IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A COMPREHENSIVE STUDY OF MALARIA DETECTION USING MACHINE LEARNING

[1]Prajna KB, [2]K Shreshta Subodh Shetty

[1]Assistant Professor, [2]Assistant Professor

[1]ECE Department,

[1]Nitte Meenakshi Institute of Technology, Bengaluru, India

Abstract: Malaria is a disease which is transmitted to people by the bites of female Anopheles mosquitoes. The major symptoms of malaria are fever, tiredness, vomiting, headaches. If it is very severe it can cause the patient to go to coma or cause death. It is preventable and curable. With the normal malaria detection methods, requires a lot of effort in differentiating infected and uninfected cells. Typically, it is really difficult to carry out this process in rural areas. Machine Learning is application of Artificial intelligence which that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly. In this paper we have modeled using four models basic CNN , SVM , and other two models with deep transfer learning . The pre-trained with image augmentation gives an accuracy of 96.1 %.

***Index Terms*- Machine Learning, SVM, CNN, Deep Transfer Learning.**

I. INTRODUCTION

Malaria also called as Plasmodium infection is a disease caused by a single celled plasmodium parasite which are transmitted through the bites of infected female Anopheles mosquitoes called as malaria vectors.

Its symptoms include fever, tiredness, chills, headache, sweating, cough, chest and abdominal pain, nausea and vomiting. It can also lead to yellow skin, coma or even severe i.e., death. Usually, symptoms begin ten to fifteen days after bitten by infected mosquito.

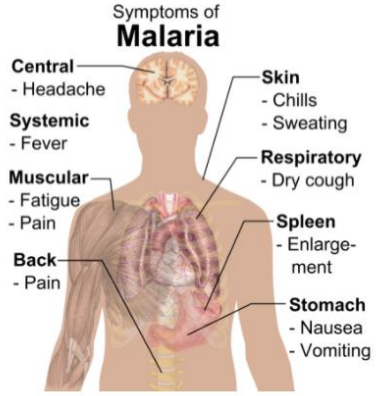


Figure: Symptoms of Malaria

If a person gets bitten by an infected mosquito, parasites carried buy it enter your body and starts destroying the Red Blood Cells (RBC). Initial symptoms are similar to flu, which starts few days after the mosquito bite. Sometimes the parasites, don't show any symptoms and can stay in human body for period of one year. As result of this there is a delay in treatment. According to the latest report of world malaria which was released on 30th November 2020 there were 229 million malaria cases in 2019 and 228 million cases in 2018. The number of death due to malaria in 2019 is 40900 and in 2018 is 411000.

1.1 Malaria Detection

There are several methods to detect malaria. General method is examining a patient's blood under microscope. The test result will be accurate depending on the quality of microscope and experience of the lab technician. Rajaram et al have discussed about another method rapid diagnostic test (RDT). These tests require offer an alternative where good microscope facility is not available. Generally, they recommend to do microscopy test, after RDT. As a result, it will take time to detect malaria. Many other tests like serology, drug resistance tests (DRT) are available but have disadvantages like serology don't detect the current infection and DRT is time consuming.

1.2 Complications

Malaria has serious complications like respiratory distress, pneumonia, anaemia and kidney failures if not treated properly. If infected to pregnant women lead to stillbirths, infant mortality and miscarriage. Brain and lungs are the most affected organs in malaria. Because of all this malaria is considered a medical emergency and should be diagnosed immediately. Delay in diagnosis leads to death. So, finding the malaria in infected patients is much important. In this paper we will be using Machine Learning in order to detect malaria in early stages.

1.3 Why machine learning?

Machine Learning (ML) provides the systems to learn automatically and improve from previous experiences. Machine learning consists of automated data processing algorithms which help in decision making. There are three categories:
Supervised ML – uses labeled input and output data, algorithm learns from training dataset.
Unsupervised ML – no need to supervise the model.
Reinforcement ML – models are trained to make decision sequences.

Figure: Categories of Machine Learning

There are four major machine learning models which are depicted along with the examples for each algorithm in the figure below:

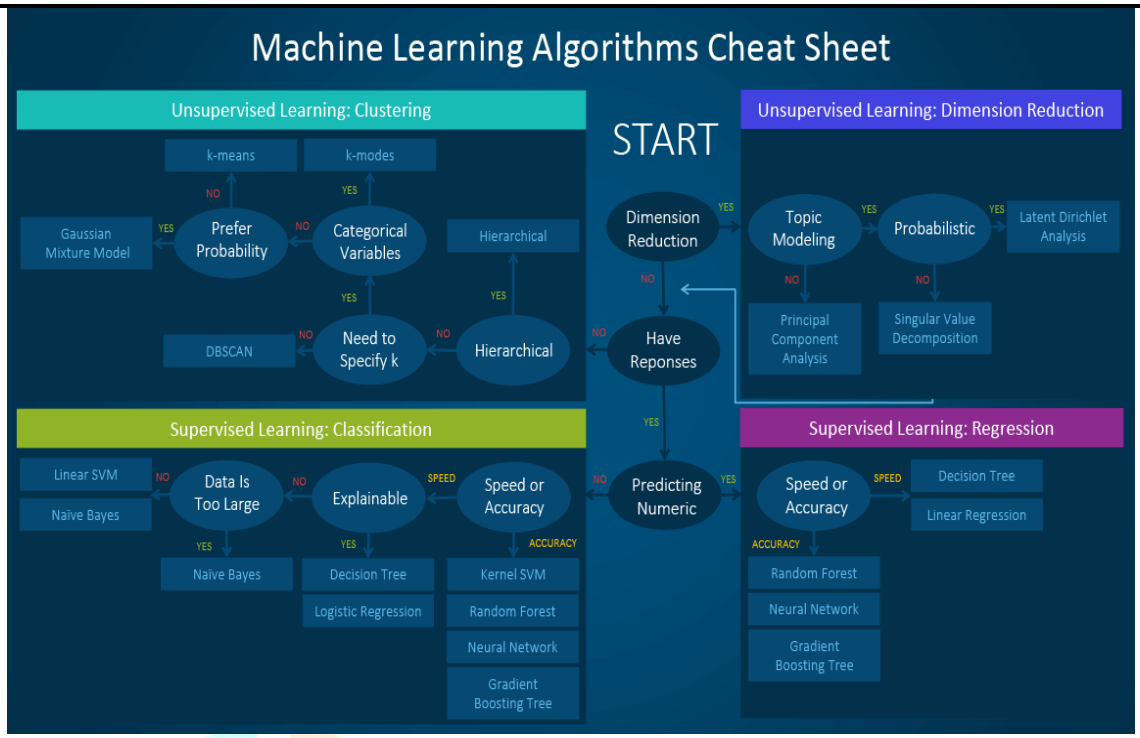


Figure: Machine Learning models with examples

Machine learning algorithms can be used in Banking and financial sectors, Healthcare, Public safety, Retail and e-commerce, Oil, gas and energy sector, Automotive applications, Advertising technologies etc.,

1.4 ML in Healthcare

Machine learning in healthcare helps in identifying diseases and diagnosis, drug discovery and manufacturing, medical imaging, personalized treatment, smart health records and predicting diseases. ML algorithm detects the patterns associated with the health conditions and diseases by studying the data base of health records of patients then provides outcomes and risk scores timely.

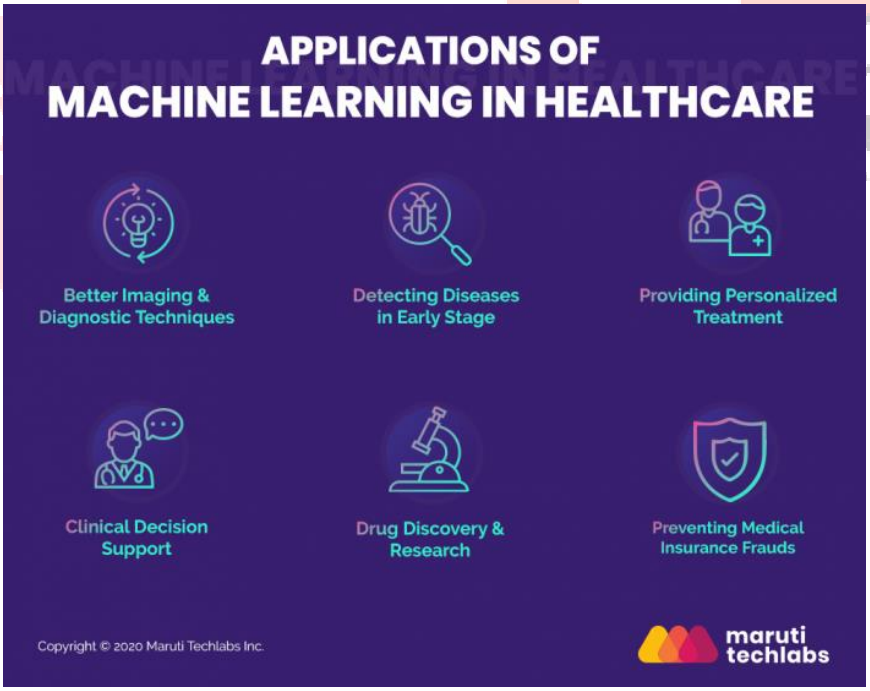


Figure: Applications of ML in healthcare

II DATASET

In this work we have used the malaria dataset of National Institute of Health (NIH) (https://ceb.nlm.nih.gov/proj/malaria/cell_images.zip.) There is total 27,598 cell images of thin blood smear on a slide. The data base has two segregations for uninfected and parasitized. Each folder consists of 13,799 cell images.
This is the sample resized view of the cell images as seen in jupyter notebook.

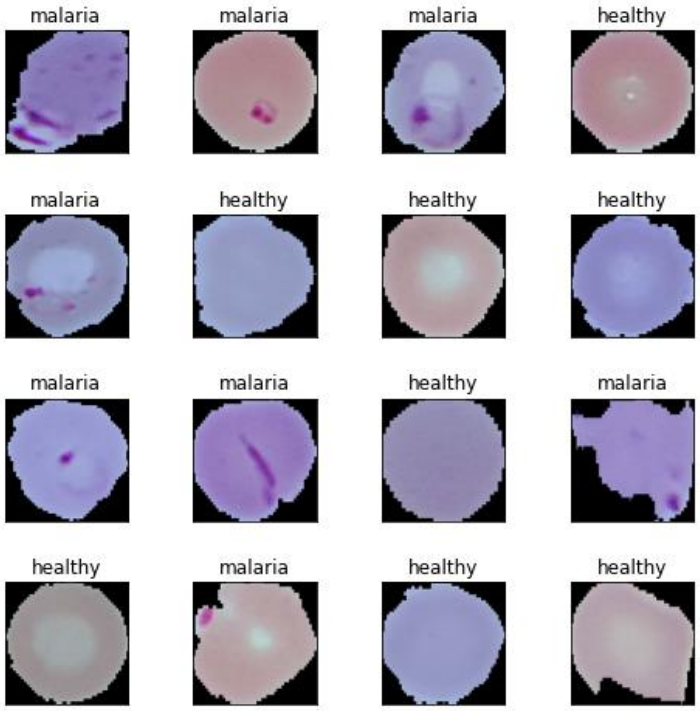


Figure: Malaria dataset view

III DEEP LEARNING MODEL TRAINING

In our paper we have used four deep learning models, trained them with our training data and compare the performance of them.

