Journal of Critical Reviews
ISSN- 2394-5125 Vol 7, Issue 11, 2020

COMPARATIVE STUDY OF SVM AND CNN IN IDENTIFYING THE TYPES OF SKIN CANCER

P. Gokila Brindha[1], Dr.R.R. Rajalaxmi[2], S. Kabhilan[3], C. Sangitkumar[4], L. Sanjeevan[5]

[1]Assistant Professor, Department of CT-UG, Kongu Engineering College, Perundurai.
[2]Professor, Department of CSE, Kongu Engineering College, Perundurai.
[3,4,5]Student, Department of CT-UG, Kongu Engineering College, Perundurai.

Received: 09.04.2020 *Revised: 16.05.2020* *Accepted: 07.06.2020*

Abstract
The objective of the paper is to analyze the performance of the machine learning algorithms in detecting the types of skin cancer. There are different types of skin cancer and some of them may lead to death. So, the early prediction of skin cancer helps in reducing the death rate. The dataset is downloaded from Kaggle website. The train data consists of 2637 images of benign and malignant images and the test data consists of 660 images of benign and malignant images. The aim of the paper is to identify the algorithm, which gives maximum accuracy in detecting the types of skin cancer when applied on the image dataset. The algorithms used are Convolution Neural Network (CNN) and Support Vector Machine (SVM). The algorithms are executed using Tensor Flow in Python.

Keywords: CNN, SVM, TensorFlow, Deep Learning, Image Processing.

DOI: http://dx.doi.org/10.31838/jcr.07.11.117

INTRODUCTION

Skin cancer is caused due to the mutation of the skin cells that grows abnormally and forms the cancer cells. The skin cancer will affect the people irrespective of the skin tone. Most of the skin cancer will occur in the parts of the skin which is exposed to the sun such as face, ears, neck, chest, arm, fingernails or toenails palms, etc., The skin cancer can be reduced by limited exposure to the Ultra Violet (UV) rays. Not only the exposure to the sunlight but also to the exposure to the toxic substances and conditions that is against the immune system will cause skin cancer. Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC) and Melanoma are the different types of skin cancer. Identification of the skin cancer at the earlier stage can help in making a successful treatment but if it starts to spread it is hard to cure. BCC is an abnormal cell that is developed in the region of the body which is affected by the UV radiation from the sun. The second type of the skin cancer is SCC which occurs in ears, face and hands of the body. It will slowly spread in the tissues and bone which will be hard to cure. The last type and the most dangerous skin cancer is the melanoma which will spread fast in the tissues and it is difficult to cure and is a deadliest cancer. Kaposi sarcoma, Merkel cell carcinoma, Sebaceous gland carcinoma are some other types of skin cancer. The symptoms of skin cancer are bump, unusual growth, unusual color and edges on the skin. So the early detection of skin cancer may reduce the treatment stress. This paper focuses on applying the machine learning methods in predicting the type of skin cancer.

The objective of this paper is to identify the occurrence of skin cancer by analyzing the images using machine learning classifiers. Since the earlier detection of skin cancer may provide the possibility of getting cure and which helps in reducing death rate. Earlier detection also helps in reducing the intake of the medicines and the reduce the cycles of treatment procedures so that patient gets relived from psychological tension. This makes to bring an idea in analyzing the skin cancer. After applying the classifiers, a comparison is made to identify which algorithm is best in the prediction.

LITERATURE SURVEY

In this paper [1] the authors used to classify the skin cancer types using classifiers like Support Vector Machine (SVM), Probabilistic neural network (PNN), Adaptive Neurofuzzy Interface System (ANFIS) and made a comparison based on the performance of those classifiers. The dataset contains 200 images that include three types of skin lesions. The processes involved in classification are image pre-processing, lesion segmentation, feature extraction and classification. The experimental result shows that ANFIS performs better compared to SVM and PNN classifier.

