Malaria Detection in Cell Images using Convolutional Neural Networks

Abstract

Malaria is one of the primary healthcare concerns that India faces today. Other than that it is also an important part of the wider socio-economic problem that plagues India. In 2017, there were a recorded 0.84 million cases of malaria which led to a recorded 194 deaths[1]. While over 100 countries have already eradicated malaria, a report published by the Lancet Commission(September 2019) puts India at 4th in the number of malaria cases worldwide. The problems were further compounded by the fact that India was one of the only countries that had recorded cases of urban malaria[2]. One of the most important parts of the malaria treatment procedure is swift and accurate diagnosis. The gold standard in this aspect remains microscopy of stained thick and thin blood smears, which not only sensitive also helps in distinguishing between various malaria species and their different stages. The other prevalent method is RDTs or Rapid Diagnostic Tests[3]. This research paper aims to come up with a neural network based model to detect malaria by processing stained blood smear images for application in areas where conventional diagnostic tools are not as readily available so as to facilitate its diagnosis and subsequent treatment.

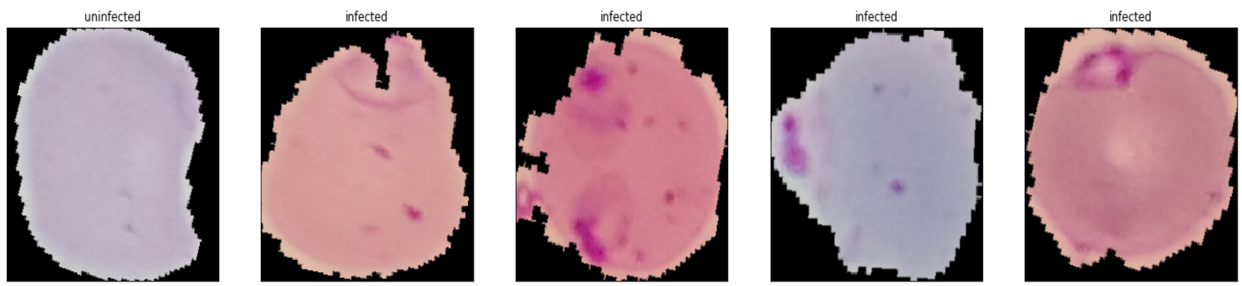
Introduction

Malaria diagnosis is done using multiple methods with clinical diagnosis, microscopic diagnosis and Rapid Diagnostic Tests being some of the more common ones. Clinical diagnosis are done on the basis of the patients signs and symptoms observed in the initial examination. Although conclusive, the results of the clinical diagnosis should always be followed by laboratory test before the administration of any medication. Microscopic diagnosis is a laboratory based test that involves the examination of a stained blood smear of the patients blood. This technique remains as the gold standard in malaria diagnosis. Another method is antigen detection using Rapid Diagnostic Tests. These tests are normally used in scenarios where reliable microscopic diagnosis is not available. Our objective with this research paper is to develop and test a new approach to diagnose malaria which aims to deliver consistent and accurate results both quicker and more cheaply than conventional approaches and at the same time takes away the requirement of trained medical professionals being on-site. The primary application area of this approach are remote areas where conventional diagnostic tools or medical professionals to use and interpret the diagnostic tools are in short supply. It's in these areas that our approach, if successful, could be used to substantial effect. This approach attempts to identify characteristics that differentiate infected and non infected cells. For this purpose we are using a dataset containing images of patients blood smears both infected and uninfected. The methodology involves building a convolutional based approach that identifies key characteristics and differences between the two subsets and later attempts to classify images based on these characteristics. In this paper we will first analyze the dataset and subsequently look at the various techniques and concepts used in our hypothesis with the results of our model and their analysis at the very end.

Dataset

The dataset is obtained from https://ceb.nlm.nih.gov/repositories/malaria-datasets/. The dataset is divided into two folders, namely: Infected and Uninfected. The dataset contains 27,558 images. The two categories, Infected and Uninfected contain exactly 13, 779 images. The dataset is evenly distributed among the two categories thus removing the change of bias towards any side. The images are in colour and are not of uniform size, hence before passing it into a Neural network, we reshape all images to a uniform size. The cell images are segmented cells from the thin blood smear slide images from the Malaria Screener research activity.The researchers at the Lister Hill National Center for Biomedical Communications (LHNCBC), part of National Library of Medicine (NLM), have developed a mobile application that runs on a standard Android smartphone attached to a conventional light microscope, which has been used to collect images for the dataset. .Giemsa-stained thin blood smear slides from 150 P. falciparum-infected and 50 healthy patients were collected and photographed at Chittagong Medical College Hospital, Bangladesh[4].

