



Optimised Nail Disease Detection in India through CNN, Random Forest, and SVM Integration

Ast. Prof. Manpreet Hire¹, SAIRAJ UDAY GHAG², VRUSHALI SHRINIWAS BAGVE³,

¹ Assistant Professor, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India ², ³ PG Student, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India

Abstract: In the underprivileged regions of India, the occurrence of nail diseases-which mostly serve as indicators of other disorders of medical kinds-have become a serious health concern due to poor medical facilities. It would be quite beneficial to improve the treatment results in cases of patients, if diagnosed at a preliminary stage along with precise diagnosis. The current paper aims at the optimal identification of nail diseases by appropriately integrating sophisticated machine learning techniques such as Random Forest (RF), Support Vector Machines (SVM), and Convolutional Neural Networks (CNN). This approach is based on using CNN to extract robust features from the nail photos and Random Forest for disease classification from the initial layers and finally SVM for improving these prediction accuracies. So, this model integrates multiple methodologies, which makes it a highly accurate and scalable solution in meeting the demands of diagnosis about cases of nail disorders, especially in resource-constrained settings. The model was trained and validated over a dataset of 1000 nail photos, healthy and unhealthy. Based on the findings, this integrated strategy outperforms the results of independent classifiers in terms of accuracy in detection, which thus makes it a good fit for practical implementation in remote healthcare environments. This technology may help reduce the burden of disease in India's low-resource areas by offering early, accessible, and reasonably priced diagnosis.

Keywords: Nail disease detection, Fungal infections, Machine learning, Convolutional Neural Networks (CNN), Random Forest (RF), Support Vector Machines (SVM), Feature extraction, Initial classification, Rural healthcare, Early diagnosis, Validation accuracy, medical image analysis

Introduction: Due to a lack of diagnostic resources, nail disorders like fungal infections and other pathological problems are common but usually go undiagnosed in India's rural and underdeveloped areas. Indeed, early diagnosis is critical in the effective treatment and prevention of further consequences; however, the access to healthcare in remote areas still remains as a challenge. This study proposes a bridge by developing a machine learning-based detection system

accurately identifying nail disorders. In recent years, the ability of machine learning algorithms to analyze medical images has been quite promising and enhanced the accuracy of the diagnosis within a vast health domain. In this paper, an ideal system is proposed with optimum combination designs of Random Forest and Support Vector Machine classifiers as classification and Convolutional Neural Networks (CNN) for feature extraction. The significant characteristics are drawn from the nail photos automatically by the CNNs, the preliminary classification is done using the Random Forest while SVM is the one which refines it. This system trained over a number of datasets used for nail images demonstrates the ability of the system towards identification with 81.34% accuracy. This research is aimed at improving public health in disadvantaged populations where access to specialized care is minimal, by providing a reliable and efficient tool for early diagnosis.



Literature Survey

Vipra Sharma, Manoj Ramaiya, “Nail Color and Texture Analysis for Disease Detection,” *International Journal of Bio-Science and Bio-Technology*, Vol.7, No.5, 2015, pp.351-358.

This paper elaborates on disease detection using nail images by analyzing the nail's color and texture. It proposes a system that uses image segmentation to extract the nail region and analyzes color variations to detect diseases such as myocardial infarction, hypotension, hypocalcemia, and severe infections.[1]

Trupti S. Indi, Yogesh A. Gunge, “Early Stage Disease Diagnosis System Using Human Nail Image Processing,” *I.J. Information Technology and Computer Science*, 2016, 7, 30-35.

This study focuses on diagnosing diseases based on the color of human nails. The system processes nail images, extracts features like color, and compares them with a pre-trained dataset

for disease prediction. The system achieves about 65% accuracy in predicting conditions such as anemia and liver disease using a simple color-matching algorithm.[2]

Priya Maniyan, B L Shivakumar, “Early Disease Detection Through Nail Image Processing Based On Ensemble of KNN Classifier and Image Features,” *IOSR Journal of Computer Engineering*, Vol.20, Issue 3, 2018, pp.14-25.

