VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

Smart Bridge Externship Program in Artificial Intelligence

Project Title: Early Stage Disease Diagnosis System Using Human Nail Image Processing

Team: 223

20BAI1304 – T S HARSHINI

20BRS1125 – T S VARSHINI

20BRS1113 – MADHUVANTHI SANKAR GANESH

Campus: Vellore Institute of Technology, Chennai

Table of Contents

1. Introduction

1.1 Overview

Human nails have special characteristics that make early illness detection always possible. Physical changes in the patient's nails are the earliest signs of disorders affecting the liver, lungs, heart, kidneys, and several other organs. Examples of diseases whose initial symptoms can be seen through the patient's nails include anemia, diabetes, and arthritis. It has been noted that this built-in protection system of the human body is frequently disregarded and seen in a humorous way. Our goal is to create an AI-based-automatic nail image classifier that can examine an image and identify a patient's vulnerable disease. In this research we have used five Transfer Learning models – VGG16, DenseNet121-ResNet50 (DN121-RN50), DN121-RN50 with custom ANN layers, DenseNet121, DenseNet121-VGG16.

1.2 Purpose

With this approach, our project can assist someone exhibiting the same symptoms in receiving appropriate care from a medical professional. Additionally, our study seeks to indicate emergency medical care to people who appear to have severe nail infections. Every user will receive simple medical support, and we hope to increase their understanding of their medical situation.

2. Literature Survey

2.1 Existing problem

Sugondo Hadiyoso et al. [1] used nail pictures and the VGG-16 Transfer learning model to categorize nail illnesses such as Koilonychia, Beaus Line, and Leukonychia. Despite the fact that this study's transfer learning-based nail disease classification model has a 96% accuracy rate, it did not contain a large number of disorders and hence has a limited understanding of the disease.

CNN (Convolution Neural Network) based disease classification was conducted on human nails by Lahari N et al.[2] . In this investigation, the patient's condition was identified using the color and different patterns on their nails. They had used GCLM feature extraction to extract the features, and then they had fed those features into CNN and ANN (Artificial Neural Network) models. The nail disease has been divided into four groups based on color: black nail, yellow or greenish nail, bluish nail, and white or light nail. Their classification of the patterns into vertical, horizontal, and yellow lines let them further describe the disease as having nervous disorders, asthma, liver issues, poor digestion, thyroid issues, diabetes, and other conditions. They further classified the same using DenseNet121 model which gave a better accuracy of about 90%. However this model again did not classify major nail diseases.

The paper Priya Maniyan et al.[3] proposes a system called Nail Image Processing System using Multiclass SVM (NIPS-McS) for analyzing human nail images and predicting various diseases. The system takes human palm images as input and focuses on the analysis of the nail portion. The methodology of the proposed system involves several steps. Firstly, the back sides of both palms are scanned using a camera with proper lighting. The palm region is then extracted from the background using image segmentation techniques. Next, the nail region is segmented, and features related to nail color, shape, and texture are extracted. The extracted features are combined to form a feature vector, which is compared with existing datasets for disease prediction. The system utilizes a Multiclass SVM classification method for disease classification

and prediction. The proposed system aims to assist doctors in the early diagnosis of diseases by leveraging the color, shape, and texture changes in nails, which are often early symptoms of various diseases. By using digital image processing techniques, the system provides a more accurate and efficient way to detect and predict diseases based on nail analysis. It offers the advantage of cost-effectiveness and convenience, as patients can potentially receive disease predictions from their own homes without the need for extensive laboratory tests. Overall, the paper highlights the potential of using nail image processing and machine learning techniques for disease prediction with an average accuracy of 89% and emphasizes the benefits it can bring to medical practitioners and patients in terms of early detection and diagnosis.

T S Indi et al.[4] proposed a system for the early diagnosis of diseases by analyzing human nail images. The system utilizes digital image processing techniques to extract meaningful features from nail images and applies machine learning algorithms for disease classification. The proposed system aims to leverage the changes in the color, texture, and shape of nails, which can serve as early indicators of various systemic and dermatological diseases. By analyzing these changes, the system can aid in the early detection and diagnosis of diseases, potentially leading to better treatment outcomes. Machine learning algorithms are employed to classify the nail images into different disease categories. The system is trained on a dataset of labeled nail images, where each image is associated with a specific disease or a healthy condition. The trained model can then predict the presence of diseases based on new, unseen nail images. However, this research provided an average of 65% accuracy only.

The paper R Mente et al.[5] presents a comprehensive overview of the use of fingernail images as a valuable tool in disease detection. Fingernails offer a unique and non-invasive window into our health, as they can exhibit visible changes in response to various systemic and dermatological diseases. The review highlights the significance of fingernail abnormalities as potential indicators of underlying health conditions such as nail psoriasis, onychomycosis (fungal nail infection), and other dermatological disorders. These changes can include alterations in nail color, texture, shape, thickness, and the presence of pits or ridges. To harness the diagnostic potential of fingernail images, the paper explores different imaging techniques and image processing methodologies employed in previous studies. These methods involve capturing high-resolution images of the fingernails and applying computer vision algorithms to analyze and extract relevant features from the images. Several studies have demonstrated promising results in disease detection using fingernail images. By leveraging machine learning techniques, such as support vector machines (SVM) and artificial neural networks (ANN), researchers have developed classification models to differentiate between normal and diseased nails based on the extracted features.

