**“Malaria Diagnosis Using Deep Learning”**

**A**

**Minor Project Report**

***Submitted by***

**Prabhat Rawat**

**17BCON094**

***In partial fulfilment for the award of the degree***

***of***

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**At**

**JECRC UNIVERSITY, JAIPUR**

**May 2020**

# Acknowledgements

Many people have supported me, in different ways, during the work with the thesis. I’d like to thank my guide Dr. Bhawana Sharma & HOD Dr. Naveen Hemrajani for their kind and active support and valuable guidance during the work process. My family has as always offered me their unconditional support, thank you! I have taken efforts in the Minor Project. However, it would not have been possible without the kind support and many individuals and organizations. I would like to extend my sincere thanks to each and every members related to JECRC University.

Prabhat Rawat

17BCON094

**Candidate’s Declaration**

I, **Prabhat Rawat**, bearing roll number **17BCON094**, hereby declare that the work which is being presented in the Minor Project, entitled **“ Malaria diagnosis using Deep Learning ”** in partial fulfilment for award of Degree of “**Bachelor of Technology**” in Department of **Computer Science Engineering** is submitted to the Department Computer Science & Engineering, JECRC University is a record of Minor Project work carried under the Guidance of Guide name, Department Computer Science & Engineering.

I have not submitted the matter presented in this work anywhere for the award of any other Degree.

Prabhat Rawat

Computer Science

Enrolment No.: 17BCON094

**JECRC UNIVERSITY JAIPUR**

Ramchandrapura, Sitapura Industrial Area Extn., Jaipur-303905(Raj.) India www.jecrcuniversity.edu.in

Date : 20th May 2020

**CERTIFICATE**

Certified that the Project Report entitled “**Malaria Diagnosis Using Deep Learning**” submitted by **Prabhat Rawat**  bearing roll no. **17BCON094**. In partial fulfilment of the requirements for the award of the degree of Bachelor of Technology of JECRC University, Jaipur is a record of the student’s own work carried out under my supervision and guidance. To the best of our knowledge, this Minor Project work has not been submitted to JECRC University or any other university for the award of the degree. It is further understood that by this certificate the undersigned does not endorse or approve of any statement made, opinion expressed or conclusion drawn therein but approve Minor Project for the purpose for which it is submitted.

(Guide) (HOD, CSE )

. **Introduction** .

Malaria is a deadly, infectious, mosquito-borne disease caused by Plasmodium parasites that are transmitted by the bites of infected female Anopheles mosquitoes. There are five parasites that cause malaria, but two types—P. falciparum and P. vivax—cause the majority of the cases.

If an infected mosquito bites you, parasites carried by the mosquito enter your blood and start destroying oxygen-carrying red blood cells (RBC). Typically, the first symptoms of malaria are similar to a virus like the flu and they usually begin within a few days or weeks after the mosquito bite. However, these deadly parasites can live in your body for over a year without causing symptoms, and a delay in treatment can lead to complications and even death. Therefore, early detection can save lives.

The World Health Organization's (WHO) malaria facts indicate that nearly half the world's population is at risk from malaria, and there are over 200 million malaria cases and approximately 400,000 deaths due to malaria every year. This is a motivatation to make malaria detection and diagnosis fast, easy, and effective.

Manual diagnosis of blood smears is an intensive manual process that requires expertise in classifying and counting parasitized and uninfected cells. This process may not scale well, especially in regions where the right expertise is hard to find. Some advancements have been made in leveraging state-of-the-art image processing and analysis techniques to extract hand-engineered features and build machine learning-based classification models. However, these models are not scalable with more data being available for training and given the fact that hand-engineered features take a lot of time.

Deep learning models, or more specifically convolutional neural networks (CNNs), have proven very effective in a wide variety of computer vision tasks.

Convolution layers learn spatial hierarchical patterns from data, which are also translation-invariant, so they are able to learn different aspects of images. For example, the first convolution layer will learn small and local patterns, such as edges and corners, a second convolution layer will learn larger patterns based on the features from the first layers, and so on. This allows CNNs to automate feature engineering and learn effective features that generalize well on new data points. Pooling layers helps with downsampling and dimension reduction.

