A COMPARATIVE ANALYSIS OF DEEP LEARNING MODELS TO PREDICT DERMATOLOGICAL DISORDER

Prakash.J
Department of Computer Science & Engineering
PSG College of Technology, Coimbatore, Tamil Nadu, India
jpk.cse@psgtech.ac.in

Vinoth Kumar.B
Department of Information & Technology
PSG College of Technology, Coimbatore, Tamil Nadu, India
bvk.it@psgtech.ac.in

Shyam Ganesh.C.R
Department of Computer Science & Engineering
PSG College of Technology, Coimbatore, Tamil Nadu, India
shyamganesh1999@gmail.com

Abstract- Skin diseases have a significant impact on people's life and health. Despite being common its identification is very troublesome due to its complexities of skin tone, color and hair. This chapter analyses the feasibility for using existing convolutional neural network architectures to diagnose such diseases at its earlier stages prior to consulting a dermatologist. This chapter aims to compare the existing image recognition models like Inception v3, Inception Resnet v2 and VGG -19 for diagnosing skin diseases. The reason for comparing the above three models is to examine the impact of the depth and architecture of a model on performance in this domain. These models were trained using different strategies in order to determine the best one which suits the given problem for the available dataset. Also, hyperparameter optimization is done to get the best results possible. The Dermnet dataset which contains about 23000 images belonging to 23 classes for training is used. The output of the trained models is the identified disease. Our system can achieve as high as 68% test accuracy for Top-1 prediction.

Keywords – Skin disease diagnosis, Convolutional neural networks, transfer learning, Deep learning

I. INTRODUCTION

Every year, millions of people are diagnosed with several thousand classes of skin diseases. Most of these diseases are temporary, but some are permanent and can also be fatal if left unnoticed. The most common cancer worldwide is skin cancer. Melanoma, which is a type of skin cancer where the pigment producing cells mutate and begin to divide uncontrollably, accounts for only 1% of skin cancers but is most fatal. The diagnosis and treatment of skin diseases should be done as soon as possible.

Due to ignorance, these diseases are left untreated which might lead to serious effects. This ignorance has grown a lot due to the indistinguishability between normal infections and several serious diseases. Thus, as a first step of diagnosis, computer vision-based techniques can be used to preliminarily detect how serious an infection is, then according to the results the patient can be either given medication or visit a doctor accordingly.

During traditional diagnosis of a skin disease by a doctor, the affected part of patient's skin is examined for spots or eruptions color and texture of the affected areas. This observation is often combined with other details such as patient's age, medical history or other symptoms reported to diagnose the disease. This method varies according to the dermatologist's level of expertise. Some uncommon skin conditions are difficult to diagnose even by highly experienced dermatologists. Thus, it is necessary to build automated methods which preliminarily identifies the class of skin disease and can also be used by the doctors to verify their diagnosis.

Literature reveals that soft computing techniques are used in many applications such as image processing [21-32], Anomaly detection [33], Prediction [34-39]. Therefore, a suitable soft computing technique can be used for this application. Based on the skin symptoms, separate methods can be engineered for each disease. A set of features was found, which primarily identifies a particular disease and try to separate such features from a given image to predict the presence of that skin condition using techniques like segmentation. However, segmentation lacks scalability as separate methods cannot be devised for each and every skin condition. To build a universal method to classify many skin diseases, deep learning can be used.
Convolutional neural networks (CNN) is used to analyze visual imagery. With the help of a sufficiently large dataset, CNN will able to train and classify a disease if the model is able to generalize properly. Instead of creating our own CNN architecture, A well-known models like Inception v3, Inception Resnet v2 and VGG19 are used. These networks were trained using the ImageNet dataset and the results show that these models can compete with humans for object classification. This chapter analyses the feasibility for using existing convolutional neural network architectures to diagnose such diseases at its earlier stages prior to consulting a dermatologist. Thus, a comparative analysis on such image recognition models for skin disease diagnosis was performed.

II. Related Work

Nisha Yadav Et al. [3] studied the possibility to apply image processing techniques like segmentation on plant diseases and applies it to human skin disease recognition. The input is a picture or video, a photograph of the pathologic skin and output is additionally another image having same characteristics as input image. Principally, Image processing models take input samples as 2-D signals and at that time they apply fastened signal process strategies to them. the total project had the below major components. Image segmentation and have extraction, disease declaration and medical recommendation. a picture is taken and noise is removed by applying filters. Then the image is metameric to extract important info and so the unwellness is assessed by mistreatment applicable classifier.
Delia-Maria and Adriana proposed a diagnosis system having two distinct parts:

1. Network development which comprises of determining the input features, data collection, preprocessing and finalizing architecture. Architecture refers to the number of layers, weights, transfer functions etc. The learning algorithm is established and the network is trained. Finally, the accuracy is determined. This step is performed at the beginning and only once.
2. Diagnosis for a new patient – it is made each time a patient should be diagnosed. Here, the required features of the patient are selected and passed to the neural network. Once the diagnosis is done, it is evaluated by a human physician.