The research paper SA Yama et al.[6] proposes an ensemble model approach for the detection of nail diseases. Nail diseases can be indicative of underlying health conditions, and early detection is crucial for effective treatment. The study aims to improve the accuracy of nail disease classification by combining multiple classifiers through majority voting. The paper first discusses the importance of automated nail disease detection systems in clinical practice. It highlights the challenges associated with manual diagnosis and emphasizes the potential of computer-aided systems in improving diagnostic accuracy and efficiency. The proposed ensemble model consists of multiple classifiers, each trained on a specific set of features

extracted from nail images. The feature extraction process involves capturing various characteristics such as color, texture, shape, and contour information from the nail images. Popular machine learning algorithms like support vector machines (SVM), k-nearest neighbors (KNN), and random forests are used as classifiers. Each classifier makes predictions based on the extracted features of nail images. To make the final prediction, the researchers use a majority voting approach. Each classifier's prediction is considered as a vote, and the prediction with the most votes is chosen as the final result. This ensemble method benefits from the diversity of classifiers to enhance the overall accuracy and reliability of the nail disease detection system. By combining multiple classifiers and using majority voting, the model improves the identification of nail diseases.

2.2 Proposed Solution

The mentioned surveys have the intention of providing a remedy based on nail color and shape characteristics. The majority of models had incredibly few classifications, and several models omitted severe nail illnesses. In this study, we seek to offer a comprehensive answer that would give the patient a clear and distinct understanding of their medical situation. In this study, we present five Deep Neural Network models that operate on two various datasets. Our first model employs a dataset1 that divides the user-input image into 17 classes using a CNN (convolutional neural network)-based VGG-16 Transfer Learning model. The second model divides information into 17 classes uses a Transfer Learning model that combines DenseNet121 and ResNet50(DN121-RN50). The third model combines the DN121-RN50 model with a custom ANN(Artificial Neural Network) layers and performs on a modified Dataset2 which classifies it into 12 classes. The fourth model works on Dataset1 and uses DenseNet121 Transfer Learning model with RMSProp optimizer. The Fifth model works on Dataset1 with a concatenated model of DenseNet121 – VGG16 with RMSProp optimizer.

2.2.1 Dataset

The dataset provided by SmartBridge contained 17 classes which includes, Aloperia areata, Beau's Lines, Bluish nails, Clubbing, Darrier's disease, Eczema, Lindsay's nails, Koilonychia, Leukonychia, Muehrcke's Lines, Onycholysis, Pale nails, Red Lunula, Splinter Hemmorrage, Terry's nail, White nails, and Yellow nails. In our research we have named this dataset as Dataset1. This dataset is used to evaluate our Model1(VGG-16 Transfer Learning model) and Model2(DN121-RN50).

However, during our research, we observed that though the dataset performed well on both the models, the models could not test most of the nail diseases accurately and was either overfitting or underfitting.

Thus, we further classified the diseases that have similar visual characteristic into one class to overcome this problem. We finally classified the Dataset2 into 12 classes which includes,

AlopeciaAreata_DariersDisease, BeausLines, BluishNail, HalfMoonRedBands,
HealthyNail, Koilonychia, Leukonychia, Onychogryphosis,
Onycholycis_NailPsoriasis_Onychomycosis,
SplinterHemmorrage_AcralLentiginousMelanoma,
TerrysNail_WhiteNails_LindsaysNails_PaleNails, YellowNails

This dataset contains an addition class called 'HealthyNail', which also informs the patient that their nails are healthy too. We have made these classifications based on the visual characteristics of the diseases.

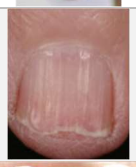
Say, in most cases it is found that the visual characters of the nail diseases are similar for Alopecia Areata and Darier's Disease. Images Fig.1.a and Fig.1.b describes the same.