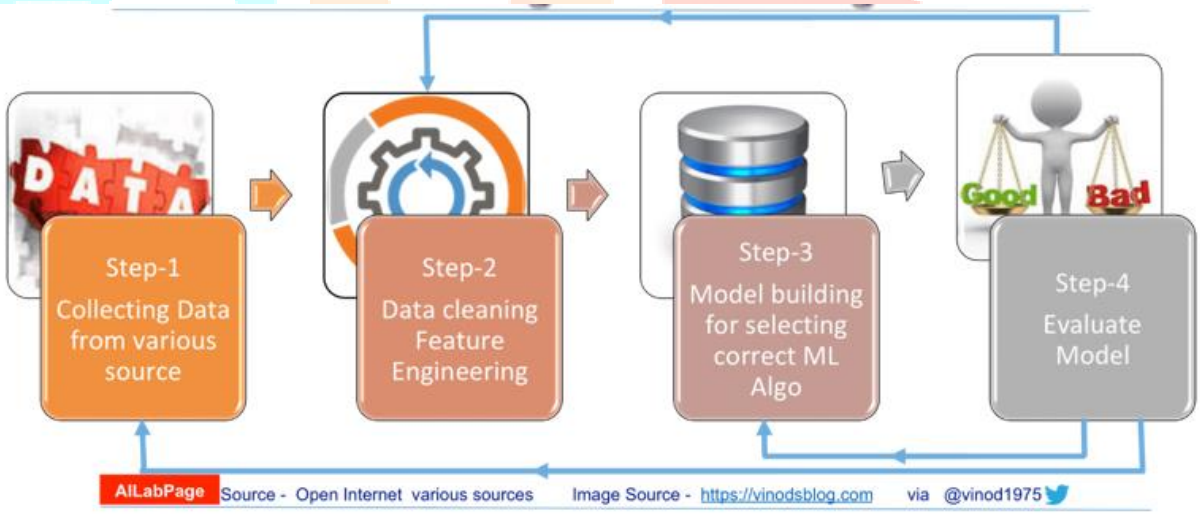


Figure: Machine Learning Process Flow

3.1 Support Vector Machine

There are several algorithms under Supervised, Unsupervised and Reinforcement machine learning. One such algorithm which comes under Supervised learning is Support Vector Machine (SVM). It is classified as linear SVM and non-linear SVM depending on the distribution of the data. SVM can be used for regression and classification problems. However, it is used mostly as the classification algorithm in order to make two-groups in classification problems depending on the attributes.

The algorithm is based on finding a decision boundary called as hyper line or hyper plane which segregates the dataset into two groups as shown below.

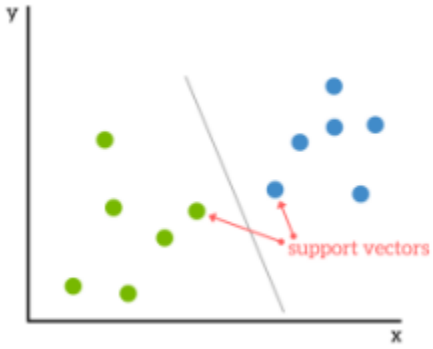


Figure: SVM as classification algorithm

SVM finds a hyperplane in a N dimensional array such that the data points can be easily classified. The number of features play a major role in the number of hyperplanes. In our case there only two or three components, as a result there is no line or plane that can divide the data. We have used nonlinear SVM classifier with Radial Basis Kernel (RBF) to differentiate between uninfected and malaria infected images. A function called SVM kernel takes low dimensional input and transfigures it to higher dimensional space. The Kernel converts it to a separable problem from a non-separable one.

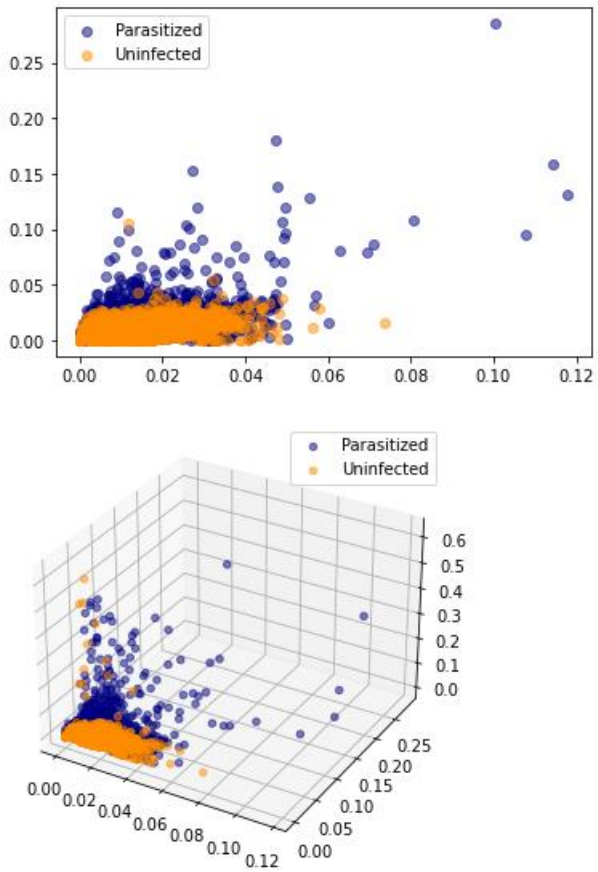
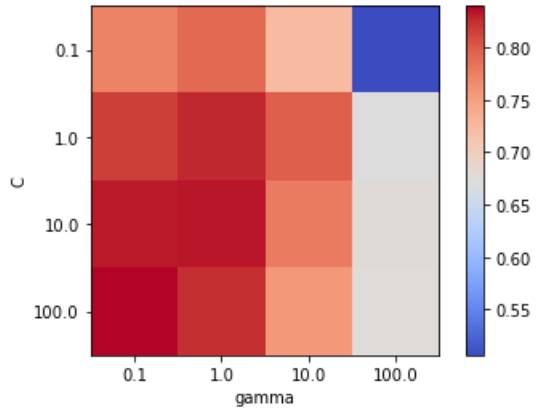


