MALARIA DIAGNOSIS USING CONVOLUTIONAL NEURAL NETWORK

Chhaya Vankhede[1]

[1]Student, CE dept., SAL Institute of Technology Engineering and Research, Ahmedabad, Gujarat, India.

Abstract:

This study has been done to make the diagnosis of malaria easy, faster and less expensive. Conventional microscopy is one of the best ways for diagnosis of the disease. These techniques including other techniques are time-consuming, expensive and require expert pathologists. Thus in this work semi-automatic malaria diagnosis system is developed which uses stained thin blood smear images. We have developed this kind of system using Convolutional Neural Network(CNN), which is trained and tested using 20000 images of the stained blood smear. This system provides 96% accuracy on test data.

Keywords:

Malaria Diagnosis, Machine Learning, CNN, Image classification, Accuracy, *Parasite.*

1. INTRODUCTION

According to WHO in 2017 malaria caused the death of 435000 people. More than two-thirds of all malaria deaths occur in under 5 years of age group [1].Malaria is a disease caused by Plasmodium parasites that are transmitted to people by the bite of infected female Anopheles mosquitoes. There are several techniques used to diagnose malaria in laboratory 1)Microscopic diagnosis using stained thin and thick peripheral blood smears (PBS): Malaria is diagnosed microscopically by staining thick and thin blood films on a glass slide, to visualize malaria parasites, 2)QBC technique: In this technique finger-prick blood is collected in a hematocrit tube containing acridine orange and anticoagulant. The tube is centrifuged at 12,000 g for 5 min and immediately examined using an epi-fluorescent microscope, 3) Rapid diagnostic tests (RDTs): RDTs are all based on the same principle and detect malaria antigen in blood flowing along a membrane containing specific anti-malaria antibodies, 4)Serological tests[2] : Immunofluorescence antibody testing (IFA) has been a reliable serologic test for malaria in recent decades. IFA is simple and sensitive, but time-consuming. It cannot be automated, which limits the number of sera that can be studied daily. It also requires fluorescence microscopy and trained technicians; readings can be influenced by the level of training of the technician, particularly for serum samples with low antibody titers.

The proposed system is designed to help people diagnose malaria without any special knowledge. As a manual examination of a thin blood smear for detecting parasite in it requires experience and knowledge. It allows the user to capture the image using the camera with an embedded microscope. The system will predict the image and will show the parasites in blood smear if present and can also generate a report with the patient's details. It uses Image classification concepts to recognize Plasmodium[3] in the blood smear. The user will require a FoldScope [4][5] or a microscope[6] and a mobile application to take pictures of the blood sample.

2. LITERATURE SURVEY

Prior state of the art reveals lots of work is done in this field of diagnosis of disease using different techniques of Artificial Intelligence.

In [7], the author has proposed a technique to detect malaria parasites in the blood smear. In this technique Images of infected and non-infected erythrocytes were acquired, pre-processed, relevant features extracted[8] from them and eventually, the diagnosis was made based on the features extracted from the images. They are using an artificial neural network classifier[9]. The system model is implemented using six main processes, namely; image acquisition, image preprocessing, image segmentation[10], feature extraction, comparison, and classification. The system recorded 99.68 % accuracy in detecting the presence of Plasmodium parasites.

Another comparative study[11] is about analyzing the efficient method of malaria diagnosis by implementing the work using ANN and CNN. This system includes detection of parasite life stages using SVM multi-classifier to the previous study[7]. They stated that SVM, ANN, and CNN gives the accuracy of 0.9434%, 0.9620%, and 0.9820% respectively.

Another work[12] presents an overview of different Artificial Intelligence technique with a review of important clinical applications. The artificial neural network, fuzzy expert systems[13], evolutionary computation[14] and hybrid intelligent systems are included in this research paper. Neural networks are mainly concerned with learning, fuzzy logic with imprecision and evolutionary computation with search and optimization[15].Genetic algorithms[16] exploit the mechanism of natural evolution to search efficiently in a given

space. They can be applied to the task like diagnosis, prognosis, medical imaging and signal processing, and planning and scheduling. Genetic Algorithm can be used for automatically training and generating neural network architectures.ANN model which was subsequently validated in prospective studies had a diagnostic accuracy of 90%, with a sensitivity of 81% and specificity of 92%.
They state that Fuzzy logic performed better than multiple logistic regression analysis in diagnosing lung cancer using tumor marker profiles, to predict survival in patients with breast cancer.