This paper used simple neural networks (ANN with 33 input units and one hidden layer) which predicts the skin disease only using the symptoms and not the image of the disease. But the accuracies achieved are fairly high.
Andre Esteva Et al. [2] proposed a CNN solution to diagnose carcinoma employing a dataset of 129,450 clinical images. It additionally tested its performance against twenty-one board-certified dermatologists on biopsy-proven clinical pictures with 2 essential binary classification use cases. The performance of CNN is on par with all tested specialists across each task. This shows that neural networks are capable of classifying skin cancerwith tier of competency equivalent to dermatologists. Thus, Inception v3 used in this paper seems to suit skin disease classification problems.
Brinker used CNN for both types of skin disease classifications such as lesion and disease targeted. For the disease-targeted diagnosis, a multi-class CNN classifier was trained, and multi-label CNN classifier was trained for the later. A new fully connected layer was replaced instead of the existing one. The model was trained using both strategies such as fine tuning and training from scratch. It is found that, for this scenario, fine-tuned works better than those trained from scratch. As investigated in this paper, this chapter will also perform all three strategies to train Inception models.
H.Liao [5] took dataset from OLE and Dermnet. In transfer learning, this paper took a pretrained network and fine-tuned its , instead of training from randomly initialization. This paper used Caffe framework to effectively train the models. VGG16, VGG19 and GoogleNet are the pretrained models used here. These CNN models can achieve high Top-5 accuracy when testing on the Dermnet dataset. This chapter will seek to improve on these models for accuracy.
Doaa A. Shoieb Et al. [4] modelled this problem in three stages. The image is first captured and segmented to locate the infection. This is then inputted to the feature extractor. These features are finally fed to a classifier. The segmentation methods used were region-based segmentation and texture segmentation for locating skin lesion. High-Level Intuitive Feature extraction was also used for feature extraction.

III. Dataset

The dataset used here is Dermnet. Dermnet provides information on a wide variety of skin conditions through innovative media. It has about 23,000 images on various skin conditions labelled in two hierarchical levels. It has 23 broad classes and more than 600 bottom level skin conditions. The dataset was downloaded from internet via a python script which raised HTTP requests as suggested in [5]. The server often stopped responding as continuous requests were given (DOS attack was detected). Thus, the requests were separated by considerable time intervals. The images were watermarked as it is the free version. Figure 1 represents the distribution of number of images per top level disease. To keep the experiment simple, the bottom level diseases were not considered for classification and only the top level 23 diseases were taken into consideration. Also, the distribution of images among these 23 classes are uneven. In order to normalize, the images were duplicated by distorting and rotating randomly for diseases with lesser images. The available dataset Dermnet, is fairly large but not large enough to train a model from scratch. So, transfer learning is used. We take pretrained CNNs like Inception v3, Inception Resnet v2 and vgg19 which were trained on ImageNet dataset and perform feature extraction and fine tuning which are two strategies and in the third, the entire model is trained from scratch.