Figure: Normalized colour histogram data

We have used 80 % of data for training and 20 % for testing the data. In our work grid search was used in order to fit the multiple SVM models on the training data to get best gamma and C for the SVM RBF kernel. Gamma is the kernel co-efficient for 'rbf', 'poly' and 'sigmoid'. The higher the value of gamma, we will get best fit as per the training data set. C: Penalty parameter C of the error term. Pre-processing done by using pipeline before fitting each model.




```
Best accuracy 0.839 using params: {'svc__C': 100.0, 'svc__gamma': 0.1}
Accuracy 95% confidence interval: 0.831 - 0.848
```

Figure: heatmap within the range of gamma and c

The figure show heatmap of within the range of gamma and C combinations explored through grid search, the highest mean accuracy of 83.9% was achieved using gamma=0.1 and C=100. Finally, the SVM model was evaluated on the testing set with gamma =0.1 and C=100 to get the performance of unseen data. The result is shown in figure.

```
Accuracy: 0.833
Matthews CC: 0.665
F1 Score: 0.830
Sensitivity: 0.820
Specificity: 0.841
```

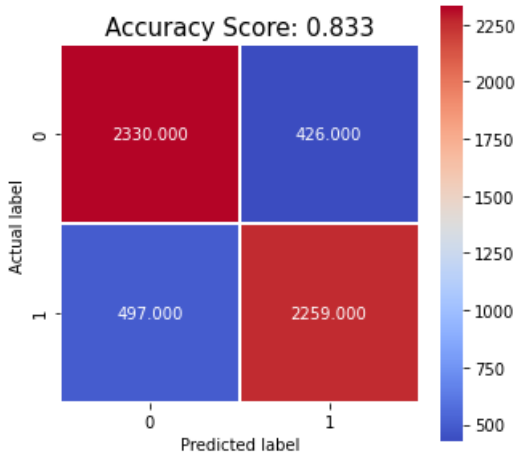


Figure: SVM accuracy

3.2 Convolutional Neural Network

Convolutional Neural Network (ConvNet or CNN) is a common algorithm for deep learning having images, videos. CNN comprises of three layers:Input layer, Output layer and Hidden layers in between input and output layers.

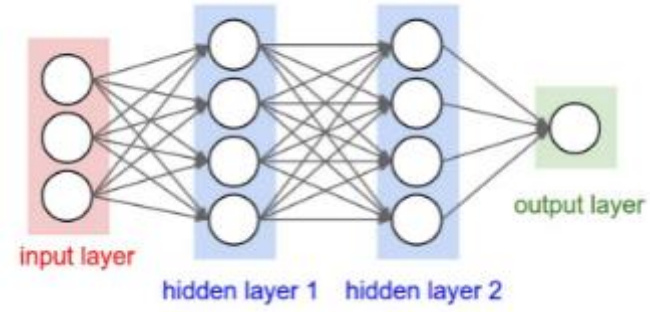


Figure: A three-layer neural network

A grid like topology is used in convolutional neural network to analyze the images. Any image we need to first feed the image in the form of arrays as input to the neural network. There are many hidden layers which perform different calculation and manipulation to obtain image feature extraction.

There are important hidden layers are

1.**Convolution layer:** This is the first step in feature extraction. This layer has several filters to perform convolution operation. Here every image is considered as a matrix of pixel values.

2.**ReLu layer:** In Rectified linear unit layer an element-wise operation is performed and all negative pixels are set to 0.

3.**Pooling layer:** Pooling is a down-sampling operation that reduces the dimensionality of the feature map.

4.**Flattening:** Flattening converts all the resultant pooled featured maps into single long continuous linear vector from a two-dimensional array.