A survey [2] about the skin cancer which speaks about the analysis of skin cancer detection using Image processing techniques. Types of Skin cancer are Basal Cell Carcinoma (BCC), Squamous-cell skin cancer (SCC), Melanoma. Symptoms of BCC looks like a raised, smooth, pearly bump on the sun-exposed skin of the head, neck, or shoulders. BCC often appears as a sore that does not heal. SCC is in red, scaling, thickened patch on sun-exposed skin. Ulceration and bleeding may occur. Melanoma differs from change in the size, shape, color or elevation of a mole.

This paper uses the CNN with the LeNet-5 architecture for classifying the images[3]. The application was coded in Python using keras library and TensorFlow backend. The dataset is taken from ISIC where 110 is melanoma and 110 is non - melanoma. The image is rescaled into 32x32 and data augmentation is used for the increasing the volume of the data. The training is initially conducted for 50 epochs but the accuracy is increased when 100 epochs applied on training. so here the 100 epoch is considered as optimal one.

The proposed work in this paper [4] is to improve the identification of the skin cancer by applying image processing techniques. The tool used here takes the color, area perimeter and diameter etc by texture, size, shape and various melanoma parameters for image segmentation and feature extraction. SVM classifier is used for identifying cancer using the extracted features. Dataset has been downloaded from Kaggle dataset. K

fold cross validation is applied for the evaluating the performance of the classifier.

In this paper [5] the authors have used Gabor filter to extract features from both the types of images that is the normal image and cancerous image. Support vector machine is used for training the model and the input images are given to model in order to test the model. Artificial Neural Network technique is also used for image classification. The classifiers classifies the input image as either skin lesion image or normal image. ANN is used for making a comparative study with SVM. SVM produces better results when compared with ANN

DATASET

The Dataset for this proposed work is downloaded from Kaggle website which was processed by ISIC. The Dataset consists of 1440 images of benign and 1197 images of malignant images used for training and 360 images of benign and 300 images of malignant for testing. The dataset used here is a balanced dataset. The size of each image is 224x224. The images are taken by using Dermoscopy tool. The images are RGB images. The data is placed in two different folders named train and test inside each folder, the images are stored in separate folders named benign and malignant

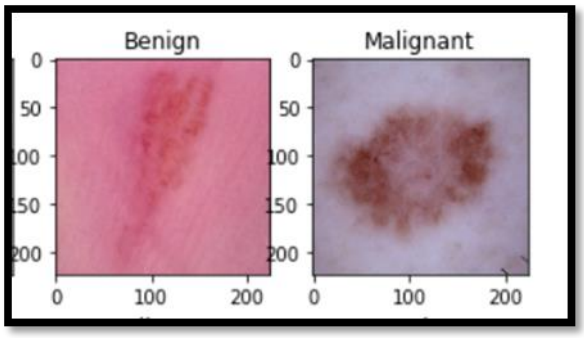

Figure 1: Skin Cancer Dataset

MATERIALS AND METHODS

The proposed work of this paper is to classify the types of the cancerous images using Support vector Machine (SVM) and Convolutional Neural Network (CNN). First step is taking the images and converting them into grayscale so as to reduce the number of channels. The RGB image will have 3 channels whereas the Grayscale image have 1 channel. Reducing the number of channels will increase the processing speed. The code is developed using Python having keras library and tensor flow at its backend. After the transformation of all the images into grayscale, the data frame will be created. The data set is divided into training set and testing set. The training set has 2637 images and the testing set has 660 images.

A. Support Vector Machine (SVM)

Support vector machine is one of the most popular and widely used supervised machine learning algorithm. Even though SVM provides higher accuracy compared to other algorithms but for huge dataset it won't give high accuracy. The concepts of SVM are relatively easy. The classifier splits the data points using a hyperplane having maximum margin. SVM helps in classifying new data points using the optimal hyperplane. SVM is generally a classification approach but it can be used for both the classification and regression problems. The main use of SVM is to divide the dataset into classes by using the maximum marginal hyperplane (MMH).