Fig1.a. Alopecia areata

Fig1.b. Darier's Disease

Hence the below tabular column gives an idea to the visual characteristics of the nail diseases that are aimed to be diagnosed

Table1. Nail Disorders included in the Dataset

Si.No	Nail Disorder Name	Figure
1	Alopecia Areata	
2	Beau's Lines	
3	Bluish Nails	
4	Clubbing	
5	Darier's disease	
6	Half Moon Red Bands	
7	Healthy Nails	

8	Koilonychia	
9	Leukonychia	
10	Lindsay's Nail	
11	Muehrcke's Lines	
12	Onychogryphosis	
13	Onycholycis	
14	Nail Psoriasis	
15	Onychomycosis	
16	Splinter Hemorrhage	
17	Terry's Nails	
18	White Nails	
19	Yellow nails	

Note: The above table shows the nail disorders that our research as a whole that is used to solve and is not the number of classes used in the research

3. Theoretical analysis

3.1. Block diagram

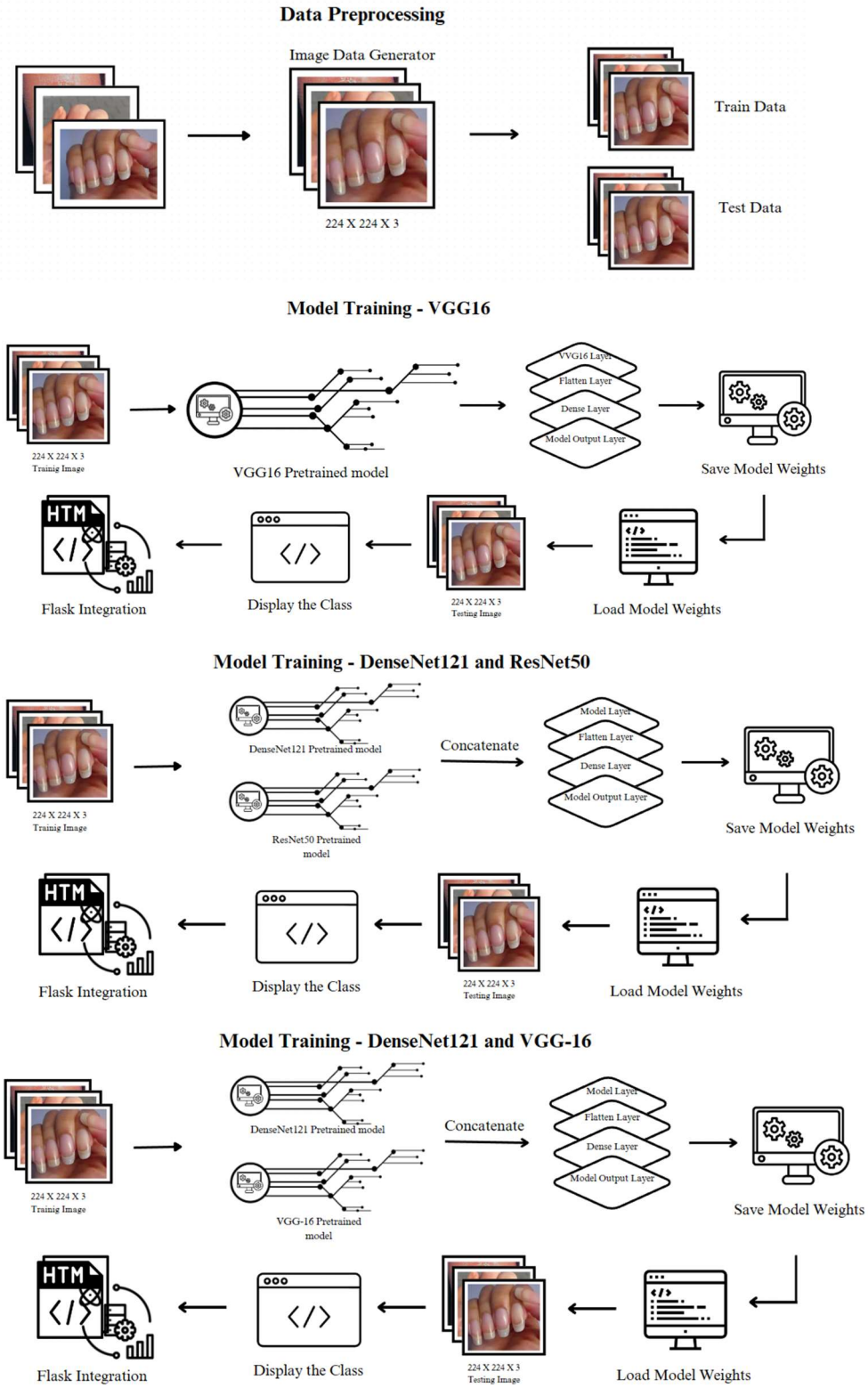


Fig2. Diagrammatic Representation of the models – Model 1 and Model 2

The above Fig2 represents the complete model integration, starting from the data preprocessing, where all the images are made into a uniform input size of 244x244x3.
The model is then load into VGG16 architecture, a pretrained CNN – Transfer Learning model trained on ImageNet dataset. The dataset used for this model is Dataset1 which was provided by SmartBridge. The model1 was trained for 10 epochs with batch_size as 22. The model was saved and was further integrated to Flask for the web application.
The second model was a custom Transfer Learning model with DenseNet121 and ResNet50 concatenated layers as an input layer. DenseNet121 and ResNet50 are also pretrained Transfer Learning model trained on ImageNet dataset. The model2 was trained for 100 epoch with batch size as 22 and learning rate as 0.0001.