Also, if dense layers are used in the end it helps in image classification

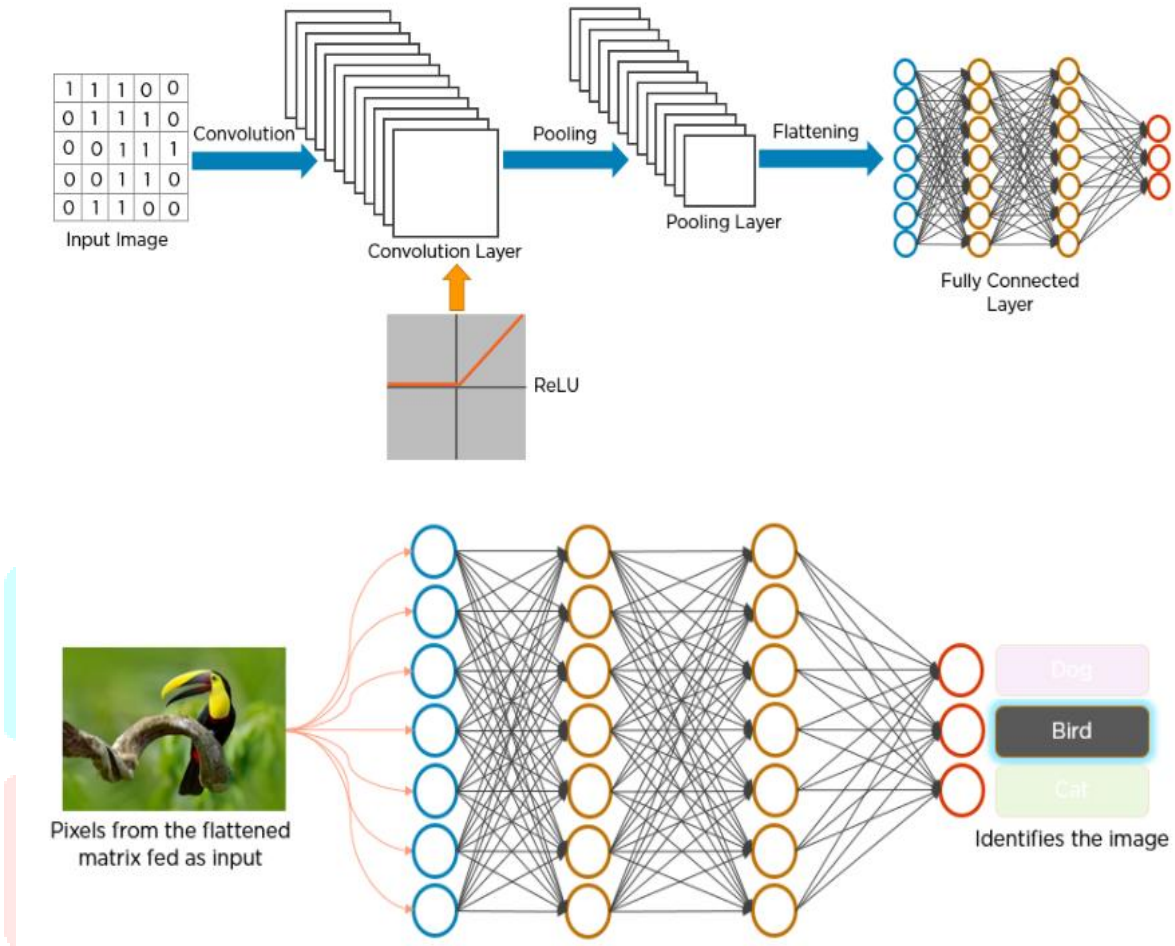


Figure: CNN used in bird recognition

The Steps used for training CNN (Convolutional Neural Network).

1: Upload Dataset
2: The Input layer
3: Convolutional layer
4: Pooling layer
5: Convolutional layer and Pooling Layer
6: Dense layer
7: Logit layer

Our CNN model has three convolution, three pooling layers, along with two dense layers and two dropouts for regularization. We got a validation accuracy of 95.49 % and testing accuracy 95.44%, but our model slightly over fitting with training accuracy of 99.88 %.

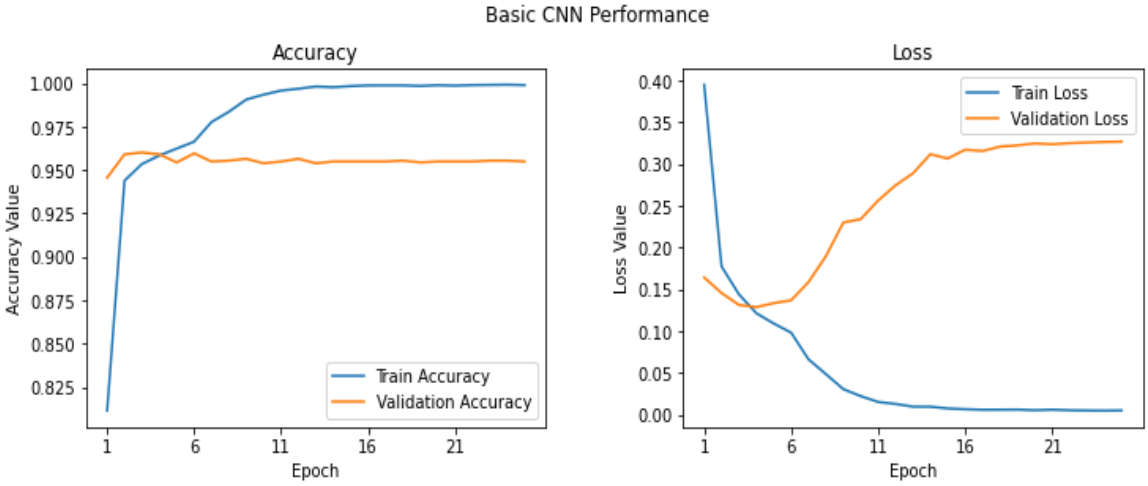


Figure: Accuracy and Loss of basic CNN

3.3 Deep Transfer Learning

It is an approach in machine learning and deep learning, where the knowledge is transferred from one model to the other. It a technique where a model is trained first on a similar problem to the problem of our interest. From the trained model one or more layers are used in a new model on the problem which is being solved. The main advantage of transfer leaning is it saves training data time, performance is better, computation power required is less and does not require more data for training.