Sample images from dataset:



Data Preprocessing and Augmentation:

The images are of non-uniform sizes, hence we transform all of them to a standard size before passing them into the Neural Network. The images are represented as a tensor, a 3D matrix, where the depth will be 3 (Red Green Blue) and the values of the matrix will be in the range of 0 to 255 to represent intensity of a colour in that particular pixel. To pass the images into a Neural network, we have chosen to reshape all images into 224,224 size. Apart from these steps, we have also applied various Augmentation techniques like:

1. transforms.RandomHorizontalFlip() - Horizontally flip few images
2. transforms.RandomVerticalFlip() - Vertically flip few images
3. transforms.RandomRotation(20) - randomly rotate few images by 20°.

This data preprocessing and augmentation is done with the aim of creating more variety within our own dataset. CNNs achieved state-of-the-art results in a variety of classification tasks, but despite wide perspectives, they still have some challenges to deal with. They are mainly driven by the large size of the networks reaching millions of parameters as well as the lack of reliable training data sets, have a problem with overfitting as well as with generalization abilities[5]. Most popular and proven as effective current practice for data augmentation is to perform traditional affine and elastic transformations: creating new images by performing rotation or reflection of the original image, zooming in and out, shifting, applying distortion, changing the color palette[5].

Neural Networks

Neural networks are a set of algorithms loosely based upon the human brain. Such networks work well when used to cluster and classify large sets of real world datasets i.e they help group unlabeled data according to similarities in certain predefined characteristics among the inputs and classify it when it has a pre labeled dataset to work with. Neural networks typically form a part of a larger machine learning based application. They accept inputs through perceptrons that label the raw input and interprets patterns in this input which are then stored as vectors, which contain representations of all real world data types such as images, sounds, numerical values, text etc.

Methodology

In this paper we have attempted classifying images using both pre-trained models and our own CNN architecture. A convolutional neural network is a subdivision of the wider neural network application in that it works primarily on images. A CNN takes an input images and attempts to identify and assign weights(importance) to various aspects of the image in order to effectively differentiate one input from the other. Convolutional Networks (ConvNets) are currently the most efficient deep models for classifying images data. Their multistage architectures are inspired from the science of biology. Through these models, invariant features are learned hierarchically and automatically[6]. Convolutional Neural Networks are able to learn to extract features through different layers.

In input, images are presented as a matrix of pixels. It has 2 dimensions for a grayscale image. The color is represented by a third dimension of depth 3 to represent the fundamental colors (Red, Green, Blue)[7].

Approach with Pre-trained models

In this approach we use transfer learning to improve the performance of our model. The idea behind transfer learning is to use the knowledge from a related task and apply it to our task. This is done with the help of pre-trained models. The goal of transfer learning is to improve learning in the target task by leveraging knowledge from the source task. Transfer learning helps improve our model in three ways, firstly, the initial performance achieved by purely transferred knowledge and without any further learning, secondly, the time taken in training the model, and lastly, the final performance of our model[8]. In this paper we have used both VGG16 and Resnet152 as pre-trained models. In both models we have added our own fully connected layer in the end and trained its weights during the training process.

VGG16

VGG16 is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition"[reference]. The VGG16 model achieves an accuracy of 92.7% top 5 on testing on the ImageNet dataset. The VGG16 architecture takes the input image in a 224x224 RGB image. The convolutional layers use ReLU as a non-linear activation function. The advantage of using this model over others is its usage of small sized kernel filters. Spatial pooling is carried out by five max-pooling layers, which follow some of the conv. layers (not all the conv. layers are followed by max-pooling). Max-pooling is performed over a 2×2 pixel window, with stride 2[9].

Resnet152

Resnet152 is the second pre-trained model used in this paper. Resnet152 was also trained on the ImageNet dataset. The model consists of an architecture with upto 152 layers depth. The special feature of a residual network is its short connections applied to a plain network.[reference]. The Resnet152 model also takes 224x224 sized images as input.The convolutional layers mostly have 3×3 filters and
follow two simple design rules: (i) for the same output
feature map size, the layers have the same number of filters; and (ii) if the feature map size is halved, the number of filters is doubled so as to preserve the time complexity per layer[10].

Approach with Own CNN architecture

In addition to the pre trained model we also decided to build our own CNN based model for reference. We built this model using pytorch library. In the following sections we have explained the structure of the model.