3.2. Hardware/ Software Designing

Table2. Hardware and Software specifications

Hardware and Software Requirements	Characteristics
Processor	AMD Ryzen 7 5800H with Radeon Graphics Intel(R) Core™ i7-10750H CPU @ 2.60GHz 2.59 GHz
OS System	Microsoft Windows 11
Display system	NAVIDIA GeForce GTX 1060 – AMD Radeon™ Graphics
RAM	16 GB
Software apps	Google Colab, Microsoft Visual Studio Code (User)
Language	Python, HTML, Java Script, CSS

Systems that meet the previously mentioned machine characteristics – Table2 have been employed to develop this project in this research. We have employed the Python programming language to further enhance the AI-based deep neural networks. Using the Tensorflow Keras libraries, we implemented the project using Google Co-Lab and Microsoft Visual Studio Code (User Version). We used pre-templates from GitHub to create the web application's user interface. These templates were implemented using HTML, Java script, and CSS. The model and web application were integrated using the Flask application.

4. Experimental Analysis

In our research we have trained our dataset with the below model specifications in Table3

Table3. Model specifications – Comparisons

Si.no	Model name	Dataset Classes	Optimizer	Epochs	Batch size	Learning Rate	Results
1	VGG-16	Dataset1 – 17 class	Adam	100	22	-	loss: 0.2905 - accuracy: 0.9500 - val_loss: 0.1709 - val_accuracy: 0.9848
2	DenseNet121-Resnet50	Dataset1 – 17 class	Adam	100	22	0.0001	loss: 0.3449 - accuracy: 0.9634 - val_loss: 0.2680 - val_accuracy: 0.9836

3	DenseNet121-Resnet50 – Custom ANN layers	Dataset2 – 12 Class	Adam	99	15	0.01	loss: 2.5001 - accuracy: 0.7217 - val_loss: 2.6986 - val_accuracy: 0.7042
4	DenseNet121	Dataset1 – 17 class	RMSProp	40	8	0.001	loss: 0.2738 - accuracy: 0.9420 - val_loss: 0.3216 - val_accuracy: 0.9290
5	DenseNet121-VGG16	Dataset1 – 17 class	RMSProp	30	8	0.001	loss: 0.0496 - accuracy: 0.9908 - val_loss: 0.1039 - val_accuracy: 0.9836

Thus from the above observations it is clear that the model DenseNet121-VGG16 with RMSProp optimizer gave the best results with an average accuracy of 98.72% and a loss difference of about 0.0543.

Below Fig3. Shows the plot between training loss and validation loss of the above models

