Comparative Study of Deep Learning Algorithms for Skin Disease Diagnosis Systems: An Analysis of Their Efficiency and Effectiveness

Author: Rahul Sunil Nair
Email: rahulsunilnair1074@gmail.com
Contact: 9975514723
Whatsapp:9975514723

Topic: Comparative Study of Deep Learning Algorithms for Skin Disease Diagnosis Systems: An Analysis of Their Efficiency and Effectiveness

Researcher: Rahul Sunil Nair

Abstract:

Skin diseases are a prevalent issue among people worldwide and early detection plays a crucial role in effective treatment it is also very hard to notice and detect hence many a times get missed out. In this context, Deep Learning (DL) models have shown remarkable performance in detecting various skin diseases. To address these issues, this research paper focuses on using Artificial Intelligence and Deep Learning technology to determine the best algorithm for skin disease detection which can be used for classifying between 9 different diseases. Main focus of this paper lies in the comparison of 3 models Convolution Neural Network (CNN), ResNet50V2 and VGG19 which were both implemented using Transfer Learning Technique. From this research paper I can conclude that ResNet50 V2 is a better model in comparison with VGG19 and CNN model in terms of accuracy with which they determine the diseases.

Keywords*: Deep Learning, Image Processing, Transfer Learning, Convolution Neural Network (CNN) Model Creation, Data Augmentation.*

Introduction:

With the rise in population India has seen a significant rise in the number of cases of Skin Diseases especially Skin cancer for the past decade. Traditionally Skin diseases has been neglected by human kind in comparison with other diseases but it is to be noted that skin disease if not detected in early stages can be fatal. Another factor being the ratio of health care workers to citizens is drastically separated. Also, the lack of medical facility available in rural area causes delay in detection. Apart from all this the cost of skin disease treatment even though it is comparatively less than other countries still a vast majority of citizens can't afford the cost for treatment due to not having an insurance. In this research paper, the use of Artificial Intelligence and Deep Learning technology is explored for skin disease detection. Three algorithms are analysed: Convolution Neural Network (CNN), Resnet50V2, and VGG19. The CNN algorithm is custom-built with 3 convolution layers, 3 pooling layers, and 2 dense layers. The other two algorithms are implemented through transfer learning using the pretrained MobileNet weights. The Resnet50V2 algorithm uses the Deep Residual Neural Network architecture and consists of 50 layers. The VGG19 algorithm consists of 19 layers, including 13 convolution layers and 3 fully connected layers. The research focuses on 9 skin diseases, both cancerous and non-cancerous, using both primary and secondary data collection methods. The primary method uses healthy skin samples and the secondary dataset is from the HAM10000 dataset, which includes 10000 samples of 7 skin diseases, and a portion of data from the Derment dataset for 2 additional diseases.

Objective of the Study:

The main objective of this study is to find the best and optimal algorithm for classification of skin disease between the three aforementioned algorithms which can then be used in a mobile application for easy and rapid prediction of skin diseases so that the users can get better guidance for further treatment or consultancy.

Research Methodology:

- ***Problem Formulation***: Like any research first problem statement was determined. After which study was conducted on the working of different Deep Learning technology and algorithms.

- ***Data Collection***: In this phase of my research, I had used both Primary and Secondary Data Collection methods. The HAM10000 dataset was used for procuring most of the images along with this dataset Derment Dataset was used for Additional diseases. The use of Primary Data collection

