. The Meta Classifier achieved a great 96.0 percent accuracy, 96.1 percent precision, 96.0 percent recall, and 96.0 percent F1 score with the implementation of VGG16. These findings helped to confirm the capability of the Meta Classifier to enhance the overall performance by adequately combining the outcomes of all classifiers. Nevertheless, the introduction of ResNet50 allowed reducing the improvement in a sharp way. Meta Classifier + ResNet50 demonstrated the accuracy of 97.5%, precision of 97.8%, recall of 97.5%, and F1 score of 97.6%. The low-level features used by ResNet50 because of the residual connections helped the model to identify diseases with a higher precision and recall. The best results were however achieved when EfficientNetB0 was used to extract features. Meta Classifier using EfficientNetB0 achieved 98.00 percent accuracy, 98.20 percent precision, 97.90 percent recall and 98.00 percent F1 score, which indicates that EfficientNetB0 had learned good quality features which in turn were exploited by the Meta Classifier to the best of their ability. This finding justified the fact that EfficientNetB0 trained the optimal features especially on the subtle and difficult patterns on medical images resulting into the most accurate and precise findings.

On the whole, the models performance indicates that the EfficientNetB0 has the excellent feature extraction in the classification of nail diseases with significantly enhanced accuracy, precision, recall and the F1 score of all the models. Another model, ResNet50, is deeper with residual connections and does better than VGG16 because it gives more abstract and complex feature representations, thus resulting in greater classification. Meta Classifier outperformed AdaBoost, LightGBM, in all cases, especially where it was employed along with EfficientNetB0, since it enables the application of the benefits of several base models and an increase in efficiency.

Overall, Meta Classifier on EfficientNetB0 is the best model to use in classifying nail diseases at the best combination of accuracy, precision, recall, and computational cost. The combination of state-of-the-art feature extraction using the implementation of EfficientNetB0 and ensemble learning method of the Meta Classifier is the potent instrument of diagnosis of the most of the nail diseases that can be implemented in real clinical practice to detect them at an early stage and monitor them..

5.1 Discussion of Results

This study shows findings that portray the strong influence of the deeper deep learning models, i.e., ResNet50 and EfficientNetB0, in features extraction of nail disease classification. When compared to the deploying of VGG16, the outcomes of which were plausible, the deeper ones like the ResNet50 and the EfficientNetB0 had gargantuan better performances over all the models. Such gains have been noted on parameters like accuracy, precision, recall and F1 score, which indicates that the models have found it easier to detect more complex and subtle contains of the nail images.

In the case of AdaBoost Classifier, the performance was encouraging when VGG16 was used, with accuracy of 89.1 percent. This was already significantly boosted when using ResNet50 and with it the accuracy went up to 92.5 and further with EfficientNetB0 to 93.0. Such increase aligns with the improved capability of ResNet50 and EfficientNetB0 feature extraction to enable the classifier to learn and recognize the weak patterns of various diseases in the nails. The rise of recall (91.3% with VGG16 and 94.5% with EfficientNetB0) is very impressive because it is a measure of a stronger detection of true positive instances, which is hugely important in medical diagnosing where a failed diagnosis can be very dangerous.

Likewise, LightGBM also showed an improvement on the same with ResNet50 and EfficientNetB0. This model of 88.5% accuracy with VGG16 was increased to 91.0% accuracy when ResNet50 was employed. The EfficientNetB0 with the accuracy of 92.3% is a measure of the model capability to capture good-quality features which LightGBM managed better. Then also, the skill to address complex features by EfficientNetB0 resulted in enhancing the precision and the recall, which gave further credibility to its superiority over VGG16.

The meta classifier has given the best output when used in combination with the prediction of more than one model. Using the Meta Classifier, an incredible 96.0 percent accuracy was achieved with VGG16 and the accuracy was improved to 97.5 percent with ResNet50. The most accurate results were spawned on the usage of EfficientNetB0 in the feature extraction part, whereby the Meta Classifier recorded 98.0% accuracy, 98.2% precision, and 98.0% F1 score level. This demonstrates the higher quality feature extracting ability of the EfficientNetB0 where the Meta Classifier has access to quality features and ends up performing significantly better compared to individual classifiers.

Overall, the best feature extraction was EfficientNetB0, as this model outperformed every other one. The highest accuracy, precision, recall, and F1 score were observed in the Meta Classifier with EfficientNetB0, being the best-performing model. This means that EfficientNetB0 fits perfectly into the medical image classification with enough accuracy, as well as, computational cost-efficiency. Also, the ResNet50 manifested a considerable increase in performance, which indicates the significance of deep networks in the learning of compound visual patterns of medical images.