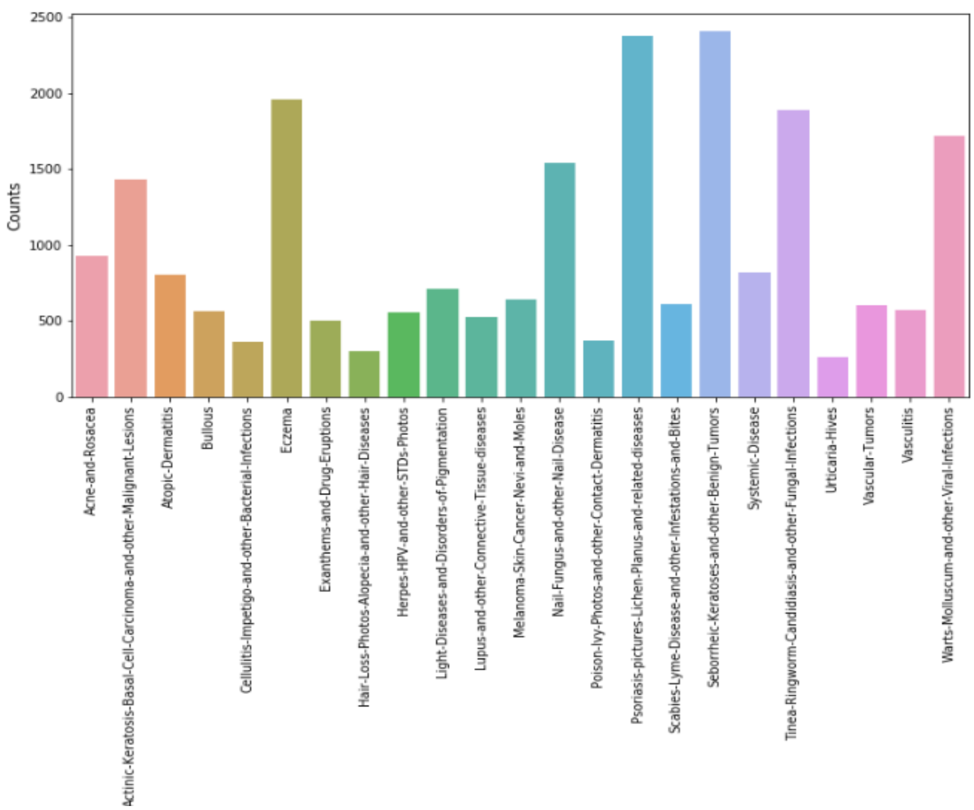

Figure 1. Dataset with 23 classes of diseases

IV. Deep Learning Models

4.1. Inception V3

The first model used for training is the Inception v3. Before training all the models, we should finalize the strategy (Fine tuning, feature extraction or end to end training) which will be used and the configuration for training. So, we test the first model with all three strategies and different configurations to determine the best. This will be applied on all models.

4.1.1. Feature Extraction

In this part of training, we use the existing weights of the Inception model, add a dense layer at the top which predicts the class label of the disease, and train the top layer only leaving the rest of the weights frozen. This actually makes use of the weights learnt by it while trained with ImageNet dataset. Thus, feature extraction process is effectively retraining the model to classify skin diseases purely based on its understanding of extracting features from images it was trained earlier with. This feature extraction will be fully efficient only when the domain of the images it was trained with is similar to the domain of the dataset we are using. But ImageNet is not a dataset which has various classes based on recognizing skin patterns. Thus, the accuracy of this phase of training is expected to be very low as seen in Figures 2 and 3. We did hyperparameter optimization and the best possible accuracy was achieved at a learning rate of 0.001 and 50 epochs.

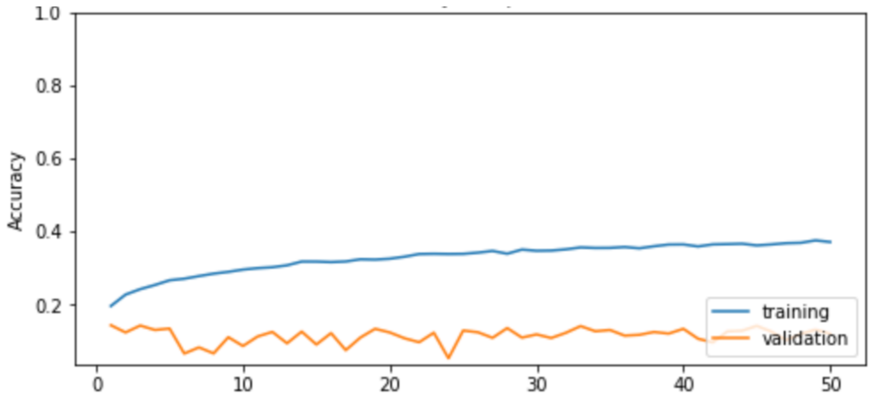

Figure 2. Accuracy comparison for feature extraction.

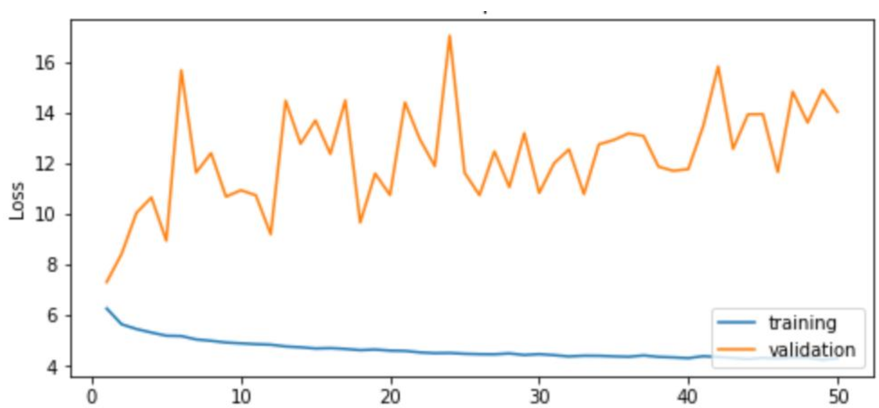

Figure 3. Loss comparison for feature extraction.

4.1.2. Fine Tune

To improve performance even more, we repurpose the top-level layers by fine-tuning with Dermnet dataset. That is, we tune our weights such that we learn high-level features like skin texture, tone etcspecific to our dataset. The Inception v3 model has a total of 311 layers, in which we will train the layers 250-311. This fine tuning will improve feature selection based on our dataset.

Even though some of the inner layers were trained again, the validation accuracy did not see any improvement. This is because of the fact that the first 100-150 layers have weights adjusted to identify real world objects only and not texture and patterns in skin. Thus, the model overfits the data by memorizing the features of training data. This implies the high training accuracy (93%) but very low validation accuracy (~25%) at the end of 25 epochs.