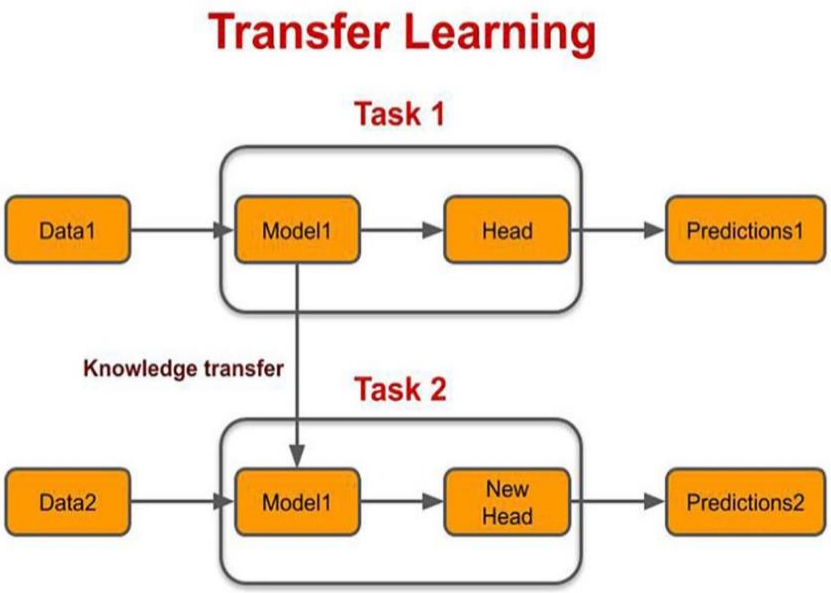


Figure: Illustration of Transfer learning

The is a difference between traditional learning and transfer learning. In traditional learning, learning generally happens on specific tasks and using models for training. In transfer learning, we can use the knowledge (features, weights etc) from previously trained models for training newer models.

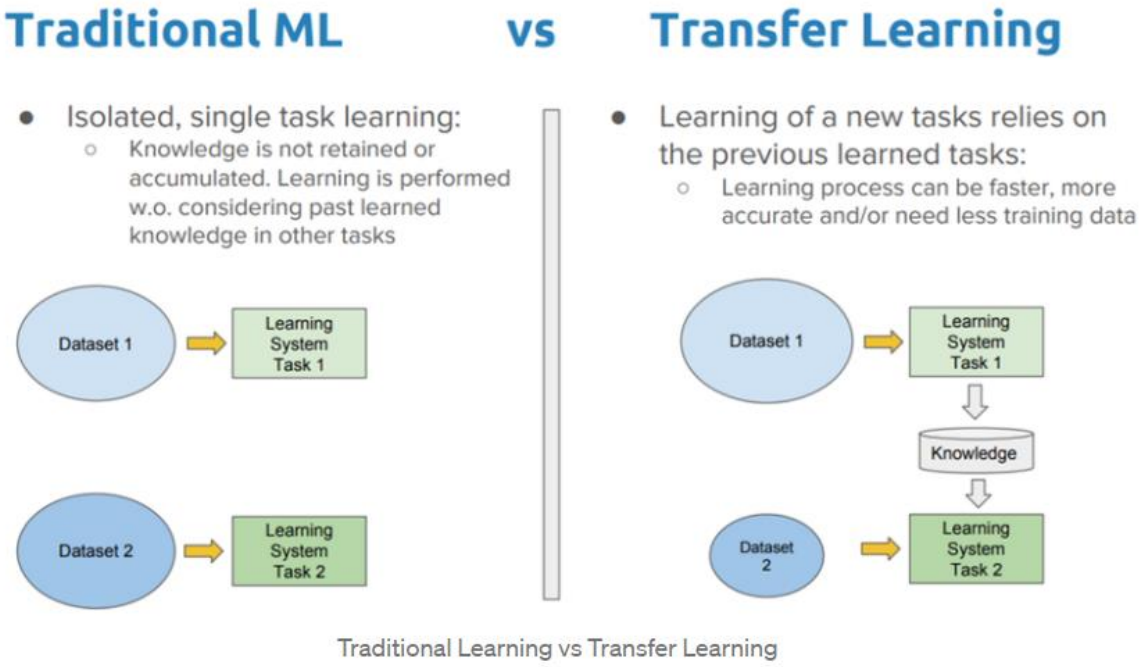


Figure: Traditional learning vs transfer learning

Layers taken from previously trained model are frozen in order to avoid corrupting of the information of the layers during training the new model. The below figure is the Size similarity matrix with the help of which we can decide the strategy.