B. Convolutional Neural Network (CNN)

The CNN is mostly commonly used deep learning algorithm for image processing. Python provides the built in packages for deep learning. The methods of CNN is implemented in the Keras libraries. Tensor flow is used

because it gives a beautiful computational graph. The CNN can be applied in image and video recognition, recommender systems, image classification, medical image analysis and natural language processing. CNN gives maximum accuracy on image pre-processing compared to other algorithms. The CNN involves a math operation and multiple convolution layers. The layers in the CNN are Convolve layer, pooling layer, normalization layer, fully connected layer and output layer. The main purpose of convolution layer is to extract the features from the images. The matrix obtained after applying the convolve method is given to the Rectified Linear Unit (ReLU) activation function. ReLU is a linear function which produces the same value or zero for the (positive or negative) values given in the input matrix.

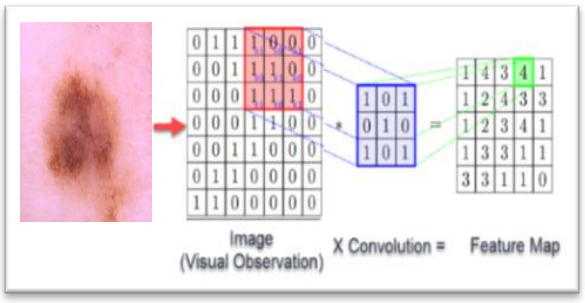

Figure 2: Feature Extraction of Skin Cancer Image

The pooling layer is applied to reduce the computational function in the network. Max pooling and average pooling are the two types of pooling layer and max pooling is the most common approach. Max pooling is used to select the maximum value at each spot of the image from the feature map. Here the Max pooling is applied with the pooling window of size 2x2.

The dropout function is to remove the connection between nodes from one layer to another layer. There may be an information loss when dropout is applied. The flatten is used to convert the matrix to the single dimensional array.

The flattened data is then applied to the sigmoid activation function. When there is a two class problem then the sigmoid function is the suitable method for prediction. The softmax function can be applied if it is multiple class function. Since in this problem we are going to identify the types of cancer(i.e., two types) the sigmoid function is used.

EXPERIMENTAL RESULT ANALYSIS

First, skin cancer images of size about 224X224 as shown in Fig 1, are taken. The images are converted to the grey scale images. The training and testing data are separately taken in the X_train and X_test variables and along with the labels in Y_train and Y_test. Here the filter size is taken as 3x3 and 64 kernels are used in the convolution layer.

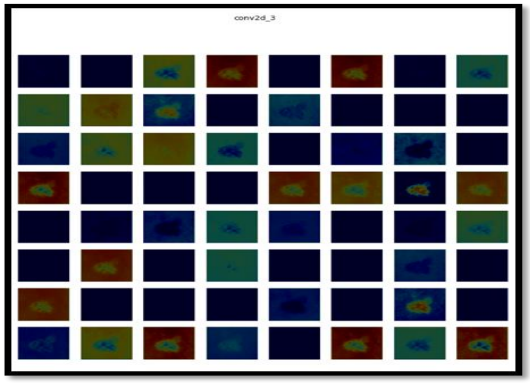

Figure 3: Convolve 2d Layer

Here the padding is same means that the size of the output image is same as the size of the input image. The 64 feature maps are generated after the convolve layer. In maxpooling the pooling window size of 2x2 is taken and size of the image is reduced to 112x112.The Droupout of 25 % is applied. Again the convolution layer is applied with the same filter size to the images which is about 112X112 in size. Then the image is ready for the pooling layer which will compress the image to make the process quickly. After maxpooling is applied t the size of the image is reduced to 56x56 as shown in Fig 5.

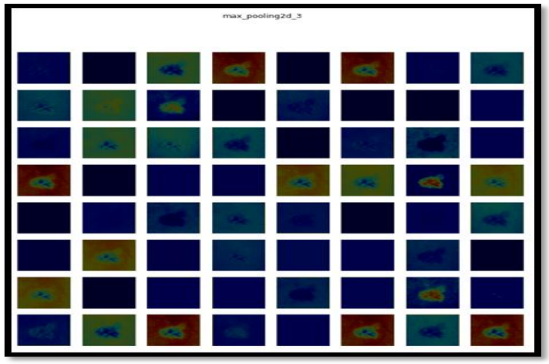

Figure 4: Max_pooling Layer

Then flattening is applied to the pooled feature map to find the class label of the image. Then in dense layer, the depth of image is taken as 64 and dropout is fixed as 0.25 as shown in Fig 6. Then the linearity of the image will be generated using the ReLU activation function. The sigmoid function at the last is used to identify the class of the image. Here prediction of the class label depends on the previous layer.