The only solution to this problem will be training the same Inception model without initializing the weights it learnt from ImageNet dataset. By training from scratch, there is a high possibility of improvement of the model’s ability to generalize.

4.1.3. Training from Scratch

The entire model is made trainable not by initializing the model with the weights learnt from ImageNet rather by giving random weights and unfreezing the entire model. The model is now trained for 38 epochs to get a training accuracy of 78% and a validation accuracy of ~68%. It is evident that the model has generalized well since it is trained from scratch as evident from the accuracy and loss curves in figure 4.

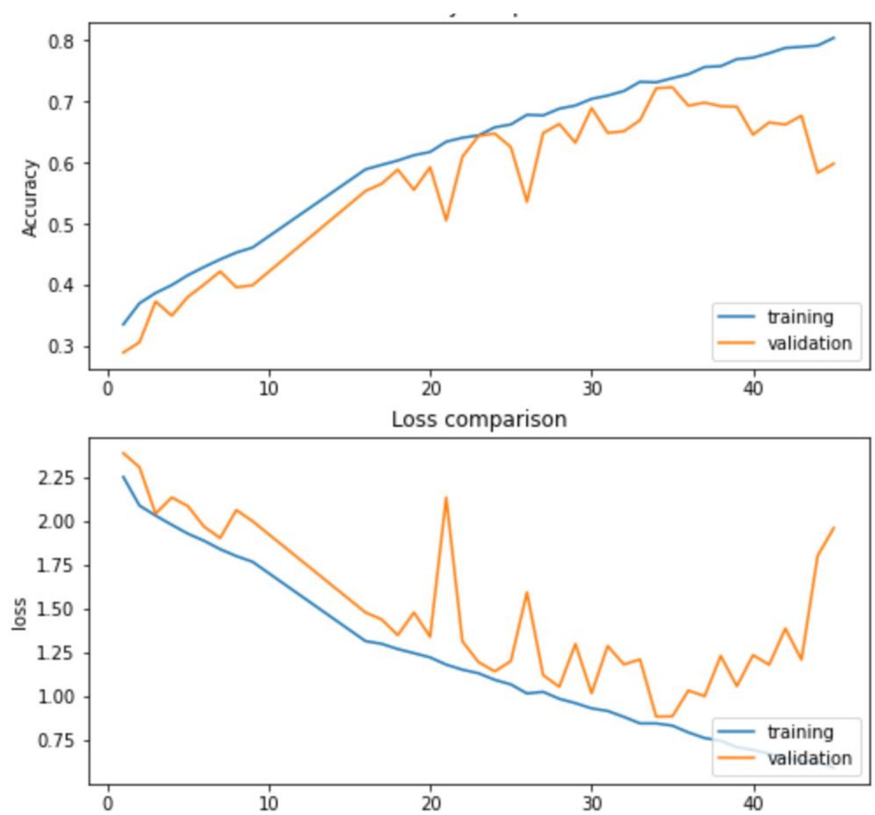

Figure 4. Training and validation accuracy and loss for training from scratch.

4.1.4. Inference

The first strategy has a very low accuracy because the weights of Inception v3 was reused and only the top layer was trained. Since ImageNet and DermNet are datasets of two different domains, a very low accuracy is obtained as expected.

The second strategy involved unfreezing some layers and the weights were modified according to Dermnet dataset. Since the lower layers' weight were untouched, the model overfitted to the available training data. This explains the comparatively low validation accuracy.

The third strategy involved training the entire model from scratch. This resulted in better results because the weights of ImageNet weren't used as it was from a different domain. Thus, this strategy proves to be the best possible one which can give the best accuracy possible.

Table -1 Comparison of accuracies for three strategies.

	Training accuracy (%)	**Validation Accuracy (%)**
Feature extraction	**43**	**14**
Fine tuning	**94**	**23**
Training from scratch	**80**	**68**

