MALARIA DETECTION USING CONVOLUTIONAL NEURAL NETWORK

Ch. Sujith[1], G. Ramya[2] ,B. Lekhanaa Jaya Sree[3] ,G. Teja[4], T.Nagamani[5]

[1]*Student, Department of Computer Science & Engineering, Gudlavalleru Engineering College, India*
[2]*Student, Department of Computer Science & Engineering, Gudlavalleru Engineering College, India*
[3]*Student, Department of Computer Science & Engineering, Gudlavalleru Engineering College, India*
[4]*Student, Department of Computer Science & Engineering, Gudlavalleru Engineering College, India*
[5]*Assistant Professor, Department of Computer Science & Engineering, Gudlavalleru Engineering College, India*
chinnamsujith1999@gmail.com,gramya0811@gmail.com,lekhanaaboyina17@gmail.com,gallateja8@gmail.com , tenalinagamani@gmail.com

ABSTRACT

Malaria is a mosquito-borne infectious disease that affects humans and other animals. Malaria causes symptoms that typically include fever, tiredness, vomiting, and headaches. In severe cases it can cause yellow skin, seizures, coma, or death. Malaria parasites can be identified by examining under the microscope a drop of the patient's blood, spread out as a "blood smear" on a microscope slide. By using machine learning we can detect whether the person is affected with malaria or not.

We train the model by giving malaria parasite images and uninfected cell images as input in the form of a dataset and then we predict whether the person is infected with malaria or not.

Keywords: Malaria, Plasmodium parasites, Blood smears, Thin films, Convolutional Neural Network, Image Processing, Automated Diagnosis Training, Testing.

I.INTRODUCTION:

Malaria is caused by Plasmodium parasites. The parasites are spread to humans through the bites of infected female Anopheles mosquitoes called "Malaria Vectors." There are 5 parasite species that cause malaria in humans, and a pair of these species – P.Falciparum and P. Vivax – it's the best hazard. Children with severe malaria frequently develop one or more of the subsequent symptoms: severe anemia, respiratory distress in reference to acidosis, or cerebral malaria. In adults, multi-organ failure is additionally frequent[1].People can develop partial immunity in malaria-endemic areas, allowing asymptomatic infections to occur. Malaria is an acute febrile illness. in a very non-immune individual, symptoms usually appear

10–15 days after the infective sting. the primary symptoms – fever, headache, and chills – is also mild and difficult to acknowledge as malaria. If not treated within 24 hours, P. falciparum malaria can accomplish severe illness, often resulting in death[2].

According to the report released by the globe Health Organization (WHO), there have been 214 million cases of malaria in 2015 and 438,000 deaths. In most cases, malaria can only be diagnosed by a manual examination of the microscopic slide. Whole slide imaging, which scans the traditional glass slides so as to supply digital slides, is that the most up-to-date imaging modality being employed by pathology departments worldwide. so as to supply a reliable diagnosis, necessary training and specialized human resource are required. Unfortunately, these resources are off from being adequate and regularly often unavailable in underdeveloped areas where malaria encompasses a marked predominance. Therefore, an automatic diagnostic system can provide an efficient solution to the current problem[1]. plasmodium in blood samples is identified by using image segmentation and have extraction.Supported Image Acquisition, Image Preprocessing, Image Smoothing, and Image segmentation is finished. A huge amount of work has been done to extract the features of malaria-infected cells. A very survey on the extraction and optimization of features of malaria cells has been well discussed. While a good feature extraction system will increase detection accuracy, this form of infection detection can not be completely automated as it still requires qualified experts to manually extract the feature vectors for the particular datasets.

Feature extraction uses two phases within the architectural model:

1) Training Phase

2) Testing (Recognition) Phase which helps to acknowledge the Plasmodium vivax. Recently,

machine learning algorithms have attracted great attention from researchers for their ability to develop an automated malaria diagnostic system.. In SVM and Naïve Bayes Classifiers were utilized to attain accuracies of 84% and 83.5% respectively. But here we use a deep convolutional neural network (CNN) which provides an accuracy of above 92% was applied to diagnose malaria within the thick blood smear. However, to tell apart infected and non-infected samples in thick films is actually difficult for pathologists, because the difference isn't as clear as those individual red blood cells cropped from whole slide images supported thin films.