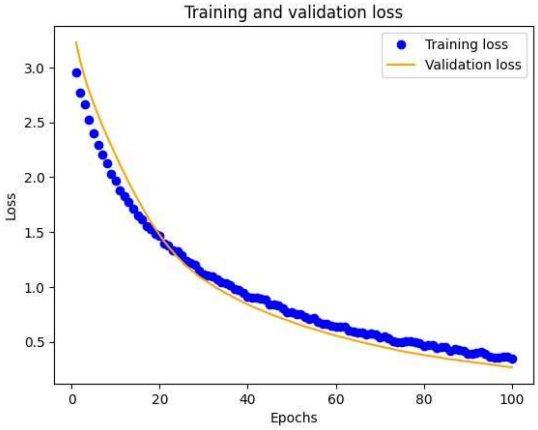


Fig3.a. DenseNet121 - RestNet50: Adam Optimizer – Dataset1

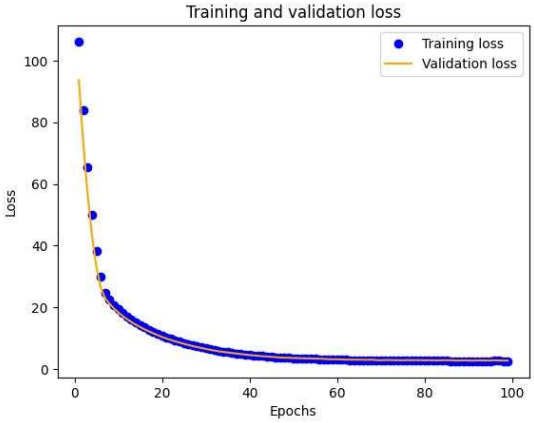


Fig3.b. DenseNet121 – ResNet50: Adam – Dataset2

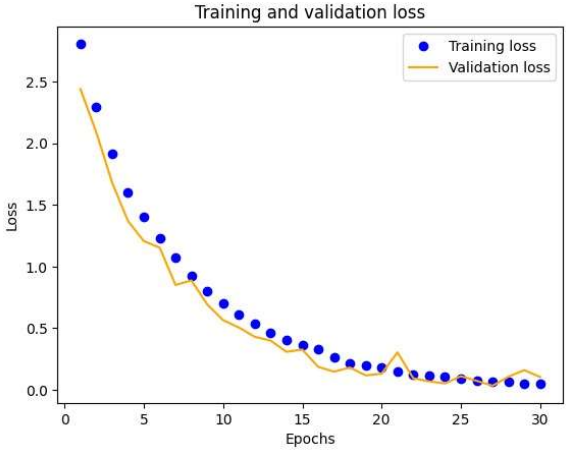


Fig3.c. DenseNet121 - VGG16: RMSProp – Dataset1

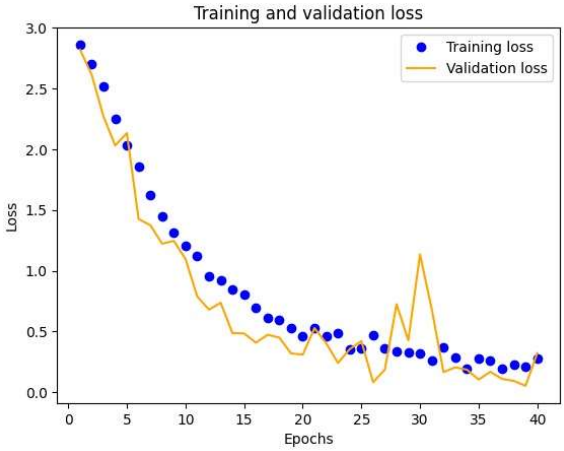
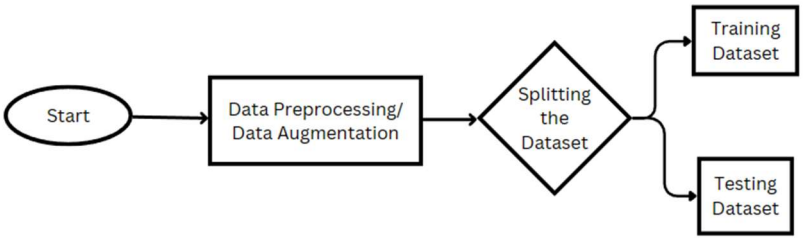


Fig3.d DenseNet121: RMSProp – Dataset1

5. Flowchart



The above Fig4. Describes our project flow. In this research we initiated by pre-process the data (Data Augmentation) where we resized the image to 224 x 224 x 3 which is the input size for VGG16 Transfer Learning model. The input size for DenseNet121 is minimum of 29 x 29. To maintain uniformity, the input size has been taken as 224 x 224 x 3 throughout the research.

The dataset has been divided into Train dataset and Test Dataset. The training dataset is then loaded into the model and the model weights are saved as h5 file format. The testing dataset is then tested using the saved model.

Once the model is tested, the saved model is then integrated to an user interface. Flask application is used for the integration.