Quinn, John A., et al.[17] has proposed deep convolutional neural network which is used for three different microscopic tasks: diagnosis of malaria, tuberculosis in sputum samples, and intestinal parasite eggs in stool samples. Used CNN model of four hidden layers with specification of 1) Convolution layer: 7 filters of size 3×3. 2) Pooling layer: max-pooling 3) Convolution layer: 12 filters of size 2×2. 4) Fully connected layer, with 500 hidden units gives very high accuracy.
Compared to this research our system has CNN model with 10 layers of combination of convolutional layer, max pooling layer , flatten layer, dense layer and dropout layer with different specification

Another research[18] includes hardware. The Processors analyze the signals to differentiate between P.falciparum malaria likely or unlikely at risk for severe/cerebral malaria, non-P falciparum malaria, and infection. A test panel and kit is used in the diagnosis of malaria and severe bacterial infection in a test sample. They include monoclonal antibodies, histidine-rich protein H5, P. falciparumaldolase and/or pan P.falciparum lactate dehydrogenase, Angiopoietin-1 and/or Angiopoietin-2, and C-reactive protein and/or procalcitonin. Processors analyze the signals to differentiate between P.falciparum malaria likely or unlikely at risk for severe/cerebral malaria, non-P falciparum malaria, and infection which is likely or unlikely to be a severe bacterial infection, in the sample.

As we are using the Convolutional Neural Network, the feature map is generated by the convolution layer. Our system gives 96% accuracy on very balanced data of 50% non-infected images and 50% infected images while training and testing the model. The model does not require any preprocessing before feeding the data to the model. We are using the mobile phone camera for taking an image of the blood sample with the help of a microscope or foldscope. Our system can only detect Plasmodium but can't classify the type of it.

3.   THE PROPOSED SYSTEM

In the proposed methodology Convolutional Neural Network is used for the detection of malaria parasites from microscopic images of Giemsa stained thin blood smear.
The blood smear slid can be prepared by cleaning finger with 70% ethyl alcohol, then allow it to dry and then the side of the fingertip is picked and a drop of blood is placed on a glass slide. A thin blood film is prepared by placing the smooth edge of a spreader slide at 45° and then smearing the blood with a swift and steady sweep along the surface. The slide is then allowed to air dry and is fixed with absolute methanol. After drying, the sample is stained with diluted Giemsa (1:20, vol/vol) for 20 min and washed by briefly dipping the slide in and out of a jar of buffered water. The slide is then allowed to air-dry in a vertical position and examined by the system with the help of a microscope.

Fig.1. shows the overall modules of the proposed system. It comprises of IMAGE DATA PREPARATION, CNN, IMAGE, UI, Report, Database module. We now briefly explain each of these modules:
1.   **IMAGE DATA PREPARATION:** This part of the system makes patches of images. It includes patches made from annotation files and images for positive patches and negative patches for training and testing of the CNN model. It also makes patches of the image that will be given for prediction. Fig.2. shows the image patches generated by this module. In this image 0 labeled images show the absence of parasites and 1 labeled images show the presence of the parasite.
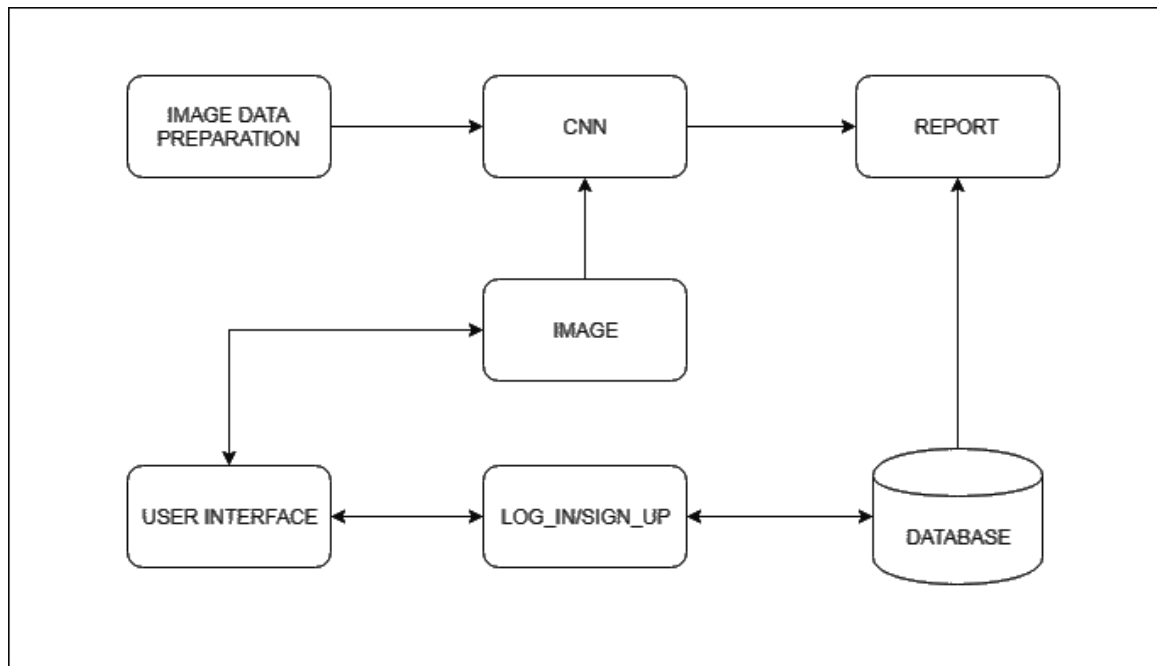
Fig.1. Block diagram of the proposed system