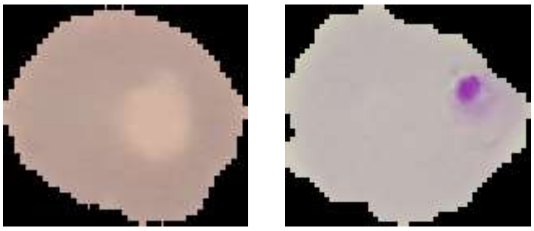

Fig 1(a): Infected Malaria Blood Smear
Fig 1(b): Uninfected Malaria Blood Smear

In this work, we focus

1) Automated detection and quantification of malaria detection
2) To determine infected image using machine learning
3) To improve the Predictive value

II.OBJECTIVES OF THE PAPER:

The main objective of the Malaria Detection is to detect whether the person is infected with malaria or not.

i. To define the rational use of microscopy and automated diagnostic system for malaria control.
ii. To identify factors that determine the choice of approaches to the diagnosis of malaria.
iii. To define the desired specifications for new diagnostic tests.

Several approaches to the diagnosis of malaria (defined for the purpose of this document as a disease caused by infection with malaria parasites) can be adopted. Each approach presents characteristics such as cost, ease of performance, and accuracy, which will determine its applicability to different situations. The approaches are[14]

i. Clinical diagnosis
ii. Microscopy Approach.
iii. Automated Diagnostic System.

i. Clinical Diagnosis:

Clinical diagnosis is that the most generally used approach. it's been the sole feasible one in many situations, particularly in rural areas and at the periphery of the health care system where laboratory support to clinical diagnosis does n't exist. Among the numerous clinical signs and symptoms related to malaria, the foremost prominent is fever, which is commonly in the midst of chills, perspiration, anorexia, headaches, vomiting, and malaise. Residents of endemic areas are often at home with this combination of symptoms and regularly self-diagnosed malaria symptoms are supported on their own. additionally to those symptoms of uncomplicated malaria, other manifestations may develop that signal severe malaria, which is nearly always thanks to Plasmodium falciparum. These include confusion or drowsiness with prostration along with severe manifestations like cerebral malaria, severe anemia, and others. Clinical diagnosis is inexpensive to perform, and requires no special equipmen t or supplies. However, the symptoms of malaria are very non-specific and overlap with those of other febrile illnesses. A diagnosis of malaria supported clinical grounds alone is therefore unreliable, and when possible should be confirmed by laboratory tests. In spite of this lack of specificity, in some settings disease management supported clinical diagnosis alone is justifiable

ii. Microscopy Approach:

Conventional light microscopy is the viable approach for the laboratory confirmation of malaria. Careful examination of a well-prepared and well-stained blood film by an expert microscope remains the "gold standard" for the detection and identification of malaria parasites.. In most settings, the procedure consists of: collecting a finger-prick blood sample; preparing a thick blood smear (in some settings a skinny smear is additionally prepared); staining the smear (most frequently with Giemsa) and examining the smear through a microscope for the presence of malaria parasites. it is used by skilled and careful technicians; microscopy can detect densities as low as 5–10 parasites per µl of blood

iii. Automated Diagnostic System:

There are many algorithms to implement the Automated Diagnostic System like SVM, CNN, and e.t.c. The dataset which contains the Plasmodium vivax blood cells are categorized into infected and uninfected groups. By taking the assistance of those algorithms test and train the photographs and detect the output..

III.PROBLEM STATEMENT:

Malaria could be a deadly, infectious, mosquito-borne disease caused by Plasmodium parasites that are transmitted by the bites of infected female Anopheles mosquitoes. Five parasites cause malaria, but two types—P. falciparum and P. vivax—cause the bulk of the cases. If an infected mosquito bites you, parasites carried by the mosquito enter your blood and begin destroying oxygen-carrying red blood cells (RBC). Typically, the primary symptoms of malaria are almost like a pestilence just like the flu and that they usually begin within some days or weeks after the sting. However, these deadly parasites can board your body for over a year without causing symptoms, and a delay in treatment can cause complications and even death. Therefore, early detection can save lives[3].This project is principally about detecting the presence of malaria parasites in your body by collecting the pictures of blood smears. By using the algorithms we conclude that whether the person is infected or not

IV.METHODOLOGY:

The nature of the difficulties likely to be encountered during the method of turning out with an data system has led to the innovation of methods to enhance the knowledge management system. to urge to the planning and implement an data system, there's a requirement to own a methodological approach so as to fulfill the time against physical and human constraints.A system development methodology refers to the steps that are used to form, plan, and control the strategy of developing a system since it's virtually impossible to drive forward a project to computerize method without prior work. The methodology employed in this project is image classification using CNN.