A CNN model is traditionally made up of three parts, the first being the convolutional layer during which the kernels scan the given image and generate various feature maps. A model can have multiple convolutional layers with the feature maps of each layer being passed as input to the next convolutional layer. Each convolutional layer can be specified with certain parameters such as stride length, input features, output features and padding. The second type of layer is called the pooling layer which is responsible for reducing the size of the convolved feature so that the computational power required to process the data is reduced. The final layer is the classification or the fully connected layer. This layer learns non-linear combinations from the

high level features extracted from the convolutional layers. Over a series of epochs, this layer is able to choose the most relevant and dominant features and then classifies the data using softmax classification. One parameter that is helpful in isolating the most dominant features is dropout in which certain features are dropped at random to see their importance during the training phase and come up with the optimal set which gives the highest accuracy during the testing phase.
Our model has three convolutional layers. Each layer has been specified with stride length=2 and kernel size=5. Each layer has been specified with same padding such that the output has the same dimensions as that of the input. The input channels for the first layer is specified by the colour scheme of the input image. Since our image is in RGB, the number of input channels would be 3. The output channels on the other hand can be specified by the user, which we have set to 16. Similarly, the second layer has 16,32 and the third layer has 32,64 input and output channels respectively. The activation function used is ReLU(Rectified Linear Unit). After each convolutional layer, there is a max pooling layer except the third layer whose output is resized and fed into the first fully connected layer. The dropout layer is called after the first and the second fully connected layer with the final classification being done by the third fully connected layer.

Results

The CNN model gave a training accuracy of 95 after 20 epochs with the minimum validation loss of 0.145 and minimum training loss of 0.128. Another variation of the model was built in which we used Log Sigmoid instead of ReLU as the activation function. This model performed substantially worse with peak accuracy hitting just 66 with 0.638 and 0.615 as the minimum training and validation loss respectively after 20 epochs. The batch size was held constant at 100 for both the models.

The models using Transfer learning have shown slightly poorer results compared to the self created CNN architecture. The VGG16 model has shown a maximum Accuracy of 90% and the Resnet152 model has shown an accuracy of 93%, both of them after 20 epochs on the training set. The models have used a ReLU activation function with an Adam optimizer and CrossEntropyLoss function. The batch size has been 100 for both these models.

Model	Accuracy	Features
Customised CNN model	95%	3 layer CNN with ReLU activation function
Resnet152	93%	ReLU activation function and 3 fully connected layers added to pretrained Resnet152 Model
Customised CNN model	66%	LogSigmoid activation function instead of ReLU
VGG16	90%	ReLU activation function and 3 fully connected layers added to pretrained VGG16 Model

Inferences

The results from the models help us understand how neural networks and Image processing can be used in the medical sciences and can help detect diseases at a much on a cellular level. Medical imaging has great scope in the near future and the cost of diagnosis can be largely reduced by implementing such models and aid diagnosis of various diseases. We also infer than models using transfer learning have performed poorly since they have tried to apply knowledge of recognising daily objects on detecting diseases on a cellular level and this has not been as effective as creating a custom architecture. The knowledge from pre-trained models cannot be directly applicable to our problem. Pre-trained models however help reduce computational costs by a great margin because of their previous knowledge.

References

1) World Health Organization, “World Malaria Report 2017”,
https://www.who.int/malaria/publications/world-malaria-report-2017/en/,
National Vector Borne Disease Control Programme,
https://nvbdcp.gov.in/index1.php?lang=1&level=1&sublinkid=5784&lid=3689, accessed Oct 26, 2019.

2) Banjot Kaur, “India Fourth in number of global malaria cases : Lancet Report”, https://www.downtoearth.org.in/news/health/india-fourth-in-number-of-global-malaria-cases-lancet-report-66595, accessed Oct. 25, 2019.

3) National Vector Borne Disease Control Programme, National Institute of Malaria Research,“Guidelines for Diagnosis and Treatment of Malaria”, https://nvbdcp.gov.in/Doc/Guidelines_for_Diagnosis___Treatment.pdf, accessed Oct 25, 2019.

4)Sivaramakrishnan Rajaraman1, Sameer K. Antani1, Mahdieh Poostchi1, Kamolrat Silamut2, Md. A. Hossain3, Richard J. Maude2,4,5, Stefan Jaeger1, George R. Thoma1, “Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images”, PeerJ 6:e4568 https://doi.org/10.7717/peerj.4568

5)Agnieszka Mikołajczyk, Michał Grochowski, “Data augmentation for improving deep learning in image classification problem’’, IEEE, 2018.

6)A. Sifaoui, A. Abdelkrim, M. Benrejeb, "On RBF neural network classifier design for iris plants", The 37th International Conference on Computers and Industrial Engineering, pp. 113-118, October 2007.

7) R. Girshick, J. Donahue, T. Darrell, J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", CVPR, 2014.

8)Lisa Torrey and Jude Shavlik, “Transfer Learning”, Handbook of Research on Machine Learning Applications, 2009.

9)K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in International Conference on Learning Representations (ICLR), pp 1-2, 2015.

10) Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition ”, arXiv:1512.03385 [cs.CV], pp 3-4 10th December 2015.