PREDICTION OF MALARIAL INFECTION USING DEEP LEARNING

[1]Janasruthi S U, [2]Lokitha S, [3]Karthiga M

Department of Computer Science and Engineering, Bannari Amman institute of Technology
Erode, India

Abstract - One of the most lethal and deadly diseases in the entire planet is malaria. Because of the plasmodium parasite, the mosquito is the main source of the disease. If it is not recognized and treated at an early stage, it could get worse and result in death. We have established procedures for diagnosing malaria, which involve experts looking at the blood cells under a microscope. One of the most popular methods for diagnosing malaria by a skilled microscopist is the rapid diagnosis test, however it takes a long time to do and can produce inaccurate results if there is a flaw or human error.We require a detecting method to raise awareness of these diseases in order to solve this issue. A component of artificial intelligence is deep learning. It includes a wide variety of algorithms that can be used to forecast any given natural event. The medical field heavily relies on artificial intelligence. Here, the picture preprocessing and feature extraction were done using a convolutional neural network. Building a model using CNN will require a huge amount of data so we have used the Transfer Learning-VGG model along with the CNN to reduce time and for better accuracy. In this paper, we are going to detect whether the patient has malaria or not by using deep learning and transfer learning techniques, then integrating it with the web application with the help of python libraries.

Keywords– Convolutional Neural Network, Transfer Learning, Image preprocessing process, VGG 19, deep learning.

1. INTRODUCTION

Malaria is one of the most vulnerable diseases in the world. It should be detected at the early stage otherwise it will become more serious to give treatment. Malaria is mainly caused by the host mosquito which carries the Plasmodium parasite. Major causes of death in many developing countries are mainly due to malaria. It is considered to be one of the most fatal diseases when compared to other diseases. So we should consider this disease as critical and we have to do research to reduce the cause of this disease. The symptoms of malaria could be fever, shivering, vomiting and diarrhea. For diagnosis malaria, we have certain traditional methods like Rapid diagnostic test and Polymerase chain response which are clinical methods. But these traditional methods will have certain limitations. Both technique requires the well trainedmicroscopist for the detection and they consumes more time for the detection, they even leads to wrong result. It is easy to solve at the early stage but we need to identify as earlier as possible when it takes long timing it may not be solved much easier they may become more complex to solve and even leads to fatal. The world health Organization identified that, mainly the region of Africa is affected by the malaria in huge amount. After the analyses done by the world health organization in 2017, it is observed that globally 434,000 deaths and 218 million of people were diagnosed.

Light microscopy is a worldwide acceptable technique which is used to diagnosis the disease by the technicians. For diagnosis of malaria we have conventional light microscopy method which uses blood smears which are thick and thin. This is a classical technique but it has few limitations like we have to spend more time and we required technicians who are skilled. As per the study of the national wild in Ghana there was a population range of 100000, found that microscopes of 1.72 per the range of population, this was considered to be inaccurate in diagnosis result delay in time.

The proper treatment after the early identification of the disease, can cure the infected people and it can also save their lives. Still many peoples all over the world are not aware of the disease and its severity, and they don't have the proper diagnosis. Here we focused on developing the malaria detection model which gives high accuracy along with it we don't have the need for the skilled technicians and their medicinal equipment

for the detection purpose. Automated way of detecting the malaria will help to reduce the time and it is used to overcome the drawbacks of the clinical diagnosis method, and there is no dependency of the human being experts to detect the malaria cell or not. By using the knowledge of automated detecting process by the conventional method practice and get efficient results as the result of implementation.