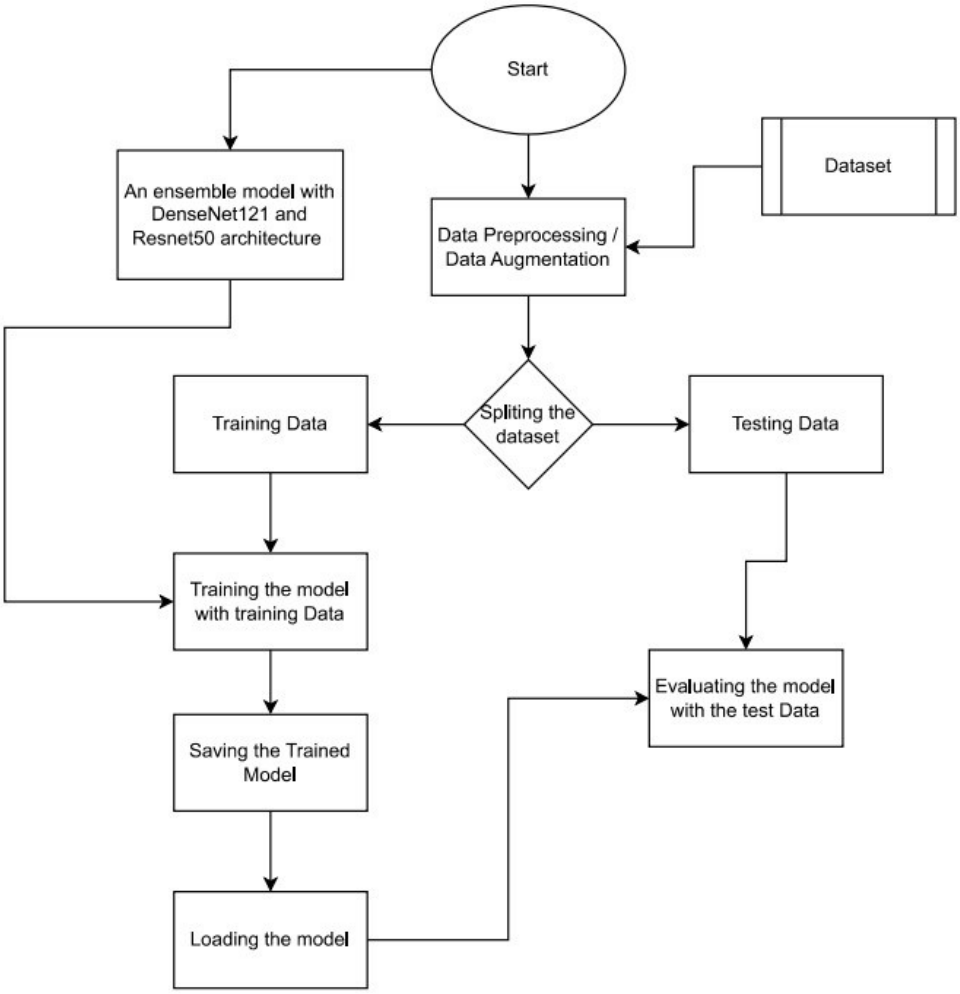


Fig4. Model flow – chart

6. Results

Table3. Model Experiment Results

Si.no	Model name	Dataset Classes	Optimizer	Epochs	Batch size	Learning Rate	Results
1	VGG-16	Dataset1 – 17 class	Adam	100	22	-	loss: 0.2905 - accuracy: 0.9500 - val_loss: 0.1709 - val_accuracy: 0.9848
2	DenseNet121-Resnet50	Dataset1 – 17 class	Adam	100	22	0.0001	loss: 0.3449 - accuracy: 0.9634 - val_loss: 0.2680 - val_accuracy: 0.9836
3	DenseNet121-Resnet50 – Custom ANN layers	Dataset2 – 12 Class	Adam	99	15	0.01	loss: 2.5001 - accuracy: 0.7217 - val_loss: 2.6986 - val_accuracy: 0.7042
4	DenseNet121	Dataset1 – 17 class	RMSProp	40	8	0.001	loss: 0.2738 - accuracy: 0.9420 - val_loss: 0.3216 - val_accuracy: 0.9290
5	DenseNet121-VGG16	Dataset1 – 17 class	RMSProp	30	8	0.001	loss: 0.0496 - accuracy: 0.9908 - val_loss: 0.1039 - val_accuracy: 0.9836

From the above, The DenseNet121 – VGG16 model when trained with Dataset1, for 30 epochs, batch size of 8 and learning rate of 0.001 gave the best accuracy of 98.72% and an average minimum loss of about 0.0543. The second-best model that gave the best results with minimum loss is DenseNet121. This model was again trained with Dataset1 with 40 epochs, batch size of 8, with a learning rate of 0.01. The Train loss vs testing loss has been plotted and is shown in Fig5.

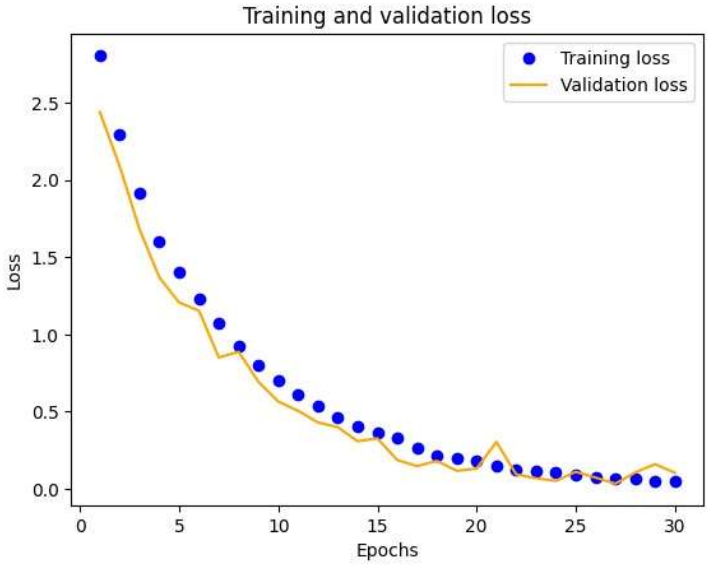


Fig5.DenseNet121 - VGG16: RMSProp – Dataset1

We further implement this model with a Web application us Flask. The model weights were used to analyze and predict the disease of the user input

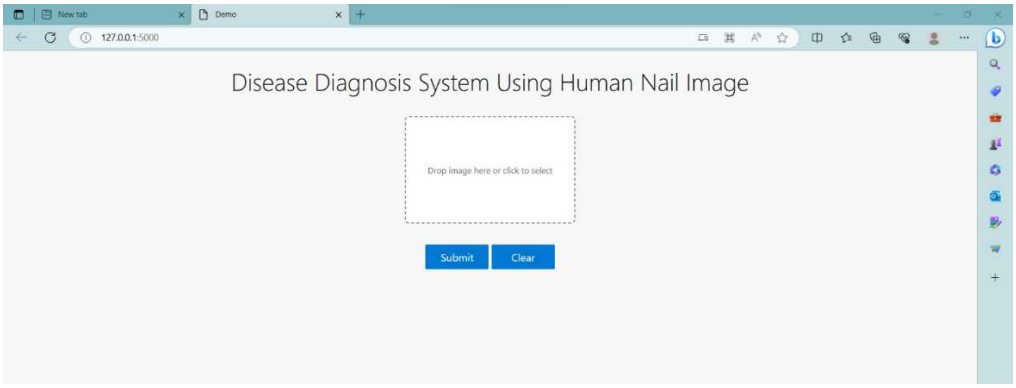


Fig6. Web user interface – User will upload the image in this portal

Fig6. Shows the user interface portal which can be used to upload the image that need to be tested and analyzed.