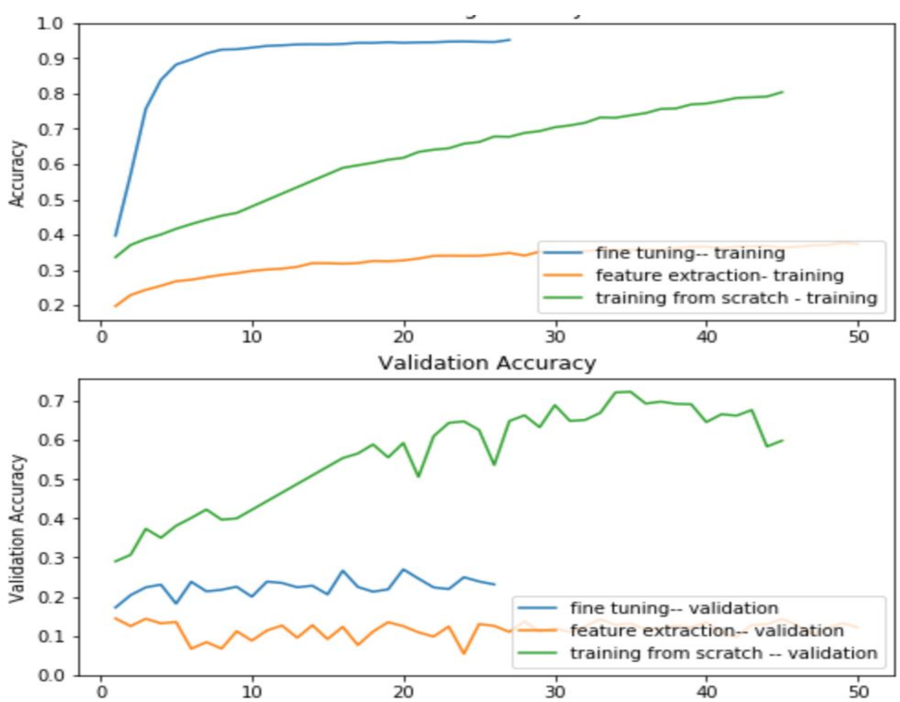

Figure 5. Comparison of Training and Validation accuracy for 3 strategies.

4.2. VGG19

According to the previous observations from InceptionV3, the weights from ImageNet dataset has no correspondence to dermnet. Thus, results will be very low when feature extraction or fine tuning is performed. So, we directly train the entire model from scratch and observe the accuracy. For testing purposes, we stop after 50 epochs. The learning rate used was 0.01. But at the end of 50 epochs, the maximum accuracy achieved was 25%.

The growth rate is reduced to almost zero after 40 epochs and there is no improvement in accuracy after that. This is because the architecture of VGG19 is less deep, thus only suitable for identifying objects (and not textures) belonging to lesser no. of classes.

4.3. INCEPTION RESNET V2

The Inception Resnet v2 showed a slow but steady increase in accuracy given its deep layers as seen in figure 10. The improvement in validation accuracy considerably slows down after a point from figure 11. This is due to the lower resolution of the images. The only possible way to improve it is to have a larger dataset with image of higher resolution.

V. RESULT ANALYSIS

The three models which were trained are tested using the same validation dataset and the results are compared below. We use a 4500-image dataset, use the models to predict the disease and visualize the predictions using a confusion matrix in heat map.

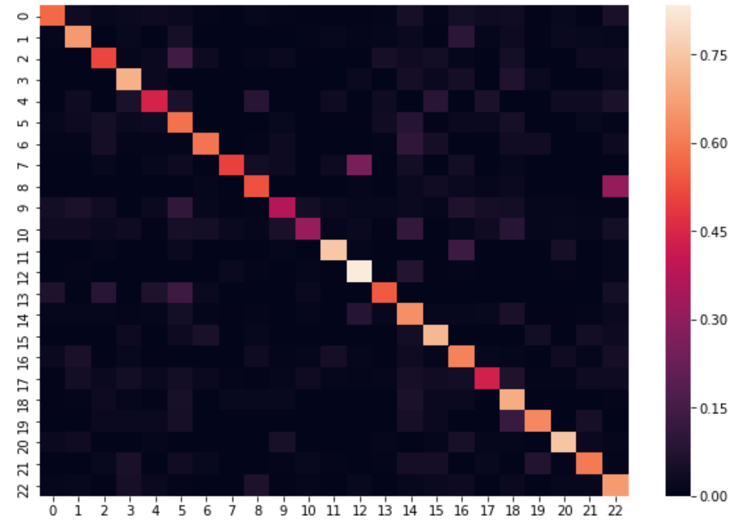

Figure 6. Confusion matrix for Inception V3.

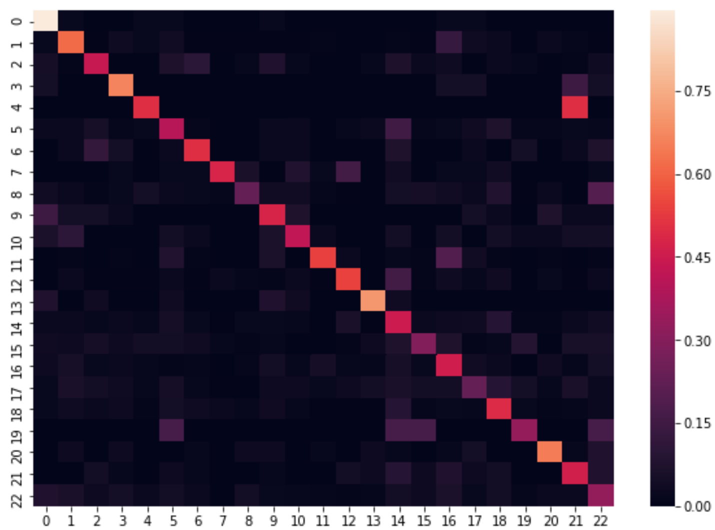

Figure 7. Confusion matrix for Inception Resnet v2.