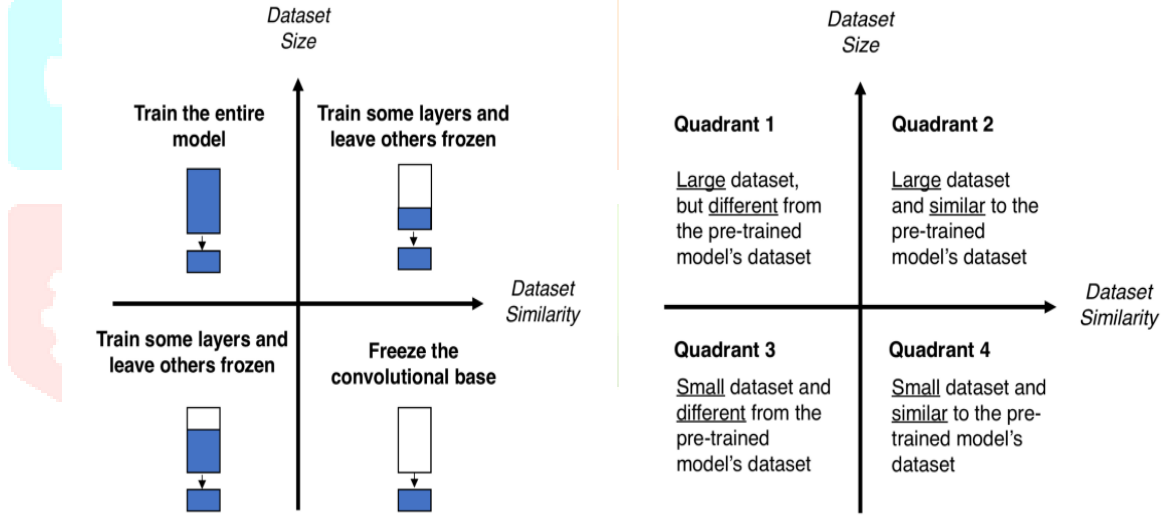


Figure: Size similarity matrix

Three popular models are as follows:

- ✓ VGG (e.g., VGG16 or VGG19).
- ✓ Residual Network (e.g., ResNet50).
- ✓ GoogLeNet (e.g., InceptionV3).

We have used pre trained VGG -19 model, which has a lot of image categories. It has 16 convolution layers with 3x3 convolution filters followed by max pooling layers for down sampling and two fully connected hidden layers.

3.3.1 Pre-trained model as a feature extractor

We have used one of the models for feature extraction by freezing the five convolution blocks. Then plugged it in to our dense layers to perform classification.

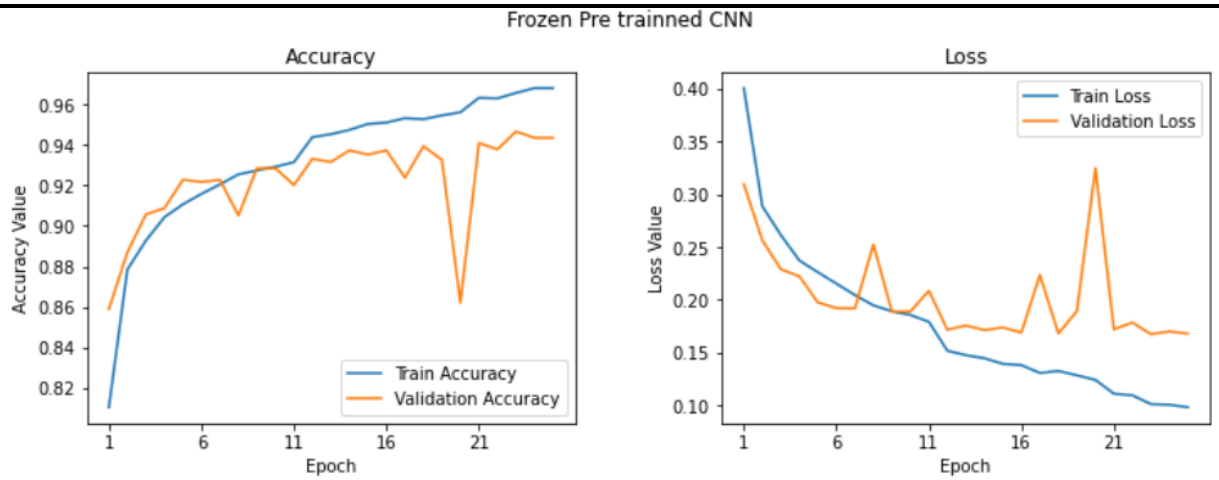


Figure: Accuracy and Loss of Pre-trained model without image augmentation

The graph shows that model is not overfitting as basic CNN, but the validation accuracy obtained is 93.3% and testing accuracy is 93.37%

3.3.2 Pre trained model with Image Augmentation

Image Augmentation is a technique which can be used to modify the images present in the dataset for training in order to have better performance and generalizing ability of the model. It increases the number of images present in the dataset. Pre-processing transformations are performed on the existing images like cropping, vertical and horizontal flipping, rotating, skewing etc.,

We fine-tuned the weights of layers of last two blocks of pre trained VGG-19 model. We have also applied image augmentation. Image augmentation means performing some transformation on image like zooming, rotating, translation etc. As a result, we get different images every time.

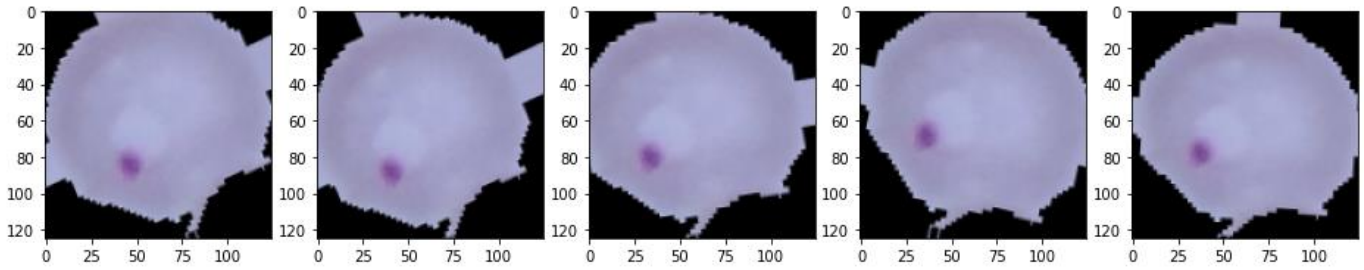


Figure: Image Augmentation

The model is not over fitting and gives a validation accuracy of 96.45, testing accuracy of 96.1 %.