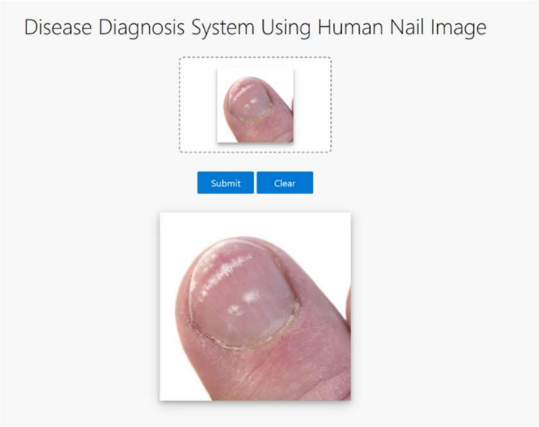


Fig7.a. Testing image – Leukonychia class

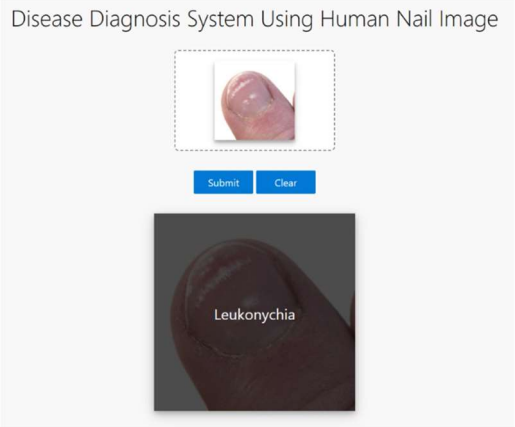


Fig7.b. Tested as Leukonychia Class

Fig7.c. Input image tested as White nail

Fig7.d. Input image tested as Yellow nail

Fig7.a, Fig7.b, Fig7.c, Fig7.d are the experimental results where the input images are being tested in a web-framework using Flask deployment.

7. Advantages and Disadvantages

The research's suggested model is more accurate and performs consistently with the dataset. Our goal was to provide a better user interface that would better estimate the image and assist users with comparable visual features in understanding any potential nail disorders. The majority of nail problems that are frequently observed and have a higher likelihood of signaling any serious medical illnesses requiring urgent care have been reviewed in this study. The main highlight of this research is that it has as its goal to cover the most common medical illnesses and nail abnormalities.

However, despite producing higher accuracies, the model at times tend to produce wrong results where there are very similar visual characters for two different nail disorders. In DN121RN50-Custom ANN model i.e. Model3 which was trained with dataset2 had an average accuracy of only 70% which can be further improved.

8. Applications

Our first goal in doing this research was to learn more about the nail anomalies and the related nail disorders that they were indicative of.

It is well known that several nail irregularities and disorders, such as dark streaks in the nail, can be a sign of melanoma susceptibility. Paronychia, or redness and swelling around the nail, may be a sign of a serious skin illness caused by a fungus. Most people with diseases like alopecia areata, which causes significant hair loss, also have nail signs. Alopecia areata is frequently indicated by nail dents or pitting. Pitted nails can also be a sign of conditions like nail psoriasis and atopic dermatitis. Additionally, studies have shown that serious pulmonary disorders can also be indicated by anomalies in the nails. Lung illnesses and rheumatoid arthritis are indicated by thickening of the nails and by yellow nails that eventually stop developing. Deep grooves called Beau's Lines run the length of our nails. They show when nail development begins to stop and our body's metabolism has slowed. Onychomadesis, a medical disease, can also be indicated by Beau's Lines. A lack of iron in our body is indicated by koilonychia, also referred to as spoon-shaped nails. When our body metabolism experiences circulation issues or when we have disorders like nail psoriasis or ichthyosis, nails frequently thicken and overgrow. This nail condition is often referred to as Ram's horn nails or Onychogryphosis. When the nails continue to curve downward, clubbing or curled nails frequently happen. Fingertips frequently expand, and when squeezed, the nails begin to feel spongy. These are the symptoms of liver, stomach, heart, and pulmonary disorders. Color changes in nails are also indicative of nail disorder or health problems. Table3 indicates the color change and the respective nail disorder.