Thus, CNNs help with automated and scalable feature engineering. Also, plugging in dense layers at the end of the model enables us to perform tasks like image classification. Automated malaria detection using deep learning models like CNNs could be very effective, cheap, and scalable, especially with the advent of transfer learning and pre-trained models that work quite well, even with constraints like less data.

**. Defining Problem .**

Classification involves predicting which class an item belongs to. Some classifiers are binary, resulting in a yes/no decision. Others are multi-class, able to categorize an item into one of several categories. Classification is a very common use case of machine learning—classification algorithms are used to solve problems like email spam filtering, document categorization, speech recognition, image recognition, and handwriting recognition.

In this context, a neural network is one of several machine learning algorithms that can help solve classification problems. Its unique strength is its ability to dynamically create complex prediction functions, and emulate human thinking, in a way that no other algorithm can. There are many classification problems for which neural networks have yielded the best results.

Artificial Neural Networks and Deep Neural Networks are effective for high dimensionality problems, but they are also theoretically complex. Fortunately, there are deep learning frameworks, like TensorFlow, keras, Pytorch, that can help you set deep neural networks faster, with only a few lines of code. You can also use deep learning platforms like to run and manage deep learning experiments automatically.

Artificial neural networks are built of simple elements called neurons, which take in a real value, multiply it by a weight, and run it through a non-linear activation function. By constructing multiple layers of neurons, each of which receives part of the input variables, and then passes on its results to the next layers, the network can learn very complex functions. Theoretically, a neural network is capable of learning the shape of just any function, given enough computational power.

Very effective for high dimensionality problems , able to deal with complex relations between variables, non-exhaustive category sets and complex functions relating input to output variables. Powerful tuning options to prevent over- and under-fitting.

The convolutional neural network (CNN) is a class of **Deep Learning neural networks**. CNNs represent a huge breakthrough in image recognition. They’re most commonly used to analyze visual imagery and are frequently working behind the scenes in image classification. They can be found at the core of everything from Facebook’s photo tagging to self-driving cars. They’re working hard behind the scenes in everything from healthcare to security.

Image classification is the process of taking an input (like a picture) and outputting a class (like “class 1”) or a **probability** that the input is a particular class (“there’s a 90% probability that this input is a class 1”). The convolutional neural networks learns the most relevant features of an image which will help in classification of image according to the probability of occurring of image in each class.

Deep Learning models, or to be more specific, Convolutional Neural Networks (CNNs) have proven to be really effective in a wide variety of computer vision tasks. CNNs help us with automated and scalable feature engineering. Also, plugging in dense layers at the end of our model enables us to perform tasks like image classification. Automated malaria detection using deep learning models like CNNs could be very effective, cheap and scalable especially with the advent of transfer learning and pre-trained models which work quite well even with constraints like less data.

**. Software Specifications .**

To build deep learning models we need a good amount of training data but we also need to test the model’s performance on unseen data. Deep Learning models, or to be more specific, Convolutional Neural Networks (CNNs) have proven to be really effective in a wide variety of computer vision tasks.

**1.) The dataset**

The data for our analysis comes from researchers at the Lister Hill National Center for Biomedical Communications (LHNCBC), part of the National Librtary of Medicine (NLM), who have carefully collected and annotated the [publicly available datase](https://ceb.nlm.nih.gov/repositories/malaria-datasets/) of healthy and infected blood smear images. They used Giemsa-stained thin blood smear slides from 150 *P. falciparum*-infected and 50 healthy patients, collected and photographed at Chittagong Medical College Hospital, Bangladesh. The smartphone's built-in camera acquired images of slides for each microscopic field of view. The images were manually annotated by an expert slide reader at the Mahidol-Oxford Tropical Medicine Research Unit in Bangkok, Thailand.

**2.) Training Specifications**

I am using a Google Colaboratory (Colab) system on the cloud having a GPU so I can run my models faster. Colab is a Python development environment that runs in the browser using Google Cloud.  Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing free access to computing resources including GPUs. The Ram Portion of 25 GB is allocated by the Google colab for intense numerical computation.