1.1 Deep learning and Transfer Learning for Malarial detection

The field of Machine learning which is useful for making decisions for getting the desire output after analyzing the data and reading the data. In the healthcare field there goes a lot of research and development to make the healthcare sector achieve significant growth. Researchers have analyzed various machine model for the diagnosis and detection process in automated way. Feature engineering, analysis of factor and tuning will be required lot in machine learning. More data need to be provided and it is not scalable. Handling the feature engineering and training the feature which isnot a handy process in machine learning. So to maintain these we are going to the concept of deep learning process which is more reliable and scalable which results in the accuracy with higher rate. To perform tasks related to the computer vision convolutional neural network gives us the effective way. To transfer the volume of image input into the class label holding the volume of the output. Images are inputted into CNN and processed through the model's many layers for the purpose of analysis. CNN is the significant deep learning tool that helps us to train the multiple layers. It has good feature extraction properties which is utilized in the level of the recognizing the pattern in the system. Although transfer learning needs less data for training purposes than other concepts, it nevertheless requires a lot of time throughout the training process. This difficulty is therefore overcome. It is a laborious procedure, too.

The idea of transfer learning, which uses information from a previous task to inform the present model. The major goal of this kind of learning is to prevent the machine learning model from repeating itself a great deal from the beginning, and the knowledge from the pretrained model is applied to other models that carry out comparable tasks.Transfer learning and deep learning both use a minimum amount of data for model training, helping us to save time and resourcesand knowledge of thepretrained model is extract a   applies to other model where it also performs the similar tasks. Transfer learning along with the deep learning utilizes the data consumption limited for the model training which help us to save the resource and time.

2. LITERATURE REVIEW

It is possible for one individual to infect another with the blood sickness that is spread from one person to another by the parasite called plasmodium. Malaria, the worst disease in the world, is one of the leading causes of death [1].

Cerebral malaria or anemia, which carry blood to the brain through capillaries that may become blocked, can occur if diagnosis and treatment are postponed. This particular strain of malaria has the potential to be fatal [2]. A reduction in the death rate will result from the disease's early diagnosis saving time and resources through training. The microscopic examination is one of the most important A popular clinical diagnosis techniques is to strain both thick and thin blood on a piece of glass in order to analyze and see the parasite. However, this procedure will take a lot of time and the expertise of qualified professionals to determine whether malaria is present or not. There can be a shortage of skilled knowledge, and remote places might not have access to it. Therefore, it is difficult for those living in rural areas to receive a timely diagnosis of their illness. Studies have shown that laboratory misdiagnosis is most likely in endemic nations like India and other nations like South Africa [3]. Malaria misdiagnosis is primarily caused by in addition to the fifteen countries, sub-Saharan Africa and India account for 80% of the world's malaria cases [4]. The high malaria cases in those countries have given us the there is not enough amount of the skilled technician for billions of the active cases they result in the inaccurate analysis of the blood cells [5]. The detection process through automation will help us to effective diagnosis and easy all over the world.

Park et al. and Das et al., authors, have utilized strategies like as support vector, k-nearest neighbour, and bayesian learning. Machine learning and logistic regression for determining whether malaria exists. To prepare the dataset for preprocessing, they eliminated the stain from the blood smear image [6]. The most crucial preprocessing step used by Mustafa the researcher was thresholding to the notion for the task's analytical goals. For the experiment involving the malarial picture dataset, they employed the c-mean fuzzy algorithm and the wolf's approach, respectively. The instruments they have utilised for feature extraction.Through study of the size and orientation of the images for the training, they have optimised each and every individual dataset.[17]

The author Var ell for computer aided diagnosis proposed methods like feature extractor based on the convolutional neural network pretrained method. The pretrained model like imagenet on the large dataset, helps to form the image feature in generic for the target task by the process of transfer the knowledge learned. The model will have the weights for the better initialization and training the model with theinitialize weights

randomly to perform better way [18].