Fig.2. Input patches generated by IMAGE DATA PREPARATION

2. **CNN:** In this, the model is generated using CNN. It is a CNN model with 10 convolution layer. Fig.3. Shows the architecture of the CNN model. RELU and sigmoid are the activation functions used here.

Conv2D:convolution mixes one function with another to reduce data space while preserving the information. Generates feature map using activation function.

MaxPooling2D: It down-sample the image by applying a max filter.

Flatten: It converts input from the previous layer to 1D vector and feeds into Dense layer

Dropout: It is a regularization technique to prevent CNN from over fitting.

ReLU(rectified linear unit) and sigmoid are the activation functions used here.

ReLU is defined as a positive part of its argument. It is the most popular activation function of 2018 for the deep neural network as it simple and gives better results compared to other activation functions. It can be defined by the below formula.

$$f(x) = x^+ = \max(0, x)$$

Sigmoid is used in the output layer. Sigmoid gives output between 0 and 1. Sigmoid is a special case of logistic function. Sigmoid is defined for all real input values. The mathematical formula for sigmoid is defined below.

$$S(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}$$

Binary cross-entropy is used to measure the performance of the model as the model does binary classification whose output is a probability of a value between 0 and 1. Define in the below formula.

$$BCE = -\frac{1}{N}\sum_{i=1}^{N} y_i * \log\big(p(y_i)\big) + (1 - y_i) * \log(1 - p(y_i))$$

Where BCE = binary cross-entropy, y is label and p(y) is predicated probability of the input for all N points.

3. **IMAGE:** This module will allow the user to capture the image of the blood smear under microscope or select it from storage. The user can edit and crop this image if needed.

4. **UI and LOG_IN/SIGN_UP:** User Interface to operate the system and User authentication, sign up and login of the user will be handled by this module

5. **REPORT:**
   The output from the model will be converted into a report and will be stored in the user's account.

6. **DATABASE:**
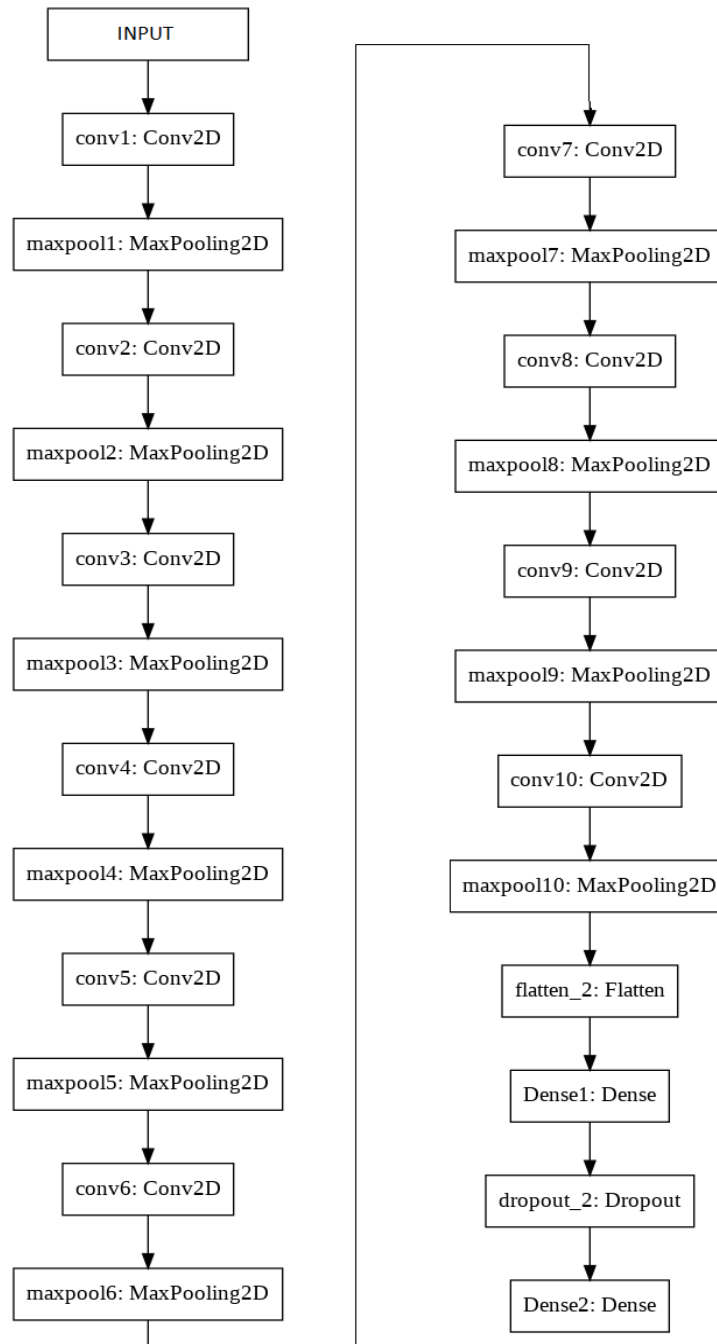   User's information and authentication details will be stored in the database.