ISSN No : 1006-7930

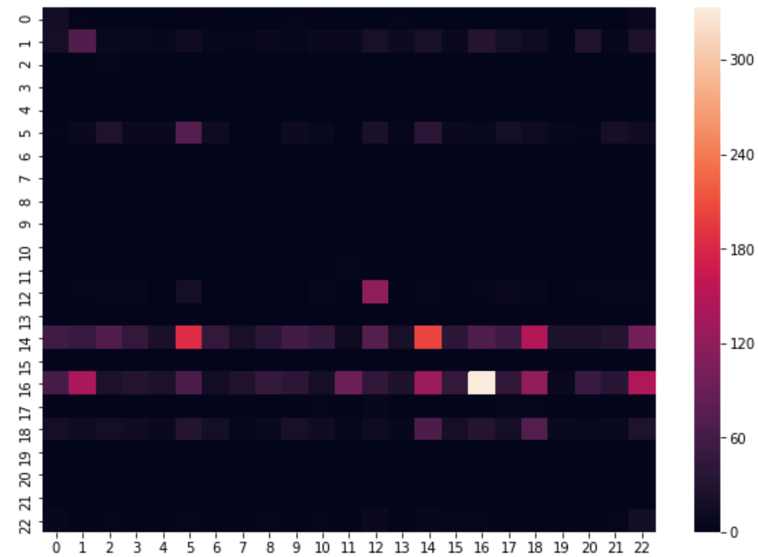

Figure 8. Confusion matrix for VGG-19.

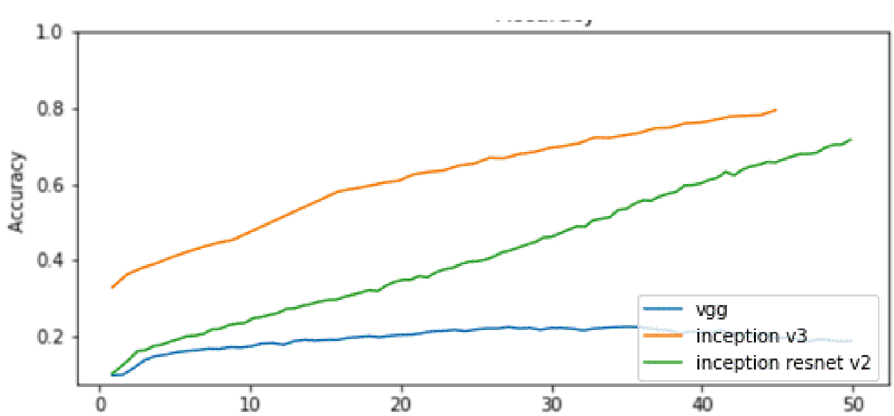

Figure 9. Training Accuracy.

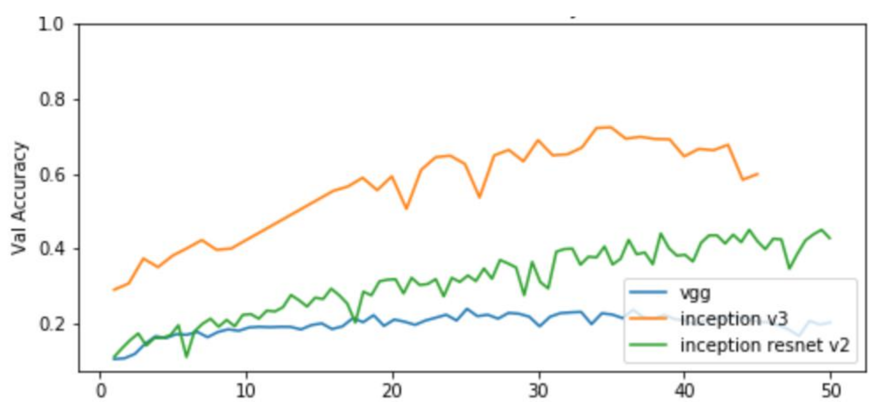

Figure 10. Validation Accuracy.

Table -2 Comparison of accuracies for three strategies.

	Training accuracy (%)	Validation Accuracy (%)
Feature extraction	**43**	**14**
Fine tuning	**94**	**23**
Training from scratch	**80**	**68**

Comparing the three models with respect to their improvement in accuracies, the inception v3, Resnet v2 are better suited for our problem than VGG-19. But we did not achieve 100% accuracy in either. By observing the confusion matrices, the errors in both models are in different diseases. The former has problems in classifying hair related diseases correctly, but the latter has better accuracies in that particular disease. Thus, we can build a new system by using the outputs of both models to improve the accuracies. We get an accuracy of 61% Top-1 accuracy after combining the results. Though this is not a big improvement from inception v3, if more models are included, the results can improve.