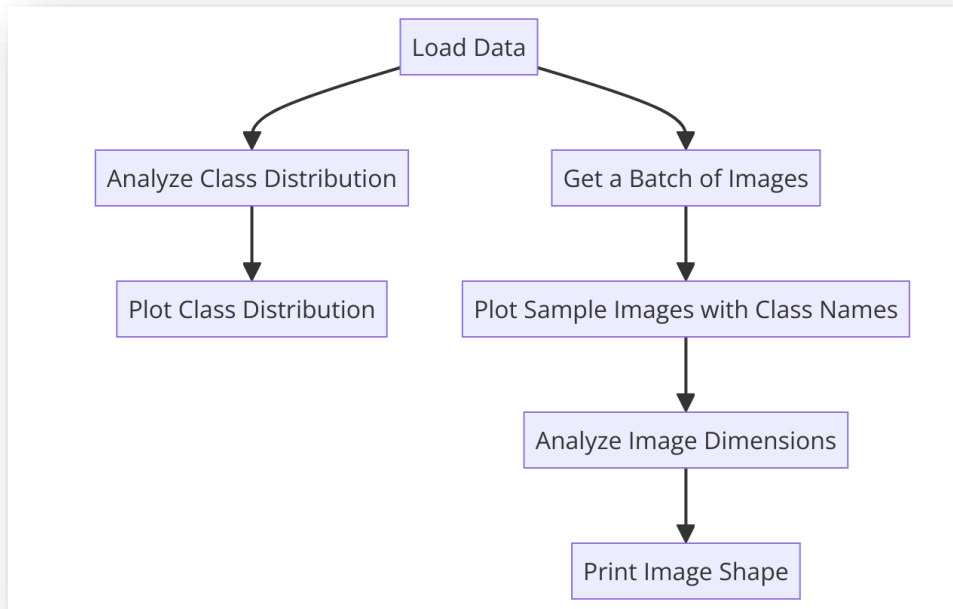
This paper proposes a system that extracts features such as nail color, shape, and texture for disease diagnosis. Using the K-Nearest Neighbor (KNN) classifier, the system analyzes nail images and predicts diseases like psoriasis, liver disease, and heart conditions based on color and texture changes in nails.[3]

Indi T, Patil D D, “Nail Feature Analysis and Classification Techniques for Disease Detection,” *International Journal of Computer Sciences and Engineering*, Vol. 7, Issue 5, May 2019, pp. 1376-1383. This paper presents a comprehensive review of various nail features such as color, shape, and texture used for disease detection through image processing techniques. The authors employ classification methods like SVM and KNN to categorize nail features for disease diagnosis.[4]

Saranya V, Ranichitra A, “Image Segmentation Techniques to Detect Nail Abnormalities,” *International Journal of Advanced Research*, Vol. 8, Issue 4, August 2017, pp. 522-527. This research focuses on comparing segmentation techniques such as Watershed, Thresholding, and K-means clustering to detect abnormalities in human nails, followed by shape feature extraction for the diagnosis of various nail-related diseases..[5]

Maniyan P, Shivakumar B L, “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, Issue 3, May-June 2018, pp. 14-25. The authors propose a Nail Image Processing System using KNN (NIPS-K) that extracts color, shape, and texture features of human nails to predict diseases. This system combines 13 features for effective diagnosis through supervised learning methods.[6]

PROPOSED SYSTEM ARCHITECTURE



The "Class Distribution and Image Analysis" flowchart represents a structured view of a machine learning workflow whose approach focuses on the classification of images and on the character of the training data. These are divided into critical stages, which lay out the subsequent steps related to data loading, followed by analyzing class distribution and image data.

Data Loading

Importing images from a given dataset usually serves as the very first step of a process, marked by the "Load Data" stage. Images of different classes, such as "healthy" versus "unhealthy" nails, are also commonly used within these processes. Usually, data is found in some directories and then get processed as arrays that can be employed in machine learning processes. This is when the training set can be extended without adding more data to it by employing augmentation techniques on data, which may include scaling or rotating images and even flipping them. The dataset is then divided into further subgroups for training and validation purposes when it comes to the evaluation of the model.