Fig.3. Architecture of CNN

4. EXPERIMENTAL RESULTS AND ANALYSIS

Experiments are performed on the system with the configuration of Windows 7, 4GB RAM, 1TB Hard disk using Spyder3IDE, Google colaboratory, MySQL. Here used library and packages arePandas, Numpy, OpenCV, Matplotlib, Keras, lxml, OS, SKlearn,Seaborn, Tensorflow. The datasetof 20000 image patches of the thin blood smear is used for training andtesting the CNN model [19]. The experiment shows that the selection of higher dataset size,higher number of filters, lower learning rate increase the accuracy of the model.

We have performed experiments to measure the accuracy y with varying values of the epoch.Fig.4. and Fig.5. show the change in accuracy and loss with every epoch.
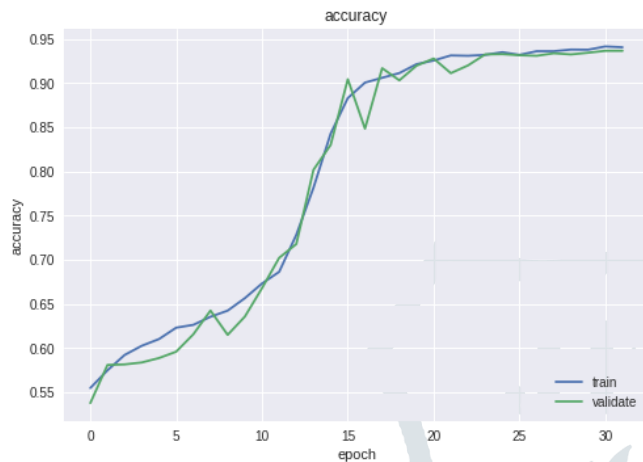

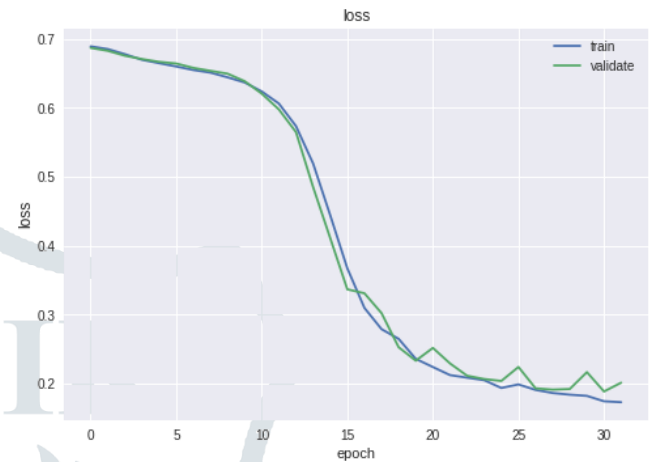Fig.4. The accuracy of the model with epochs   Fig.5. Loss with epochs

Experiments are generated using 12800 train images and 3200 validation images. From these figures, we can conclude that the model is generalized and trained properly as a train and validation graph in both the figure are similar.
The figure shows that after 25th epoch changes in both loss and accuracy are minor.
Table 1 represents the accuracy achieved by changing the number of convolution layers and other parameters. Model's performance changes as we change the size of the dataset to train, higher the size of the dataset higher the performance.