Table4. Nail color changes and corresponding disorders

COLOR	NAIL DISORDER OR HEALTH PROBLEM
Blue nails	Not enough oxygen in our blood stream
White nails	Liver disease, diabetes
Blue half-moons	Poisoning
Yellow nails	Lung diseases, nail infection
Pale nails	Anaemia
Half-pink and Half-white nails	Kidney diseases
Dusky Red half-moons	Could be Lupus, heart diseases, alopecia areata, arthritis, dermatomyositis

9. Conclusion

This research has used Five pre-trained Transfer Learning models on two different dataset. The first dataset – Dataset1 is a multiclass dataset with 17 classes. The Dataset2 is a multiclass dataset with 12 classes. Among all the models used in this research four model were trained on Dataset1 and one model was trained with Dataset2. The four Transfer Learning models that were trained on Dataset1 are: VGG16, ResNet121-ResNet50 hybrid, DenseNet121, DenseNet121-VGG16. The first two models were trained used Adam optimizer and the next two models were trained using RMSProp optimizer. The last model that was trained on Dataset2 uses a Transfer Learning model combining DenseNet121 and ResNet50 with Adam optimizer. Among the five Transfer Learning model used in this research, the model5(in Table3) performing with a concatenation of DenseNet121-VGG16 and RMSProp optimizer on Dataset1 (Dataset given by SmartBridge) gave better results with an accuracy of 98.72% and an average minimum loss of about 0.0543.

10. Future Scope

The user interface used does not store the data of the user, where they can refer it for their future consultations. The approach can be implemented on a big scale using a variety of web applications and enough database systems. We can also integrate an interactive medical-voice assistant to the user interface which can enable them interact and clear their doubts on their disease. Also the web result can give a description about the predicted disease.

11. Bibliography

a. https://www.researchgate.net/publication/333679828_Nail_Feature_Analysis_and_Classification_Techniques_for_Disease_Detection
b. https://www.researchgate.net/publication/305001583_Early_Stage_Disease_Diagnosis_System_Using_Human_Nail_Image_Processing
c. https://www.researchgate.net/publication/321426036_A_Review_Fingernail_Images_for_Disease_Detection
d. https://www.researchgate.net/publication/360994776_Nail_Disease_Detection_and_Classification_Using_Deep_Learning
e. https://www.researchgate.net/publication/360086466_Melanoma_of_nail_unit
f. https://www.researchgate.net/publication/355973020_HUMAN_NAIL_DISEASES_CLASSIFICATION_BASED_ON_TRANSFER_LEARNING
g. https://www.researchgate.net/publication/362768598_Classification_of_Koilonychia_Beaus_Lines_and_Leukonychia_based_on_Nail_Image_using_Transfer_Learning_VGG-16
h. https://www.researchgate.net/publication/371170940_CNN_Based_Disease_Identification_using_Human_Nail_Images
i. https://www.researchgate.net/publication/369401238_DETECTION_OF_NAIL_DISEASES_USING_ENSEMBLE_MODEL_BASED_ON_MAJORITY_VOTING_-_COGUNLUK_OYUNA_DAYALI_TOPLULUK_MODELI_ILE_TIRNAK_HASTALIKLARININ_TESPITI
j. https://ijesc.org/upload/524d539e54bc9e9cca328cd4007862ec.Detection%20of%20Diseases%20using%20Nail%20Image%20Processing%20Based%20on%20Multiclass%20SVM%20Classifier%20Method.pdf
k. https://www.researchgate.net/publication/362950780_Deep_Learning_Based_Classification_of_Human_Nail_Diseases_Using_Color_Nail_Images

APPENDIX

a. Model Definition for Final model – DenseNet121-VGG16, RMSProp optimizer

```
# Define your DenseNet-VGG16 hybrid model architecture
def DenseNetVGG16(num_classes):
    # Load pre-trained DenseNet and VGG16 models
    densenet = DenseNet121(weights='imagenet', include_top=False, input_shape=(224, 224, 3))
    vgg16 = VGG16(weights='imagenet', include_top=False, input_shape=(224, 224, 3))

    # Freeze the pre-trained layers
    for layer in densenet.layers:
        layer.trainable = False
    for layer in vgg16.layers:
        layer.trainable = False

    # Combine DenseNet and VGG16 models
    input_layer = tf.keras.Input(shape=(224, 224, 3))
    densenet_output = densenet(input_layer)
    vgg16_output = vgg16(input_layer)

    # Add global average pooling layer
    densenet_output = GlobalAveragePooling2D()(densenet_output)
    vgg16_output = GlobalAveragePooling2D()(vgg16_output)

    # Concatenate DenseNet and VGG16 outputs
    combined_output = tf.keras.layers.concatenate([densenet_output, vgg16_output])

    # Add a fully connected layer
    dense = Dense(units=128, activation='relu')(combined_output)

    # Add output layer
    output = Dense(units=num_classes, activation='softmax')(dense)

    # Create the model
    model = Model(inputs=input_layer, outputs=output)

    return model
```