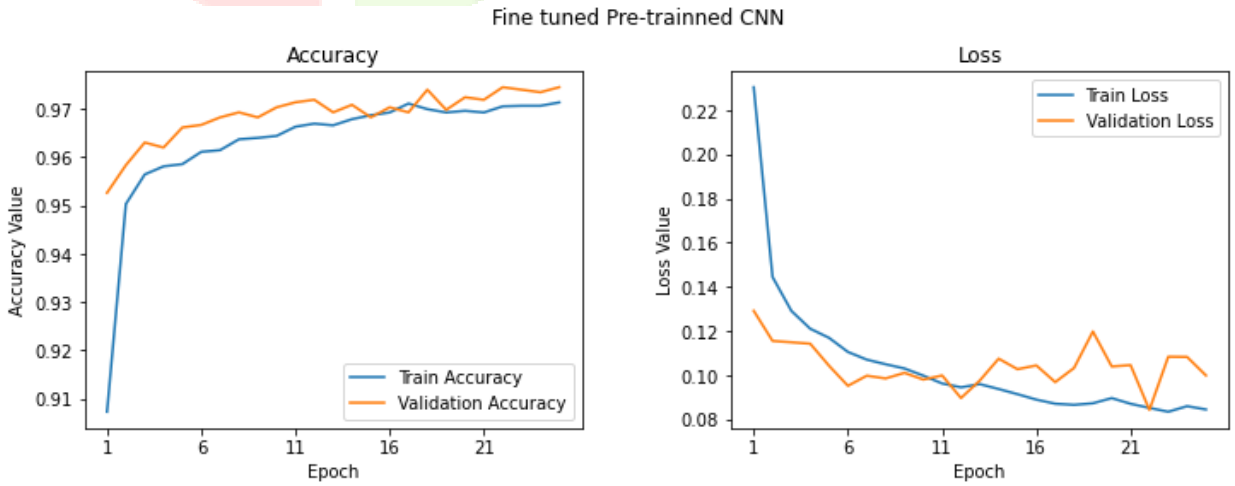


Figure: Accuracy and Loss of Fine-tuned Pre-trained CNN

IV CONCLUSION

According to the analysis made, fine-tuned pre trained CNN has the best accuracy of 96.1 %. In future we will be analyzing the other algorithms and will be checking whether there will be change in the accuracy. Using deep learning techniques in medical field will help to get the results soon and will help in treating the patients as early as possible.

Table: Comparison of various machine learning algorithms

MACHINE LEARNING ALGORITHMS	ACCURACY
Support Vector Machine (SVM)	83.3%
Convolutional Neural Network (CNN)	95.44%
Deep Transfer Learning - VGG -19 model without image augmentation	93.3%
Deep Transfer Learning - VGG -19 model with image augmentation	96.1%

REFERENCES

[1] Kaustubh Chakradeo, Sofya Titarenko, Michael J Delves, "Malaria Parasite Detection Using Deep Learning Methods", International Conference on Machine Learning Methods for Computer Vision, New York, February 2021, DOI:10.5281/zenodo.4569849

[2] Gautham Shekar; S. Revathy; Ediga Karthick Goud, "Malaria Detection using Deep Learning", 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), 15-17 June 2020.

[3] Yongli Zhang "Support Vector Machine Classification Algorithm and Its Application", International Conference on Information Computing and Applications ICICA 2012: Information Computing and Applications pp 179-186, Part of the Communications in Computer and Information Science book series (CCIS, volume 308).

[4] Anay Raj K S, "Malaria Disease Diagnosis using Machine Learning Techniques", International Research Journal of Engineering and Technology (IRJET),Volume: 07 Issue: 06 | June 2020.

[5] Nadia Jmour, Sehla Zayen, Afef Abdelkrim, "Convolutional Neural Networks for Image Classification", International Conference on Advanced Systems and Electric Technologies (IC_ASET 2018).

[6] Sumit Patil, Piyush, Rahul Kumar, Akash Raju, Susila M, "Detection of Malaria Using Image Cells", International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 03 | Mar 2020.

[7] Md. Khayrul Bashar, "Improved Classification of Malaria Parasite Stages with Support Vector Machine Using Combined Color and Texture Features", 2019 IEEE Healthcare Innovations and Point of Care Technologies, (HI-POCT), 20-22 Nov. 2019, DOI: 10.1109/HI-POCT45284.2019.8962686.

[8] Connor Shorten and Taghi M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning", Journal of Big data, Springer, 2019, https://doi.org/10.1186/s40537-019-0197-0.

[9] Muhammad Imran Razzak, Saeeda Naz, Ahmad Zaib, "Deep Learning for Medical Image Processing: Overview, Challenges and the Future Classification in BioApps", Springer (2018).

[10] Dipanjan (DJ) Sarkar "Detecting Malaria with Deep Learning AI for Social Good — A Healthcare CaseStudy", towardsdatascience (2018).

[11] Sivaramakrishnan Rajaraman, Sameer K Antani, Mahdieh Poostchi, Kamolrat Silamut, Md. A. Hossain, Richard J. Maude, Stefan Jaeger and George R. Thoma "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images", Reasearchgate, (2018), PeerJ, DOI 10.7717/peerj.4568 2/17.

[12] Miss. Shruti Annaldas, "Automatic Identification of Malaria Parasites using Image Processing", International Journal of Emerging Engineering Research and Technology Volume 2, Issue 4, July 2014, PP 107-112 ISSN 2349-4395 (Print) & ISSN 2349-4409 (Online).