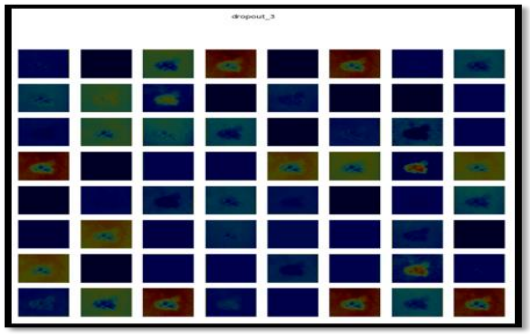

Figure 5: Dropout Layer

The number of epochs for the training the model is 50 and for each and every epoch the accuracy gets increased. The Fig 7 shows the final output after executing CNN for 50 epochs.

```
Epoch 45/50
 - 25s - loss: 0.0833 - accuracy: 0.9716 - val_loss: 0.4654 - val_accuracy: 0.8239

Epoch 00045: ReduceLROnPlateau reducing learning rate to 6.24999984211172e-07.
Epoch 46/50
 - 25s - loss: 0.0890 - accuracy: 0.9649 - val_loss: 0.4682 - val_accuracy: 0.8239
Epoch 47/50
 - 25s - loss: 0.1045 - accuracy: 0.9630 - val_loss: 0.4665 - val_accuracy: 0.8239
Epoch 48/50
 - 25s - loss: 0.0838 - accuracy: 0.9711 - val_loss: 0.4661 - val_accuracy: 0.8201
Epoch 49/50
 - 25s - loss: 0.0857 - accuracy: 0.9697 - val_loss: 0.4753 - val_accuracy: 0.8182
Epoch 50/50
 - 25s - loss: 0.0771 - accuracy: 0.9744 - val_loss: 0.4700 - val_accuracy: 0.8333

Epoch 00050: ReduceLROnPlateau reducing learning rate to 3.12499992105586e-07.
dict_keys(['val_loss', 'val_accuracy', 'loss', 'accuracy', 'lr'])
```

Figure 6: CNN Output

The accuracy of the model generated after the training and the testing phase is shown in the Fig 7.

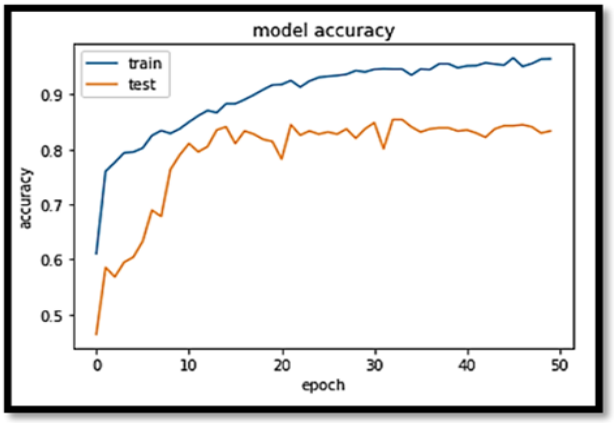

Figure 7: Model Accuracy

Considering SVM algorithm, the images are read from the folders and are converted to the grey scale images. These images are then added into the dataframes. The images are then normalized. The labels of the images are also added into the dataframes. There are separate dataframes for the training and testing images along with the class labels(in separate dataframes). Now SVM is applied on the training dataset, the model is generated and then is tested on the testing dataset. In this paper SVM is implemented to the image dataset. The accuracy obtained using SVM is 61%.