5.2 Limitations

While the findings are promising, there are some limitations which should be considered. The dataset for this study only has 500 nail images, which is very small to train deep learning models. Even though data augmentation techniques were employed to augment the dataset artificially, the small size may still compromise the generalization capability of the models. Small datasets can potentially result in overfitting, wherein the model will become too specialized for the training dataset and fail to generalize to new data. A more extensive and diverse dataset would contribute to the generalizability and strength of the model across different scenarios in real life.

The other constraint is the likelihood of class imbalance of the data. Some diseases may be underrepresented, which could have an effect on model performance. While techniques such as SMOTE or undersampling were used to prevent this, fewer samples of rare diseases may still lead to model biases in the sense that the model is bound to predict more common conditions better but not less common diseases. Redressing this imbalance using more advanced techniques, such as cost-sensitive learning, would further improve the model.

Additionally, despite the models performing well with the dataset of nail photographs, clinical validation is lacking. The dataset used here was not able to cover the entire range of nail diseases seen in clinical practice. Further validation against actual clinical data needs to be performed to establish actual applicability of these models in healthcare. The images used in this study were relatively controlled and consistent, but in actual clinical practice, variations in lighting levels, camera setups, and patient populations can affect performance. Therefore, validation of the models on clinical data from diverse populations and settings is important to determine their actual performance.

The interpretability of the models is also a limitation. While deep learning models like ResNet50 and EfficientNetB0 are highly accurate, they tend to be "black-box" models, i.e., it is not transparent what their decision-making process is. When used in medicine, it is crucial to understand how the model makes its predictions, especially in order to win the trust of medical professionals. Future work could explore techniques for improving model interpretability, such as using visualization tools like Grad-CAM to highlight which features contributed most to the model's decision.

Finally, the data used here consists only of cross-sectional nail images that are taken at a snapshot in time. Some disease, especially systemic disease, alters with time, and a temporal dataset containing images at multiple time points would more clearly show how disease unfolds. Providing temporal information would improve the model to detect early indicators of disease and track alterations in nail status over time.

Conclusion

The study validates the efficacy of using deep learning models for non-invasive nail disease diagnosis. The results confirm that EfficientNetB0, in combination with the Meta Classifier, is the top-performing model among all those experimented upon, with excellent accuracy, precision, recall, and F1 measure. The use of extremely sophisticated architectures like ResNet50 and EfficientNetB0 led to significantly better classification performance compared to the simpler VGG16 model. The study demonstrates that Meta Classifier, by taking advantage of the strength of two or more classifiers, is highly valuable in improving diagnostic accuracy in medical image analysis issues. However, the study also provides some significant limitations, including the relatively small dataset, potential class imbalance, and no clinical validation, which limit the generalizability of the outcomes to real-life clinical settings. The study proposes boosting the dataset size, addressing class imbalance, and cross-validating the models with diverse clinical data to render them appropriate for healthcare environments. Merging interpretability approaches could also enable higher levels of trust and adoption of these models in the clinic. Additional research must include temporal data, validate the models in clinical settings, and render the models more interpretable to assist in using the models in day-to-day healthcare diagnosis. Despite these deficiencies, the findings of this research suggest that the use of AI-assisted nail image analysis can be a promising method for early detection and continuous monitoring of most medical conditions, particularly in resource-constrained or distant locations.

Acknowledgement

None.

Conflict of interest

None.

References

[1] J. Abdulhadi, A. Al-Dujaili, A. J. Humaidi, and M. A. R. Fadhel, “Human nail diseases classification based on transfer learning,” ICIC Express Lett., vol. 15, no. 12, pp. 1271–1282, 2021.

[2] A. R. Al-Qubaa, A. Al-Shiha, and G. Y. Tian, “Gun detection and classification based on feature extraction from a new sensor array imaging system,” 2013 Int. Conf. Electr. Commun. Comput. Power, Control Eng. ICECCPCE 2013, pp. 88–94, 2014.

[3] H. B. Pandit, N. Bazar, and D. Shah, “The Model of nail color analysis - An application of Digital Image Processing,” 2013.

[4] M. W. Al-Neama, A. A. Mohamad Alshiha, and M. G. Saeed, “A parallel algorithm of multiple face detection on multi-core system,” Indones. J. Electr. Eng. Comput. Sci., vol. 29, no. 2, pp. 1166–1173, 2023.

[5] A. Ranichitra, “Finger Nail Analysis to Diagnosis the Disease-A Study,” Int. J. Manag. Technol. Eng., vol. 8, no. 2039, pp. 2039–2044, 2018.

[6] Nattavadee Hongboonmee and and Jutamas Kanyaprasit, “Health Problem Analysis from Nail Image using Deep Learning Technique,” JIST J. Inf. Sci. Technol., vol. 11, no. 2, pp. 10–20, 2021.