VI.CONCLUSION

The possibility of using a CNN to diagnose skin diseases was extensively explored and it is found that it can do very good predictions based on the quality of dataset used. But for the same to replace or assist actual dermatologists, it needs to be trained on a larger and diverse dataset on more granular classes of skin diseases and if it is cross verified by a dermatologist during training, we might be able to achieve a model which has the ability to correctly identify diseases. All the CNN models available today is very capable of doing any task. The only big deciding factor is the dataset it is trained on and the resources available. Our future improvements to this proposal will be the acquirement of such diverse and large dataset. Also, we will be trying to rebuild the architectures of the pretrained models and examine ways to improve the capacity. The classification accuracies for different diseases in each model is observed and predictions of different models can be combined to improve this accuracy.

REFERENCES

[1] Haofu Liao, Yuncheng Li, & Jiebo Luo. (2016). Skin disease classification versus skin lesion characterization: Achieving robust diagnosis using multi-label deep neural networks. *2016 23rd International Conference on Pattern Recognition (ICPR)*. https://doi.org/10.1109/icpr.2016.7899659

[2] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115-118. https://doi.org/10.1038/nature21056

[3] Yadav, N., Kumar, V., & Shrivastava, U. (2016). Skin diseases detection models using image processing: A survey. *International Journal of Computer Applications*, *137*(12), 34-39. https://doi.org/10.5120/ijca2016909001

[4] Shoieb, D. A., Youssef, S. M., & Aly, W. M. (2016). Computer-aided model for skin diagnosis using deep learning. *Journal of Image and Graphics*, 122-129. https://doi.org/10.18178/joig.4.2.122-129

[5] Liao, H. (2015). A Deep Learning Approach to Universal Skin Disease Classification.

[6] Devi, I., Karpagam, G., & Kumar, B. V. (2017). A survey of machine learning techniques. *International Journal of Computational Systems Engineering*, *3*(4), 203. https://doi.org/10.1504/ijcsyse.2017.089191

[7] Rajesh, R., & Mathivanan, B. (2017). Predicting Flight Delay using ANN with Multicore Map reduce framework, Communication and Power Engineering. In *Communication and power engineering*. Walter de Gruyter GmbH & Co KG

[8] Brinker, T. J., Hekler, A., Utikal, J. S., Schadendorf, D., Berking, C., Enk, A. H., & Von Kalle, C. (2018). Skin cancer classification using Convolutional neural networks: Systematic review (Preprint). https://doi.org/10.2196/preprints.11936

[9] Arevalo, J., Cruz-Roa, A., Arias, V., Romero, E., & González, F. A. (2015). An unsupervised feature learning framework for basal cell carcinoma image analysis. *Artificial Intelligence in Medicine*, *64*(2), 131-145. https://doi.org/10.1016/j.artmed.2015.04.004

[10] Arroyo, J. L., & Zapirain, B. G. (2013). Automated detection of melanoma in Dermoscopic images. *Series in BioEngineering*, 139-192. https://doi.org/10.1007/978-3-642-39608-3_6

[11] Barata, C., Marques, J. S., & Mendonça, T. (2013). Bag-of-Features classification model for the diagnose of melanoma in Dermoscopy images using color and texture descriptors. *Lecture Notes in Computer Science*, 547-555. https://doi.org/10.1007/978-3-642-39094-4_62

[12] Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1798-1828. https://doi.org/10.1109/tpami.2013.50

[13] Chang, H., Zhou, Y., Borowsky, A., Barner, K., Spellman, P., & Parvin, B. (2014). Stacked predictive sparse decomposition for classification of histology sections. *International Journal of Computer Vision*, *113*(1), 3-18. https://doi.org/10.1007/s11263-014-0790-9

[14] Burns, T., Breathnach, S., Cox, N., & Griffiths, C. (2008). *Rook's textbook of dermatology*. John Wiley & Sons.

[15] Cruz-Roa, A., Basavanhally, A., González, F., Gilmore, H., Feldman, M., Ganesan, S., Shih, N., Tomaszewski, J., & Madabhushi, A. (2014). Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. *Medical Imaging 2014: Digital Pathology*. https://doi.org/10.1117/12.2043872

[16] Cruz-Roa, A. A., Arevalo Ovalle, J. E., Madabhushi, A., & González Osorio, F. A. (2013). A deep learning architecture for image representation, visual Interpretability and automated basal-cell carcinoma cancer detection. *Advanced Information Systems Engineering*, 403-410. https://doi.org/10.1007/978-3-642-40763-5_50

[17] Jia Deng, Wei Dong, Socher, R., Li-Jia Li, Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/cvprw.2009.5206848

[18] Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. https://doi.org/1502.03167

[19] G. K. Jana, A. Gupta, A. Das, R. Tripathy, and P. Sahoo.(2010), Herbal Treatment to Skin Diseases: A Global Approach. *Drug invention today.*

[20] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., & Darrell, T. (2014). Caffe. *Proceedings of the ACM International Conference on Multimedia - MM '14*. https://doi.org/10.1145/2647868.2654889

[21] BALASUBRAMANIAN, V. K., & MANAVALAN, K. (2016). A problem approximation surrogate model (PASM) for fitness approximation in optimizing the quantization table for the JPEG baseline algorithm. *TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES*, *24*, 4623-4636. https://doi.org/10.3906/elk-1501-108

[22] Viswajaa, S., Vinoth Kumar, B., & Karpagam, G. R. (2015). Survey on Nature Inspired Meta-Heuristics Algorithms in Optimizing the Quantization Tableforthe JPEG Baseline Algorithm. *International Advanced Research Journal in Science, Engineering and Technology*, *2*(4), 14-123.