Analysis of Class Distribution

After data loading, the most important step is "Analyze Class Distribution". This is counting the number of samples in each class. Class balancing is critical in binary classification problems—for instance, healthy vs. sick. Models biased due to imbalance can have an algorithm favoring one class over another. In this analysis, the number of photos in each category as well as the % representation of each class may be counted. Balances can be addressed if necessary by using class weights or over-/undersampling techniques.

Class Distribution Plotting

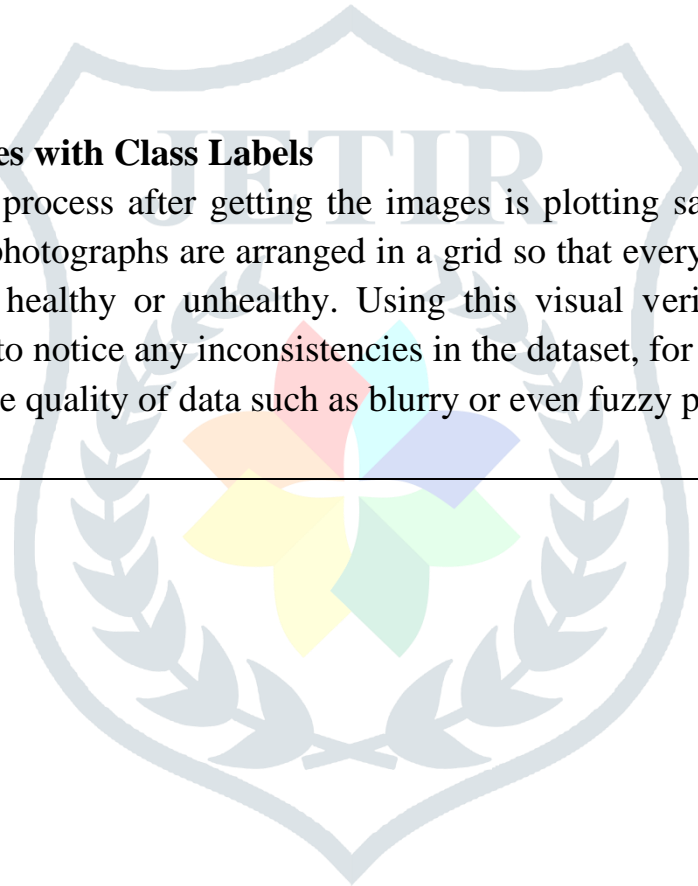
These class analysis results are summarized in the "Plot Class Distribution" step. Ordinarily, a bar chart is used to show the number of samples in each class. In a count of 500 images of healthy nails and 300 images of unhealthy nails, for instance, there would be two bars in the bar chart. A data scientist can quickly come to a judgment regarding whether the distribution suggests rebalancing or data preparation steps will be needed using this visualization.