Image Classification:

Image classification refers to a process in computer vision which will classify a picture consistent with its visual content. for instance, a picture classification algorithm is also designed to inform if a picture contains a person's figure or not. While detecting an object is trivial for humans, robust image classification continues to be a challenge in computer vision applications.

Image classification could be a supervised learning problem: define a collection of target classes (objects to spot in images), and train a model to acknowledge them using labeled example photos. Early computer vision models relied on raw pixel data because the input to the model. The raw pixel data itself cannot provide a sufficiently stable representation to enlarge the myriad variations of an object as captured in a picture. The position of the article, the background behind the article, ambient lighting, point of view, and camera focus all can fluctuate with raw pixel data; these differences are significant enough that they can't be corrected for by taking weighted averages of pixel RGB values.[4]

Architecture:

A machine learning algorithm takes examples of inputs and outputs associated with a task and produces an output that can automatically differ them. The below will show the architecture used for this research[2][3].

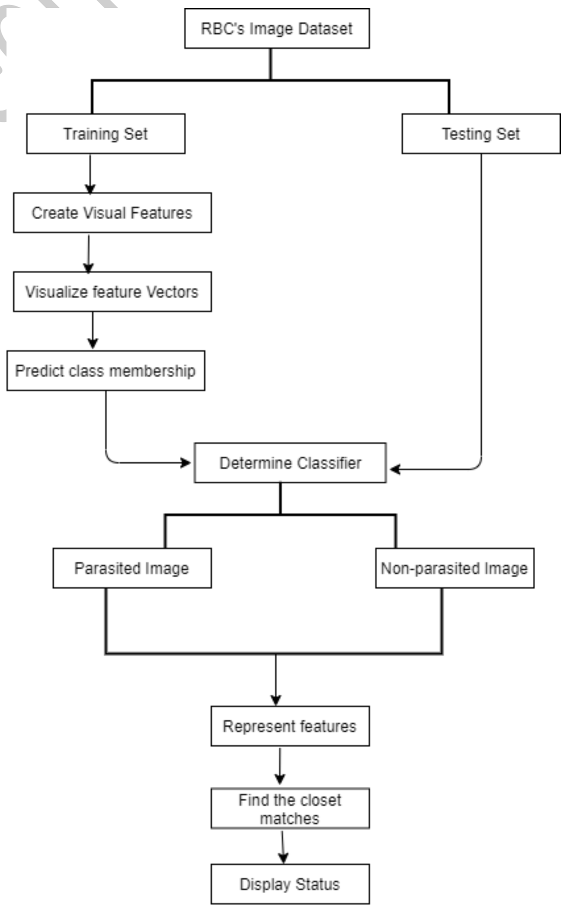

Fig-2: Architecture

Convolutionl Neural Network algorithm:

Convolutional Neural Network (**ConvNets** or **CNNs**) are a category of Neural Networks that have proven very effective in areas such as image recognition and classification. ConvNets are successful in identifying faces, objects, and traffic signs aside from powering vision in robots and self-driving cars.

It is a multilayer perceptron that's the special design for the identification of two-dimensional image information. Always it has more layers: input layer, convolution layer, sample layer, and output layer. Also, in deep network architecture the convolution layer and sample layer can have multiple. CNN is not as restricted Boltzmann machine, it needs to be before and after the layer of neurons in the adjacent layer for all connections, convolution neural network algorithms, each neuron don't need to do feel the global image, just feel the local area of the image. Also, each neuron parameter is set to the same, namely, the sharing of weights, namely each neuron with the same convolution kernels to deconvolution image. CNN algorithm has two main Processes: convolution and sampling[4].

Convolution process: use a trainable filter Fx, deconvolution of the input image(the first stage is that the input image, the input of the after convolution is that the feature image of every layer, namely feature Map), then add a bias bx, we will get convolution layer Cx.

A sampling process: n pixels of each neighborhood through pooling steps, become a pixel, and then by scalar weighting Wx + 1 weighted, add bias bx + 1, and then by an activation function, produce a narrow n times feature map Sx + 1[3].