[23] B.Vinoth Kumar and G.R.Karpagam, (2011). An Empirical Analysis of Requantization Errors for Recompressed JPEG Images, *International Journal of Engineering Science and Technology*, Vol. 3 No.12 2011, pp. 8519-8527.

[24] Naresh, S., Vinoth Kumar, B., & Karpagam, G. R. (2016). A literature review on quantization table design for the JPEG baseline algorithm. *International Journal Of Engineering And Computer Science*. https://doi.org/10.18535/ijecs/v4i10.24

[25] Vinoth Kumar, B., & Karpagam, G. R. (2019). A smart algorithm for quantization table optimization: A case study in JPEG compression. *Smart Techniques for a Smarter Planet*, 257-280. https://doi.org/10.1007/978-3-030-03131-2_13

[26] Vinoth Kumar, B., & Karpagam, G. (2018). Reduction of computation time in differential evolution-based quantisation table optimisation for the JPEG baseline algorithm. *International Journal of Computational Systems Engineering*, *4*(1), 58. https://doi.org/10.1504/ijcsyse.2018.10011810

[27] Kumar, B. V., Karpagam, G. R., & Naresh, S. P. (2016). Generation of JPEG quantization table using real coded quantum genetic algorithm. *2016 International Conference on Communication and Signal Processing (ICCSP)*. https://doi.org/10.1109/iccsp.2016.7754457

[28] Vinoth Kumar, B., & Karpagam, G. R. (2017). Single versus multiple trial vectors in classical differential evolution for optimizing the quantization table in JPEG baseline algorithm. *ICTACT Journal on Soft Computing*, *7*(4), 1510-1516. https://doi.org/10.21917/ijsc.2017.0210

[29] Vinoth Kumar, B., Karpagam, G., & Zhao, Y. (2019). Evolutionary algorithm with Memetic search capability for optic disc localization in retinal fundus images. *Intelligent Data Analysis for Biomedical Applications*, 191-207. https://doi.org/10.1016/b978-0-12-815553-0.00009-4

[30] Vinoth Kumar, B., Sabareeswaran, S., & Madumitha, G. (2019). A decennary survey on artificial intelligence methods for image segmentation. *Advanced Engineering Optimization Through Intelligent Techniques*, 291-311. https://doi.org/10.1007/978-981-13-8196-6_27

[31] Kumar, B. V., Jeneessha, P., & Nivethitha, M. (2020). A differential evolutionary algorithm for image segmentation of white blood cells in acute lymphoblastic leukaemia images. *2020 Fourth International Conference on Inventive Systems and Control (ICISC)*. https://doi.org/10.1109/icisc47916.2020.9171169

[32] Hemanth, D. J., Kumar, B. V., & Manavalan, G. R. (2019). *Recent advances on Memetic algorithms and its applications in image processing*. Springer Nature.

[33] R Sandhya, J Prakash, B Vinoth Kumar (2020). Comparative Analysis of Clustering Techniques in Anomaly Detection in Wind Turbine Data. *Journal of Xi'an University of Architecture & Technology*, Vol. 12 No.3 2020, pp. 5684-5694.

[34] Prakash J., & Vinoth Kumar B. (2021). Exploration of computational intelligence insights and data analytics to combat COVID-19. *Advances in Data Mining and Database Management*, 373-383. https://doi.org/10.4018/978-1-7998-3053-5.ch018

[35] Sivanandhini, P., & Prakash, J. (2020). Crop Yield Prediction Analysis using Feed Forward and Recurrent Neural Network. *International Journal of Innovative Science and Research Technology*, 5(5), 1092–1096. doi:10.38124/volume5issue5

[36] Prakash J. (2018). Enhanced Mass Vehicle Surveillance System, *J4R*, Volume 04, Issue 04 ,002, 5-9, June2018.

[37] Prakash, J., & Joy, E. J. (2020). A comparison of different surrogate models for delamination detection in composite laminates using experimental modal analysis. *PROCEEDINGS OF ADVANCED MATERIAL, ENGINEERING & TECHNOLOGY*. https://doi.org/10.1063/5.0019366

[38] Prakash, J., Vimal, E., Vinoth Kumar, B., & Priyadharshini, S. (2020). Impact Analysis and Combating of COVID-19 Around the World. *Journal of Analysis and Computation (JAC)*, *13*(7), 1-11.

[39] Sivanandhini, P., & Prakash, J. (2020). Comparative Analysis of Machine Learning Techniques for Crop Yield Prediction. *International Journal of Advanced Research in Computer and Communication Engineering*, *9*(6), 289-293. https://doi.org/10.17148/IJARCCE.2020.964