	0	1	2	3	4	5	6	7	8	9	...
0	0.611765	0.615686	0.611765	0.627451	0.650980	0.650980	0.662745	0.650980	0.654902	0.654902	...
1	0.729412	0.737255	0.745098	0.756863	0.756863	0.756863	0.756863	0.760784	0.760784	0.749020	...
2	0.086275	0.090196	0.086275	0.086275	0.090196	0.062745	0.407843	0.584314	0.560784	0.576471	...
3	0.690196	0.682353	0.682353	0.694118	0.698039	0.690196	0.694118	0.701961	0.705882	0.698039	...
4	0.654902	0.654902	0.631373	0.639216	0.647059	0.643137	0.666667	0.674510	0.670588	0.666667	...
...	...	...	...	...	...	...	...	...	...	...	...
2632	0.219608	0.250980	0.282353	0.298039	0.329412	0.341176	0.352941	0.380392	0.407843	0.427451	...
2633	0.670588	0.635294	0.611765	0.603922	0.627451	0.635294	0.639216	0.658824	0.662745	0.682353	...
2634	0.678431	0.678431	0.678431	0.682353	0.682353	0.690196	0.686275	0.686275	0.694118	0.698039	...
2635	0.666667	0.615686	0.498039	0.427451	0.521569	0.592157	0.545098	0.537255	0.529412	0.513725	...
2636	0.650980	0.650980	0.666667	0.690196	0.682353	0.670588	0.666667	0.662745	0.666667	0.666667	...

2637 rows × 50177 columns

Figure 8: Dataframe

In this paper two classifiers Support Vector Machine (SVM) and Convolutional Neural Network (CNN) are used in predicting the type of skin cancer. These algorithms are trained and tested on the image dataset. As per the functions performed CNN gives the maximum accuracy of 83% accuracy and SVM gives 61% accuracy.

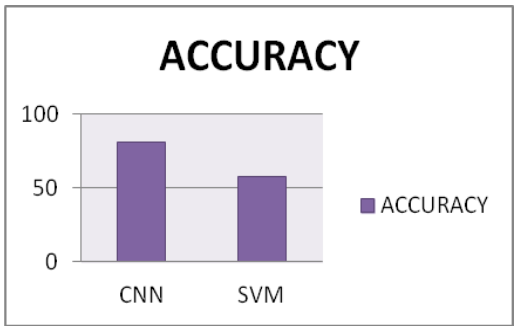

Figure 9 : Accuracy

CONCLUSION AND FUTURE WORK

In this proposed work SVM and CNN is applied for the prediction of the skin cancer types. The experiments are done using Python programming. Image processing is done with the help of tensorflow and the accuracy is calculated for the algorithms used. The result shows that CNN gives the maximum accuracy when compared to SVM. The future work can be done by optimizing the various parameters obtained when applying the deep learning method so as to improve the accuracy of the algorithm.

REFERENCES

1. Kavimathi.P and Sivagnanasubramanian.S. P (2016) "Analyzing the Performance of Classifiers for the Detection of Skin Cancer with Dermoscopic Images", International Conference on Innovations in Engineering and Technology (ICIET).
2. N. Durga Rao, Dr. G. Sudhavani (2017) "A Survey on Skin Cancer Detection System", International Journal of Engineering Research and Application.
3. Rina Refianti, Achmad Benny Mutiara, Rachmadinna Poetri Priyandini (2019) " Classification of Melanoma Skin Cancer using Convolutional Neural Network ", International Journal of Advanced Computer Science and Applications.
4. Poornima M S, Dr. Shailaja K (2017) "Detection of Skin Cancer Using SVM", International Research Journal of Engineering and Technology (IRJET).
5. Jisha Gopinath, Maya John (2015) "Feature Extraction by Gabor Filter and Classification of Skin Lesion using Support Vector Machine", International Journal of Innovative Research in Computer and Communication Engineering.
6. Suleiman Mustafa (2018) "A SVM-Based Diagnosis of Melanoma using Only Useful Image Features", National Agency for Science & Engineering Infrastructure.
7. Santhosh Dan (2016) "Image Classification using Data Mining Techniques", Advances in Computer Science and Information Technology (ACSIT).
8. Honsy KM, Kassem MA, Foaud MM (2019) "Classification of Skin Lesions using transfer learning and augmentation with Alex-net" PLoS ONE.
https://www.mayoclinic.org/diseases-conditions/skin-cancer/symptoms-causes/syc-20377605