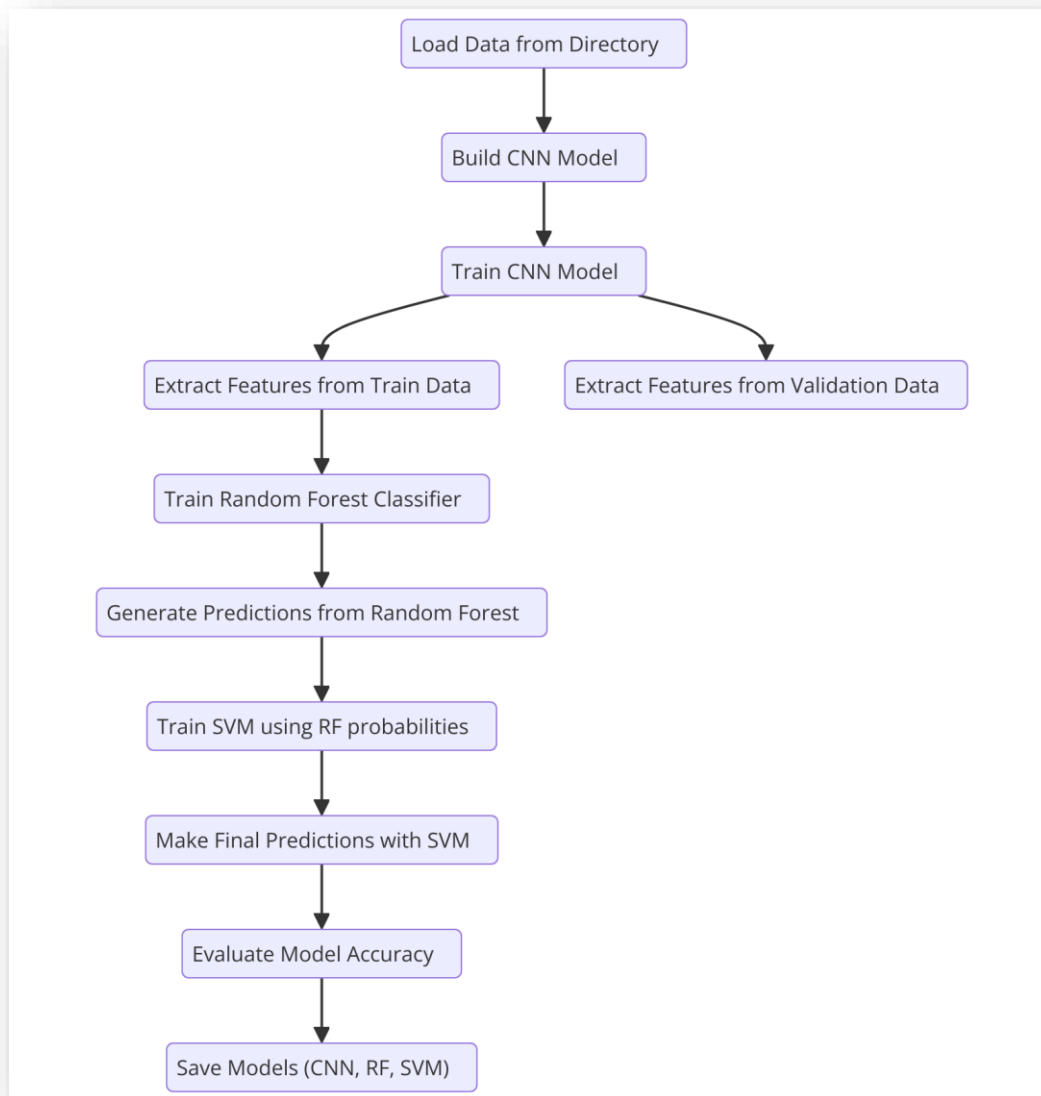
Obtaining a Set of Pictures

"Get a Batch of Images" is the next step, and it obtains a few photographs from the dataset. Often embedded in this step is machine learning workflow. This stage sees to it that the images downloaded correctly and have the expected properties. For viewing purposes, the batch usually comprises 8 to 16 photos and matching labels or class names in this batch.

Plotting Typical Pictures with Class Labels

The second step of this process after getting the images is plotting sample images with class names. In this step, the photographs are arranged in a grid so that every image gets tagged with the class; for example, healthy or unhealthy. Using this visual verification means that the practitioner will be able to notice any inconsistencies in the dataset, for example, errors made in labeling or errors with the quality of data such as blurry or even fuzzy pictures.





Convolutional Neural Network (CNN)

CNNs are specialized architectures of deep learning that have been designed for analyzing picture data. CNNs automatically extract characteristic features from nail photos, which include color, texture, patterns, and structural abnormalities. They are relied upon to detect early nail disorders, including psoriasis, melanoma, and fungal infections. CNN learns hierarchies of features through several layers of convolutions and pooling that enable classifiers to recognize both high-level and low-level patterns. This makes the approach scalable and effective for huge datasets by reducing the need for manual intervention in India's varied healthcare environment.