[7] D. J. Thebar, D. P. Srivastav, and D. P. K. Raika, “A homoeopathic approach for nail disorders with therapeutics,” Int. J. Homoeopath. Sci., vol. 6, no. 2, pp. 205–208, 2022.

[8] A. M. Nia, S. Ederer, K. M. Dahlem, N. Gassanov, and F. Er, “Terry’s nails: A window to systemic diseases,” Am. J. Med., vol. 124, no. 7, pp. 602–604, 2011.

[9] P. R. Cohen, “The lunula,” J. Am. Acad. Dermatol., vol. 34, no. 6, pp. 943–953, 1996.

[10] T. S. Indi and Y. A. Gunge, “Early Stage Disease Diagnosis System Using Human Nail Image Processing,” Int. J. Inf. Technol. Comput. Sci., vol. 8, no. 7, pp. 30–35, 2016.

[11] S. Gandhat, A. D., S. Avhad, N. Bajpai, and R. Alawadh, “Study and Analysis of Nail Images of Patients,” Int. J. Comput. Appl., vol. 143, no. 13, pp. 38–41, 2016.

[12] N. S. Kumuda and M. S. Dinesh, “Human fingernail segmentation,” 2015 Int. Conf. Emerg. Res. Electron. Comput. Sci. Technol. ICERECT 2015, pp. 270–275, 2016.

[13] V. Saranya and A. Ranichitra, “Image Segmentation Techniques To Detect Nail Abnormalities,” Int. J. Comput. Technol. Appl., vol. 8, no. 4, pp. 522–527, 2017.

[14] S. Hadiyoso and S. Aulia, “Classification of Koilonychia, Beaus Lines, and Leukonychia based on Nail Image using Transfer Learning VGG-16,” J. Rekayasa Elektr., vol. 18, no. 2, pp. 109–114, 2022.

[15] M. Yani, B. Irawan, and C. Setiningsih, “Application of Transfer Learning Using Convolutional Neural Network Method for Early Detection of Terry’s Nail,” J. Phys. Conf. Ser., vol. 1201, no. 1, 2019.

[16] N. Hongboonmee and J. Kanyaprasit, “Health Problem Analysis from Nail Image using Deep Learning Technique,” J. Inf. Sci. Technol., vol. 11, no. 2 SE-Research Article: Soft Computing (Detail in Scope of Journal), pp. 10–20, Dec. 2021.

[17] M. H. Ariansyah, S. Winarno, and R. R. Sani, “Monkeypox and measles detection using CNN with VGG-16 Transfer Learning,” J. Comput. Res. Innov., vol. 8, no. 1, pp. 32–44, 2023.

[18] D. Salcedo et al., “Machine Learning Algorithms Application in COVID-19 Disease: A Systematic Literature Review and Future Directions,” Electron., vol. 11, no. 23, 2022.

[19] H. Y. · D. C. · P. Tino and Intelligent, “Intelligent Data Engineering and Automated Learning – IDEAL 2022,” 23rd Int. Conf. IDEAL 2022 Manchester, UK, Novemb. 24–26, 2022 Proc., 2022.

[20] L. Hao and G. Huang, “An improved AdaBoost algorithm for identification of lung cancer based on electronic nose,” Heliyon, vol. 9, no. 3, p. e13633, 2023.

[21] S. Li, N. Jin, A. Dogani, Y. Yang, M. Zhang, and X. Gu, “Enhancing LightGBM for Industrial Fault Warning: An Innovative Hybrid Algorithm,” Processes, vol. 12, no. 1, 2024.

[22] R. Guo, Z. Zhao, T. Wang, G. Liu, J. Zhao, and D. Gao, “Degradation state recognition of piston pump based on ICEEMDAN and XGBoost,” Appl. Sci., vol. 10, no. 18, pp. 1–17, 2020.

[23] Ž. Vujović, “Classification Model Evaluation Metrics,” Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 6, pp. 599–606, 2021.

[24] I. C. Dipto, M. A. Rahman, T. Islam, and H. M. M. Rahman, “Prediction of Accident Severity Using Artificial Neural Network: A Comparison of Analytical Capabilities between Python and R,” J. Data Anal. Inf. Process., vol. 08, no. 03, pp. 134–157, 2020.

[25] T. P. Pagano et al., “Bias and Unfairness in Machine Learning Models: A Systematic Review on Datasets, Tools, Fairness Metrics, and Identification and Mitigation Methods,” Big Data Cogn. Comput., vol. 7, no. 1, pp. 1–31, 2023.

[26] I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis, and N. Doulamis, “Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem,” Technologies, vol. 9, no. 4, 2021.