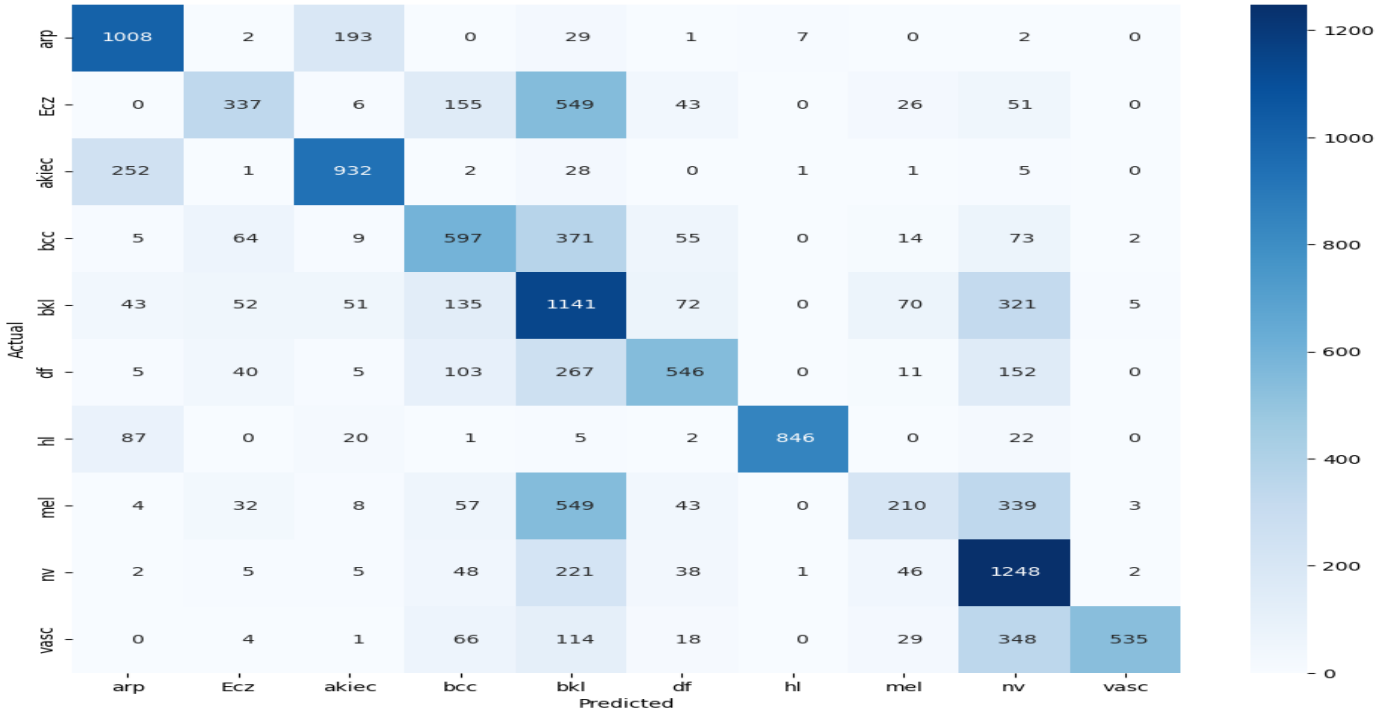
was necessary since lack of availability of healthy skin sample images therefore with the help of voluntary service healthy skin samples were collected and later processed.

- ***Data Cleaning***: In this step data was cleaned and labelled and arranged into their specific classes (12 classes). After that using data augmentation technique extra samples were created so that each class of disease have approximately same number of samples so that class imbalances can be removed from the dataset. After this the dataset was divided into 3 categories Training data, Validation data and Testing Data. with Training data having majority samples.
- ***Data Pre-processing***: In this Step the data was pre-processed finally before it can be used for training the Neural Network. In this step Resizing of images to 224 x 224 px size for having uniform image size and rescaling of image was done to maintain the aspect ratio of the image.
- ***Model Creation***: In this step I designed a custom CNN algorithm based on research knowledge which consisted of 3 Feature Extraction layers i.e., combination of Conv2D layer + Pooling layer + Dropout layer for regularization after that the model contained a Flatten Layer which converts multidimensional layer to 1D vector and finally it contained the Classification layer which consisted of 2 Dense layers of neurons.
  After that more two models namely Resnet50V2 and Vgg19 was implemented using the technique of transfer learning which is the process of using pre-processed models trained on a lot of data to solve our problem definition. For my model I have used the Mobilenet weights for training.
- ***Training:*** The input data for training were feed to the models in batches of 16/32 images. For the training step 25 epochs were used with the help of early stopping the epochs varied in all 3 models. A regularization parameter (L1 regularization) was introduced to reduce overfitting of model. After each epoch a validation step was done using the validation set to check the accuracy at each step.
- ***Testing and Analysis phase***: In this step our model was introduced with the testing set to evaluate the model with unknown data to check if the model was generalizing well. A confusion matrix was plotted to find the misclassified data.

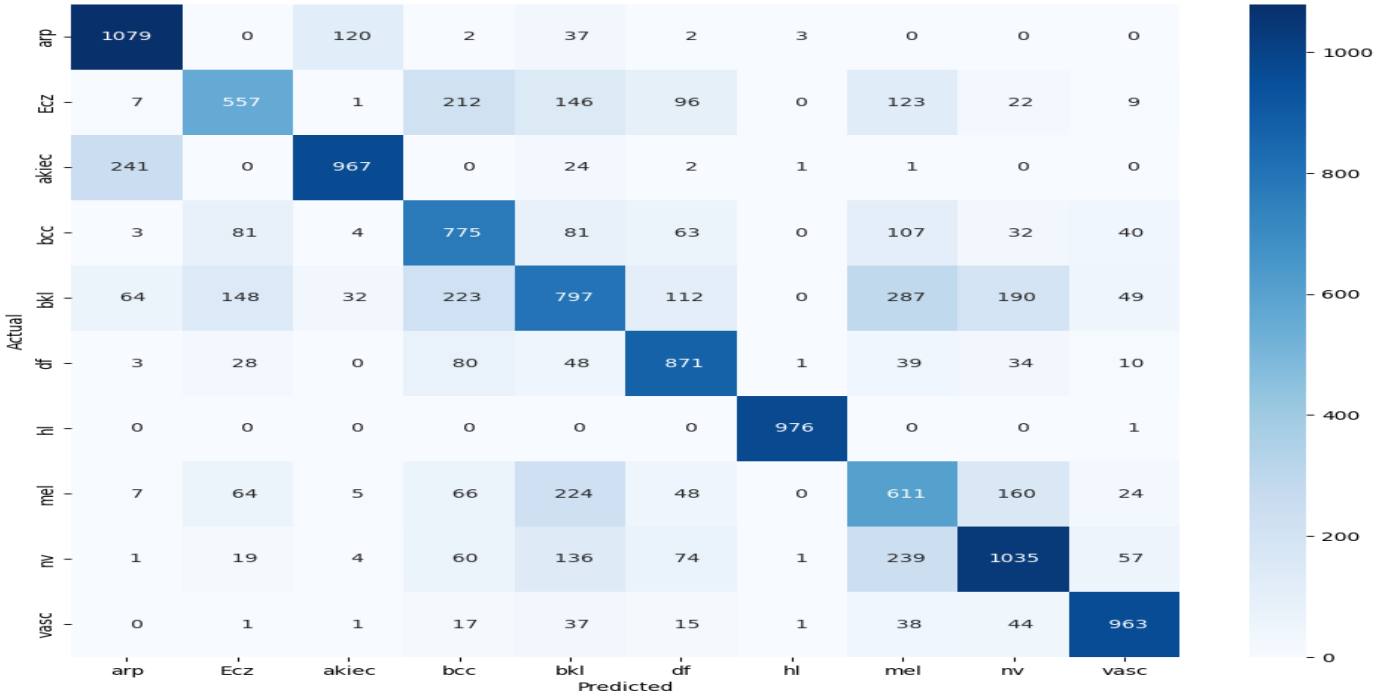
Data Analysis:

For the analysis of the model a structure called Heatmap is used to plot confusion matrix which plots the predicted values against Actual values. The values lying in the Diagonal of the square are the correct classification and is used for calculating the accuracy. The more the darker the colour the more the number of samples.

Confusion matrix for VGG19 model.



Confusion matrix of ResNet50V2 model



Confusion matrix for CNN model

Classification Reports

For VGG19

	precision	recall	f1-score	support
Ecz	0.72	0.81	0.76	1242
akiec	0.63	0.29	0.40	1167
arp	0.76	0.76	0.76	1222
bcc	0.51	0.50	0.51	1190
bkl	0.35	0.60	0.44	1890
df	0.67	0.48	0.56	1129
hl	0.99	0.86	0.92	983
mel	0.52	0.17	0.25	1245
nv	0.49	0.77	0.60	1616
vasc	0.98	0.48	0.64	1115
accuracy			0.58	12799
macro avg	0.66	0.57	0.58	12799
weighted avg	0.63	0.58	0.57	12799

For CNN

	precision	recall	f1-score	support
Ecz	0.74	0.68	0.71	1243
akiec	0.42	0.54	0.47	1169
arp	0.63	0.76	0.69	1226
bcc	0.42	0.70	0.53	1189
bkl	0.53	0.33	0.40	1894
df	0.89	0.33	0.48	1130
hl	0.92	0.93	0.92	982
mel	0.48	0.49	0.48	1249
nv	0.69	0.66	0.67	1616
vasc	0.85	0.97	0.91	1117
accuracy			0.62	12815
macro avg	0.66	0.64	0.63	12815
weighted avg	0.64	0.62	0.61	12815

For RESNet50 V2

	precision	recall	f1-score	support
Ecz	0.77	0.87	0.81	1243
akiec	0.62	0.47	0.54	1173
arp	0.85	0.78	0.82	1236
bcc	0.54	0.65	0.59	1186
bkl	0.52	0.42	0.46	1902
df	0.68	0.78	0.73	1114
hl	0.99	1.00	1.00	977
mel	0.42	0.51	0.46	1209
nv	0.68	0.64	0.66	1626
vasc	0.84	0.86	0.85	1117
accuracy			0.68	12783
macro avg	0.69	0.70	0.69	12783
weighted avg	0.68	0.68	0.67	12783

Findings and Recommendations:

Based on the precision, recall, and f1-score metrics in the tables for VGG19, ResNet50 V2, and CNN, it appears that the ResNet50 V2 model performed the best with an accuracy of 68%. This model had the highest macro average f1-score of 0.69. The VGG19 model had an accuracy of 58% and a weighted average f1-score of 0.57. The CNN model had an accuracy of 62% and a weighted average f1-score of 0.61. However, it is important to keep in mind that these results may not necessarily be representative of the models' overall performance, as the results could be dependent on the specific data used for the evaluation. Additionally, the overall performance of the models can be improved by fine-tuning the model architecture and by using techniques such as regularization to prevent overfitting.

Conclusion:

In general, these results suggest that there is room for improvement in these models, and that further research and development is needed in order to achieve better performance on a wider range of datasets. However, the performance of these models can also be influenced by factors such as the quality and diversity of the training data, the choice of evaluation metrics, and the specific requirements of the application. As such, it is difficult to make a definitive conclusion about which model is the "best" without considering these factors in more detail.

References:

https://www.kaggle.com/datasets/shubhamgoel27/dermnet
https://www.hindawi.com/journals/cmmm/2018/8145713/
www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000
https://www.kaggle.com/code/smitisinghal/skin-disease-classification