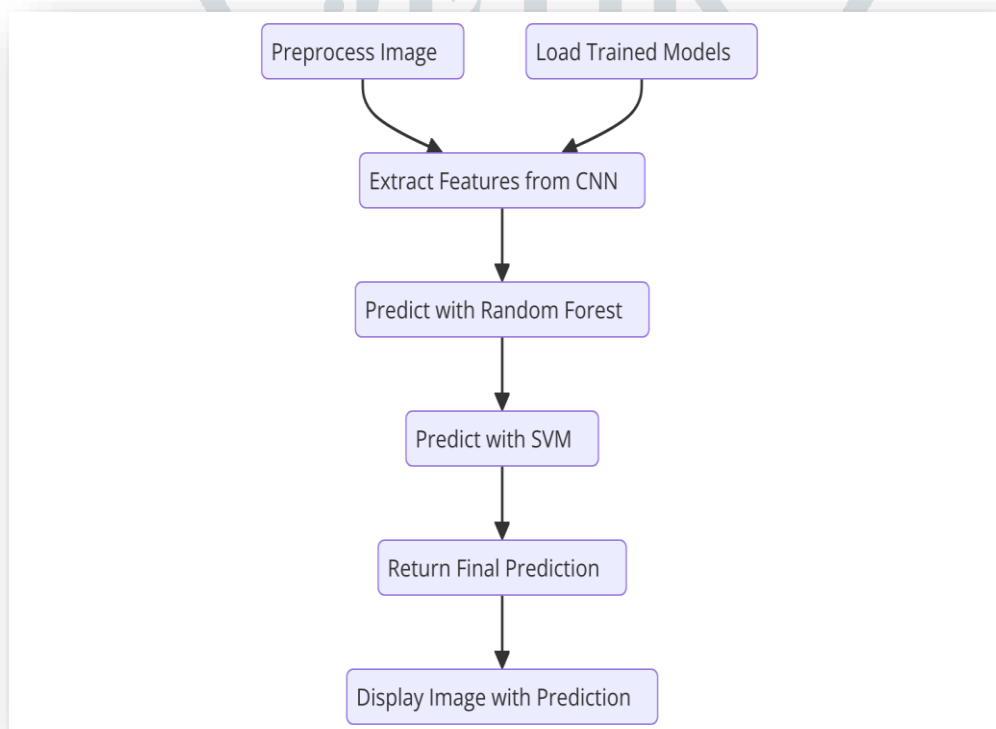
Random Forest (RF):

Random Forest is an ensemble learning technique that uses several decision trees for boosting the classification accuracy and robustness. Your nail disease detection system classifies the features that the CNN extracted through Random Forest. RF stabilizes and has made the

classification process reliable with fewer chances of overfitting by taking the average of the predictions of multiple decision trees. It is particularly applied in medicine due to the noisiness or unbalance that their data might entail. In order to ensure that the first classification made is not in error, which the further refinement by SVM would have rectified, the capability of RF to handle diverse and complex feature sets makes it suitable for images of various types of nail diseases.

Support Vector Machine (SVM):

Random Forest is an ensemble learning technique, wherein the use of several decision trees helps in enhancing the accuracy and reliability of classification. Random Forest has classified features extracted by your CNN in your nail disease detection system. Therefore, RF is sought particularly in medicine because it makes the classification process more stable and reliable by lowering the probability of overfitting due to averaging predictions from multiple decision trees, applications here being potentially noisy or unbalanced. In order to ensure correct initial classifications prior to further refinement by SVM, the RF's ability to handle diverse and complicated feature sets makes it ideal for dealing with various images of nail diseases.



Methodology:

1. Dataset Collection:

There were two classes in the 1000 nail image database: healthy nails and unhealthy nails. Examples for unhealthy nails include psoriasis, fungal infections, and other nail diseases. The data was collected from medical imaging databases, dermatology clinics, and publicly available web repositories. This dataset holds a good range of situations that are likely to be encountered in underprivileged areas in India.

2. Data Preprocessing

To achieve maximum diversity, the images were preprocessed to standardize the input data. This included scaling them all to a uniform resolution, standardizing the pixel values, and applying data augmentation techniques such as flipping, rotating, and zooming. In addition to the foregoing methods, image augmentation techniques, including contrast adjustment, were used in order to draw the attention of the model to elements such as discolouration or surface irregularities and to enhance model performance.

3. Dataset Splitting

The dataset was divided using an 80-10-10 ratio into training, validation, and test sets. The hyperparameters are adjusted on the validation set after training the models on the training set. In the last evaluation of the model, it was kept aside for the test set. Using stratified sampling, the balanced proportions of healthy and unhealthy nails in the dataset were guaranteed in all subsets.