The key technology of CNN is the local receptive field, sharing of weights, sub sampling by time or space, to extract feature and reduce the size of the training parameters.

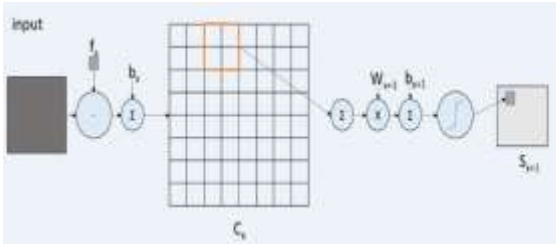

Fig-3: Main Process of CNN

The advantage of CNN algorithm is that to avoid the explicit feature extraction, and implicitly to learn from the training data; the same neuron weights on the surface of the feature mapping, thus network can learn parallelly, reduce the complexity of the network; Adopting sub-sampling structure by time or space, can achieve some degree of robustness, scale and deformation displacement; Input information and network topology can be a very good match, It has unique advantages in speech recognition and image processing[15].

CNN algorithm needs experience in architecture design, and need to debug unceasingly in the practical application, to obtain the most suitable for particular application architecture of CNN. Based on the gray image as the input of 96 x 96, in the preprocessing stage, turning it into 32 x 32 of the size of the image. Desin depth of the layer 7 convolution model: input layer, convolution layer C1, sub-sampling layer S1, convolution layer C2, sampling layer S2,hidden layer H, and output layer F.

The input image can contain multiple channels such as color, wings, eyes, beak of birds which means that the convolution layer performs a mapping from 3D volume to another 3D volume. 3D volumes considered are width, height, depth. The CNN has two components:

1) **Feature extraction part:** features are detected when the network performs a series of convolution and pooling operation.

2) **Classification part:** extracted features are given to a fully connected layer which acts as classifier[8].

Data Modeling:

Dataset may be a collection of related sets of data that's composed of separate elements but may be manipulated as a unit by a computer. the knowledge set contains 9759 lines, each of them representing either a default or not a default (binary value) of a risk once they raise a loan from a bank[10]. Default and good loans are characterized by the 0 or 1 labeled variables that are directly obtained from the businesses, financial statements, balance sheets, income statements, and income statements where the values are considered at all-time low level of granularity. After importing the info, we cleaned the variables and removed features with no pertinent information. Then, we split the info into two subsets, considering 80% of the info then 20% of those data was used for test purposes. The validation performance permits one to boost the training approach, and that we use it to supply

Vol 11, Issue 6,June/2020
ISSN NO:0377-9254

prediction performance on the test set. With in the training set, we verify if it's a balanced dataset or not. Data mining methodology starts with data preprocessing. Data preprocessing may be a data processing technique that involves transforming information into a lucid format. Real-world data is commonly incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is probably going to contain many errors. Data preprocessing consists of the steps:

Data Preprocessing:

The original image data for the entire slide contains a large amount of redundant material.In order to achieve good classification accuracy, image segmentation and de-noise must be performed in order to extract only the blood cells and remove the redundant pixels at the same time. Before performing threshold operation, each and every image tile was converted to grayscale image from the color space.However, noise as the byproduct of thresholding can degrade the quality of the acquired images and lower the classification performance. By performing image morphological operations which present in the matlab removes the noise pixels in the image. Another significant concern is that segmented whole slide images almost inevitably have a lot of red blood cells (RBCs) overlapping, which can cause poor classification.The Hough Circle transform is used to detect disk-like overlapped RBCs and then separate them[11].

V.DATA PREPARATION

Thousands of records are required to train the model effectively thus the size of the training dataset will be crucial because one of the characteristics of deep learning is its ability to perform well when trained with large data sets. The dataset was obtained from public sources like Google's Open Image, Kaggle. The records collected were prepared in a structured format.

After the data preprocessing, we randomly selected a large number of cell images the entire whole slide image dataset has been divided into four segments evenly. The cell image is considered as infected and include in dataset if and only if all the reviewers mark as positive otherwise it is excluded.The same phenomenon is applied to the non-infected cell also. After this data curation, we collected 5000 infected cells and 5000 non-infected cells. Further, this dataset is divided into two subsets those are Training set and Testing set[13][16].