**3.) Library(modules) Specifiactions**

* **Scikit Learn** :

Scikit-learn (sklearn) is probably the most useful library for machine learning in Python. The sklearn library contains a lot of efficient tools for Data Preprocessing and statistical modeling including Normalization, Standardization, Splitting and dimensionality reduction.

* **Matplotlib and Seaborn** :

These are Data visualization Library used for analysis for Different types of Plots based on different formats of Data. It includes Barplot, Scatter Plot, Pie Chart, Stacked Plot, Input Data visualization, Output metrics visualization, Result evaluation etc..

* **Keras (Tensorflow functional API)** :

Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, R, Theano, or PlaidML. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible.

* **Numpy** :

NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python.

* **Scipy**:

SciPy is a library that uses NumPy for more mathematical functions. SciPy usesNumPy arrays as the basic data structure, and comes with modules for various commonly used tasks in scientific programming, including linear algebra, integration (calculus), ordinary differential equation solving, and signal processing.

**. Software Designs .**

Start

Loading Data

Cleaning Data

Data Pre-processing

Feature Extraction

Model

Selection

Custom Model VGG 16 Model

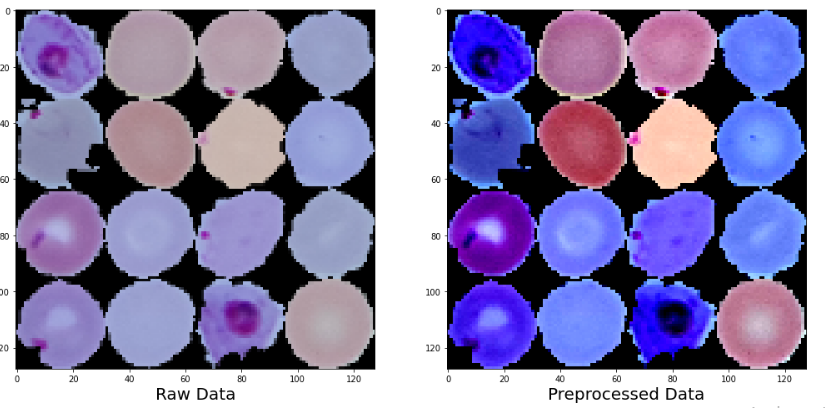
**Ensemble Model**

**Paracitized** **Prediction Healthy Control**

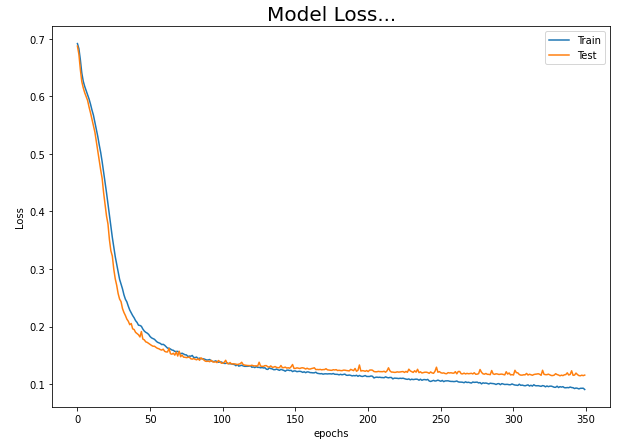
**Model Evaluation and Visualization .**

**1.)** Pre-processed Data :

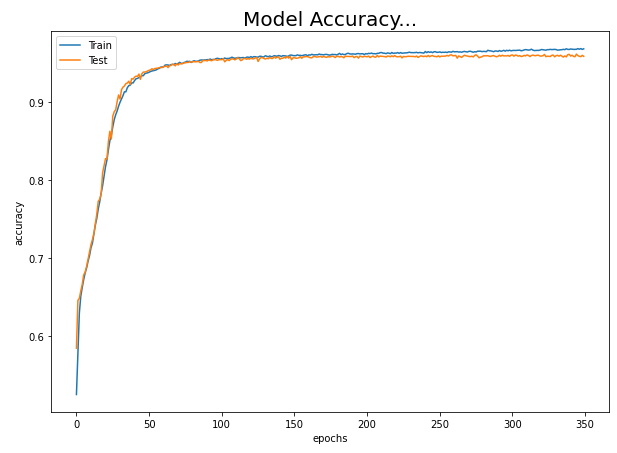
As we can notice the difference between Raw an Pre-processed Data that how the important features for classifications are highlighted after pre-processing which will make the classification results more accurate.