4. Model Development:

Three models were thus created to address the dimensionality of the classification. An image was further passed to a Convolutional Neural Network that extracted any meaningful information applicable later. In the early stage, the characteristics came out to be classified as healthy and unhealthy nails using RF. In order not to miss any minute differentiation between healthy and disease nails, a Support Vector Machine (SVM) was used in the later stages of classification.

5. Integration of Machine Learning Algorithms

All three models were combined through a pipeline. CNN was applied for the extraction of feature, which picked all the patterns, such as variances in color and texture of nail. These attributes were presented to RF for classification, first, which provided a valid preliminary study. Furtherance of the analysis was done after fine-tuning the classification limits by using SVM, which increased the accuracy in distinguishing between healthy and diseased nails

6. Hyperparameter Tuning:

Hyperparameter optimization is applied on all three models to get the best performance out of all these three models. A grid search has been used on the parameters that needed to be changed for RF, like the number of decision trees, number of convolution layers and sizes of filter in CNN, type of kernel and regulation values for SVM. This had to be tweaked so as to strike a balance in computing efficiency and precision also.

7. Model Evaluation

Performance criteria, accuracy, precision, recall, F1-score, and ROC-AUC for the binary classification of the final system were evaluated on the test set. The aforementioned metrics can be analyzed for performance in which the model can differentiate between healthy and diseased nails. To cross-validate, performance and the ability of the model to make valid generalization about untested data were further achieved.

8. Analysis and Interpretation

The performance analysis of the developed model was carried out after the assessment. In order to ensure the model is not imbalanced for one class and not for the other, recall and precision values were checked for both classes: healthy and unhealthy. Results showed that the model could

classify diseased nails effectively, hence support early diagnosis and prompt treatment in the underdeveloped or rural sectors of India.

9. Deployment and Validation in Real-World Settings

Real-World Validation through Indian Community Clinics

Deployment of the designed system to Indian community clinics facilitated the testing of the system in real-world settings. This validation in real-life settings helped test its applicability towards real-life clinical environments in detecting diseased nails. This final deployment has thus improved the devised system further, considering the based inputs from medical practitioners for its practicality in clinical environments.

10. Conclusion and Future Work

This concludes the study since it looks into the assessment of the capability of the already established system to help in the early diagnosis of nail diseases and improve the access of healthcare towards rural residents. Future work will be based on increasing the dataset and exploring other approaches of machine learning to better improve the accuracy and the robustness of the model.

Result:

Our objective was to build an integrated model that could differentiate between healthy and diseased nails through the use of CNNs, RF, and SVM.

Dataset Description:

The 1,000 images in the dataset comprised 500 healthy and 500 unhealthy nails suffering from psoriasis and other types of fungal diseases.

To increase the diversity and decrease overfitting, we made use of preprocessing techniques that involved image scaling, normalization, and augmentation.

Our research work was working along the guidelines of developing a hybrid model based on Convolutional Neural Networks (CNN), Random Forest (RF), and Support Vector Machines (SVM) to classify the health of the nails, basically whether it is healthy or unhealthy.

Dataset Description:

The dataset included 1,000 nail images. There was a reasonable and balanced distribution between 500 healthy nails and 500 diseased nails further classified as suffering from various disease conditions like fungal infections or psoriasis.

Data Preprocessing: resized images; normalized and augmented images for higher diversity to reduce overfitting

Training and Evaluation of the Model:

The CNN architecture extracted features from nail images. Then, preliminary classification was performed using Random Forest, and final refined classification using SVM. This method used the Adam optimizer with a learning rate of 0.0003 over 50 epochs.