Label	Training set	Test set
Infected cell images	4000	1000
Uninfected cell images	4000	1000

Table-1: Division of Images

Fig-4: Parasite Cell Images

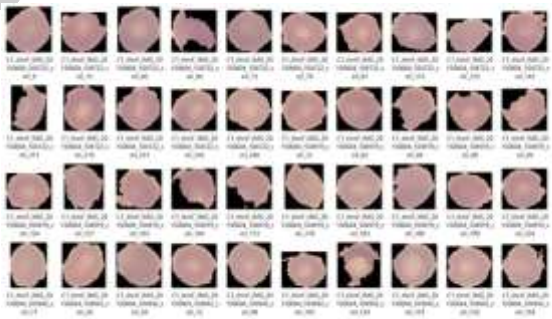

Fig-5: Non- Parasited Cell Images

Results:

The neural networks are trained using this training set and the trained models were then tested on the testing set. 75% of the images in the entire training set are collected to build a validation set. After training, validation was performed.In addition to validation, the trained neural networks were tested on a predefined test set to demonstrate their ability to handle unknown data. The accuracy is above 92%.

Fig-6: Infected cell prediction

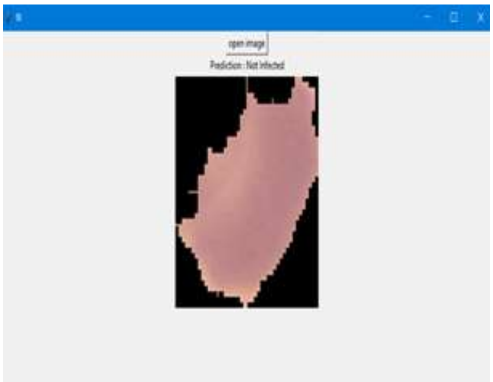

Fig-7: Non-Infected cell prediction

Discussions:

CNN is a very powerful algorithm that is widely used for image classification and object detection. The hierarchical structure and powerful feature extraction capabilities from an image make CNN a very robust algorithm for various image and object recognition tasks.

Future Scope:

By using this technology it is easy to give results as fast as possible and more accurately.

VI.CONCLUSION:

We have introduced a methodology and built an algorithm for detecting malaria, automated malaria detection, and evaluation of malaria contamination. Also, we developed a technique to coach with machine learning, adaptable to the detection of malaria with other kinds of parasites, and also discuss to extend the predictive value with results. visible of a scarcity of publicly available, high-resolution image datasets to support pattern recognition research for automated malaria diagnosis; we built a picture dataset of malaria-infected human red blood cells extracted from high-resolution whole slide images. We then used the dataset to train and test several well-known deep convolutional networks. Simulation results showed that very high recognition accuracy can be achieved by these deep learning techniques.

REFERENCES

[1].https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2688806/
[2].Ahirwar A., Pattnaik S., Acharya B., Advanced Image Analysis Based System for Automatic Detection and Classification of Malarial Parasite in Blood Images. International Journal of Information Technology and Knowledge Management
[3].Amit Kumar, Prof. Arun Choudhary, Prof. P. U. Tembhare, Prof. C. R.Pote. Enhanced Identification of Malarial Infected Objects using the Otsu Algorithm from Thin Smear Digital Images. International Journal of Latest Research in Science and Technology
[4].Y. LeCun, K. Kavukvuoglu, and C. Farabet, “Convolutional networks and applications in vision,” in Proc. International Symposium on Circuits and Systems (ISCAS’10). IEEE, 2010.
[5].https://arxiv.org/ftp/arxiv/papers/1506/1506.01195.pdf
[6].https://www.cdc.gov/malaria/about/faqs.html
[7].http://www.who.int/malaria/media/world-malaria-report-2016/en.
[8].http://who.int/mediacentre/factsheets/fs094/en/
[9].www.ece.uah.edu/_dwpan/malaria
[10].https://academic.oup.com/jn/article/147/11/2163S/4743207
[11].https://www.sciencedirect.com/science/article/pii/S193152441730333X
[12].https://cmr.asm.org/content/15/1/66
[13].https://www.kaggle.com/iarunava/cell-images-for-detecting-malaria
[14].https://www.sciencedirect.com/science/article/pii/S193152441730333X
[15].https://medium.com/@ksusorokina/image-classification-with-convolutional-neural-networks-496815db12a8
[16]. https://developers.google.com/machine-learning/practica/image-classification