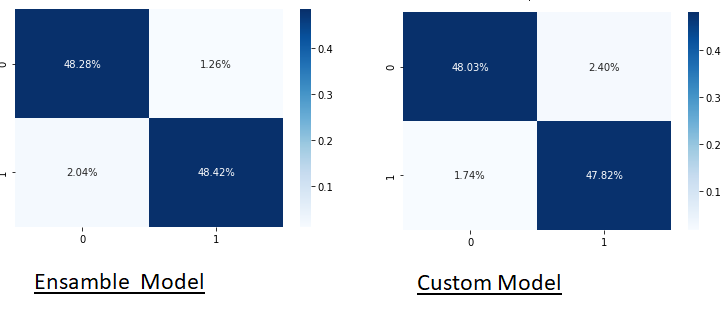
**2.)** Loss Evaluation :

****

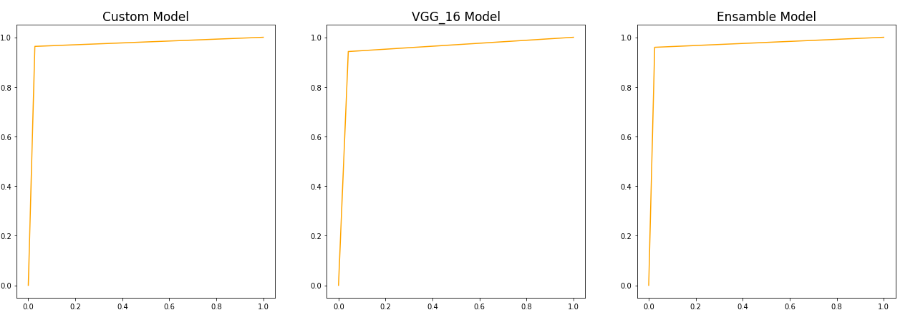
**3.)** Accuracy Evaluation :



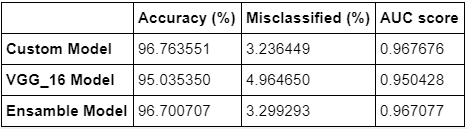
**3.)** Confusion Matrix :



**4.)** ROC - Curve :



**5.)** Comparisons :



These comparisons will show all the performance matrices for judgement of different models. Here we can see Accuracy Parameter, Miss-classification Rate, AUC i.e. Area Under The ROC curve. These parameters are enough to evaluate the model performance on our input Data

**Conclusion .**

We looked at an interesting real-world medical imaging case study of malaria detection in this article. Malaria detection by itself is not an easy procedure and the availability of the right personnel across the globe is also a serious concern. We looked at easy to build open-source techniques leveraging AI which can give us state-of-the-art accuracy in detecting malaria thus enabling AI for social good. I encourage everyone to check out the articles and research papers mentioned in this article, without which it would have been impossible for me to conceptualize and write this article. Let’s hope for more adoption of open-source AI capabilities across healthcare making it cheaper and accessible for everyone across the world.

**References .**

1.) Bibin, Dhanya, Madhu S. Nair, and P. Punitha. "Malaria parasite detection from peripheral blood smear images using deep belief networks." IEEE Access 5 (2017): 9099- 9108.

2.) Towards Data science blog : <https://towardsdatascience.com/detecting-malaria-with-deep-learning-9e45c1e34b60> .

3.) [Very Deep Convolutional Networks for Large-Scale Image Recognition](https://arxiv.org/pdf/1409.1556.pdf)

4.) K. He, X. Zhang, S. Ren and J. Sun, “Deep residual learning for image recognition.” arXiv preprint, arXiv: 1512.03385, 2015 Dec.

5.) D.K. Das, A.K. Maiti, and C. Chakraborty, “Automated system for characterization and classification of malaria-infected stages using light microscopic images of thin blood smears.” Journal of Microscopy, 2015, 257(3), pp. 5140-5144