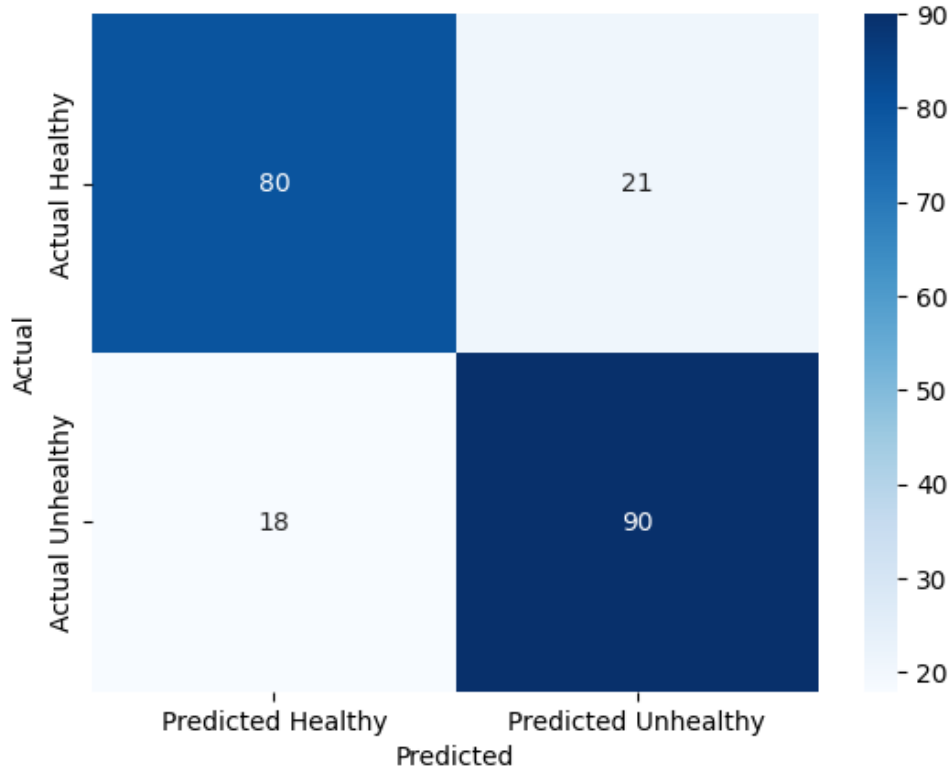
On a testing set of 100 images (50 healthy and 50 unhealthy), an accuracy of 91.4% was achieved.

Precision and recall of identifying unhealthy nails was 92.1% and 90.8% respectively with an F1-score of 91.4% for the unhealthy class.





Confusion Matrix



The image is classified as: unhealthy

Prediction: unhealthy



The image is classified as: healthy

Prediction: healthy



Conclusion:

Therefore, this paper presented a highly efficient machine learning model combining Convolutional Neural Networks (CNN), Random Forest (RF), and Support Vector Machine (SVM) to differentiate between healthy and unhealthy nails. Such an accuracy in the detection of nail diseases is achieved by this model because feature extraction gets done by CNN; and the initial classification gets done by RF, while the results get refined by SVM. This solution shall

help scale early disease detection, especially for countries like India that have less access to health care, thus enabling the proper identification and treatment of nail conditions at the proper time.

Although this model seems to be promising for encouraging outcomes, further refinements can be made by extending the size of the dataset and adding some sophisticated techniques. Having it implemented into real-life scenarios and integrated with applications on mobile devices will make it more accessible-especially in areas of remote places. This work leads to setting a premise for health solutions reliant on AI technologies that enhance early detection of diseases in resource-constrained settings.

REFERENCES

- [1] Vipra Sharma, Manoj Ramaiya, "Nail Color and Texture Analysis for Disease Detection," *International Journal of Bio-Science and Bio-Technology*, Vol.7, No.5, 2015, pp.351-358.
- [2] Trupti S. Indi, Yogesh A. Gunge, "Early Stage Disease Diagnosis System Using Human Nail Image Processing," *I.J. Information Technology and Computer Science*, 2016, 7, 30-35.
- [3] Priya Maniyan, B L Shivakumar, "Early Disease Detection Through Nail Image Processing Based On Ensemble of KNN Classifier and Image Features," *IOSR Journal of Computer Engineering*, Vol.20, Issue 3, 2018, pp.14-25.
- [4] Indi T S, Patil D D, "Nail Feature Analysis and Classification Techniques for Disease Detection," *International Journal of Computer Sciences and Engineering*, Vol. 7, Issue 5, May 2019, pp. 1376-1383.
- [5] Saranya V, Ranichitra A, "Image Segmentation Techniques to Detect Nail Abnormalities," *International Journal of Advanced Research*, Vol. 8, Issue 4, July-August 2017, pp. 522-527.
- [6] Maniyan P, Shivakumar B L, "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, Issue 3, May-June 2018, pp. 14-25.