Table 1

No. of convolution layers	Size of train dataset	epochs	Accuracy	Loss
Experimental results of CNN				
10	11200	250	0.9602	0.1182
10	12800	32	0.943	0.1959
10	6400	16	0.7195	0.5778
5	6400	16	0.89275	0.3904

5. CONCLUSION AND FUTURE SCOPE

The proposed system diagnoses malaria by examining the microscopy image of the blood sample. The System generates results using CNN which can be improved by changing its parameters and size of the training dataset. This study shows that the accuracy of the system depends on the size of the dataset, the use of optimizer, training epochs and architecture of the model. Some other techniques of machine learning can also be used to prepare model but among all of those techniques, CNN gives better results as this prediction and classification are based on image dataset.
The proposed system is for everyone who is curious about exploring their knowledge and also for medical students, pathology students, researchers, and patients. Anyone can easily operate this system and can have immediate results at home with some preparations. The proposed system saves time, money and effort of the user.

The proposed system can be improved in the future to classify the type of parasites and for the diagnosis of other diseases as well.

REFERENCES:

1. [Online] [Cited: 2 5, 2019.] https://www.who.int/news-room/fact-sheets/detail/malaria.

2. [Online] [Cited: 1 22, 2019.] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2688806/.

3. [Online] [Cited: 1 22, 2019.] https://en.wikipedia.org/wiki/Plasmodium_falciparum.

4. [Online] [Cited: 1 22, 2019.] Cybulski, J.S., Clements, J. and Prakash, M., 2014. Foldscope: origami-based paper microscope. *PloS one*, *9*(6), p.e98781.

5. [Online] [Cited: 1 22, 2019.] https://www.amazon.in/Foldscope-Instruments-Inc-Deluxe-Individual/dp/B077KMXW9C.

6. [Online] [Cited: 1 22, 2019.] https://sciencing.com/different-kinds-microscopes-uses-5024481.html.

7. Bashir, A., Mustafa, Z.A., Abdelhameid, I. and Ibrahem, R., 2017, January. Detection of malaria parasites using digital image processing. In Communication, Control, Computing and Electronics Engineering (ICCCCEE), 2017 International Conference on (pp. 1-5). IEEE.8. [Online] [Cited: 1 22, 2019.] https://medium.com/@mehulved1503/feature-selection-and-feature-extraction-in-machine-learning-an-overview-57891c595e96.

9. [Online] [Cited: 1 22, 2019.] https://ai.stackexchange.com/questions/5546/what-is-the-difference-between-a-convolutional-neural-network-and-a-regular-neur.

10. [Online] [Cited: 1 22, 2019.] Pal, N.R. and Pal, S.K., 1993. A review on image segmentation techniques. *Pattern recognition*, *26*(9), pp.1277-1294.

11. Das, Soumya, P, Sony and L, Jyothi R., 2018, May. AUTOMATED IDENTIFICATION & CLASSIFICATION OF MALARIAL PARASITES IN THIN BLOOD SMEAR IMAGES. College of Engineering, Chengannur, Kerala, India. IRJET12. Ramesh, A.N., Kambhampati, C., Monson, J.R. and Drew, P.J., 2004. Artificial intelligence in medicine. Annals of The Royal College of Surgeons of England, 86(5), p.334.

13. [Online] [Cited: 1 22, 2019.] Hayes-Roth, F., Waterman, D.A. and Lenat, D.B., 1983. Building expert system.

14. [Online] [Cited: 1 22, 2019.] https://en.wikipedia.org/wiki/Evolutionary_computation.

15. [Online] [Cited: 1 22, 2019.] https://medium.com/@devnag/machine-learning-representation-evaluation-optimization-fc7b26b38fdb.

16. [Online] [Cited: 1 22, 2019.] https://www.tutorialspoint.com/genetic_algorithms/genetic_algorithms_introduction.htm.

17. Quinn, J.A., Nakasi, R., Mugagga, P.K., Byanyima, P., Lubega, W. and Andama, A., 2016, December. Deep convolutional neural networks for microscopy-based point of care diagnostics. In *Machine Learning for Healthcare Conference*(pp. 271-281).

18. Friedland, Susan, and Kain, Kevin,2010. Diagnostic test panel for the diagnostic of malaria and severe bacterial infections,WO2010099607A1.
https://patents.google.com/patent/WO2010099607A1/en?q=Diagnostic+test&q=panel&q=diagnostic&q=malaria&q=severe&q=bacterial+infections&oq=Diagnostic+test+panel+for+the+diagnostic+of+malaria+and+severe+bacterial +infections

19. [Online] [Cited: 9 28, 2018.] https://www.microscopyu.com/gallery-images/plasmodium-vivax-infection-at-40x-magnification.