Numerous experiments have been conducted to show that among the most difficult jobs in machine learning and deep learning, detecting malaria using a thin blood smear image is one. In order to classify whether a person is infected or not, the researcher Dong et al. compared kernel-based algorithms like support vector machines with conventional neural networks. They divided the dataset for the random test and validation using the red blood cells as a guide. When compared to the kernel-based approach, they found that the CNN-based detection provides greater accuracy. When compared to the kernel-based approach, they found that the CNN-based detection provides greater accuracy. However, CNN has determined that a significant quantity of data and time would be needed for the model's training and that just a small number of qualified experts will be needed for the purpose of diagnosis[9].In another study the author performed the splitting of the smear image by randomized method and then evaluating the model with the deep belief network for the performance of the comparisonof classification algorithm for the prediction of the presence and absence of the malaria in the human begin[10]. Another author Rajaramanhave done the feature extraction from the annotated clinical image dataset in the large scale, the features are extraction by the pre trained layers and they validated it by the performance of both testing and training of the blood cell which is parasited or uninfected [19-24].

3. PROPOSED METHODOLOGY

The dataset we utilized in this analysis, which is available in an open source repository and contains smear blood cells, was created using deep learning and transfer learning. Here, the photos are being downsized to 224 X 224 pixels. Below, in Fig. 1, is an example image of a blood smear, together with a convolutional neural network and the pretrained model VGG19. For feature extraction and image processing, CNN is utilised. Transfer learning is employed to shorten training times and shrink the quantity of thedatasetfor the training model. The proposed methodology's flow is provided below in Fig. 2.

3.1 DATASET

This dataset was obtained from the National Institutes of Health's official website. The collection consists of 30,600 photos, of which 15,300 are free of infection and the remaining 15,300 are. The collection comprises of tiny smear blood cell pictures that are used to identify malaria. The dataset was gathered as part of the biomedical communication project from the National Library of Medicine.

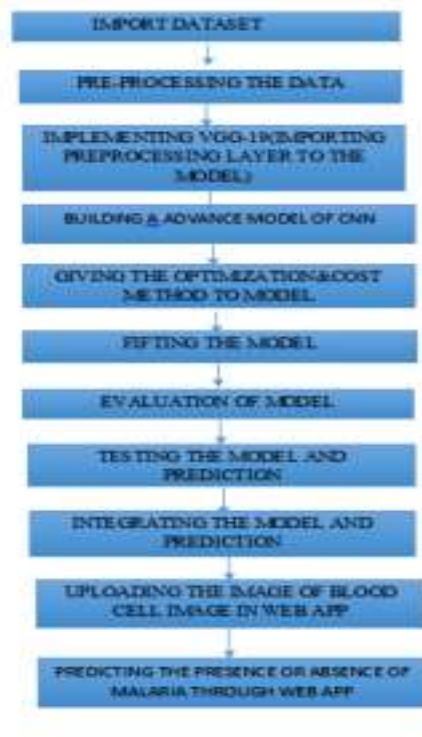**Fig(1):Uninfected and infected Sample blood smear**

Figure 2. Flow chart of the proposed solution

3.3 PREPROCESSING

To perform the computer vision related to the image as the first step we have to preprocess the image which we are going to use. This dataset consists of thousands of images for both training and testing purposes. The dataset consists of both parasited and non parasited images which are labeled as infected and uninfected cells under the respective folders. The dataset is divided into validation, testing and training. The validation to train to test ratio is 63:7:30. We have made analysis that 63% of dataset have been used for the training to maintain the level of the dataset for the computational purpose. Then 63% of the dataset is for the training purpose which helps us to detect the images of the blood cell is whether healthy or unhealthy blood cells, and then the dataset of 30% is used for the testing of the model. After the dataset analysis the images are in different different dimensions throughout the dataset. Image scaling is important step in analyzing the performance of the data that will reduce the memory space and helps us to reduce the time in computation. We know that the large amount of data will require a huge amount of the memory space to store the image as the data format but not only that it also leads to the complex building neural network model. In Order to reduce both space and time complexity we are going for the certain preprocessing techniques like label encoding, normalization, data augmentation. It is obsolete that this process will help us to preprocess the dataset according to our need for the model development purpose in the effective and faster manner.

3.4 CONVOLUTIONAL NEURAL NETWORK-VGG19

With regard to image processing and visualisation, artificial intelligence is crucial. Convolutional neural networks are mostly utilised in deep learning for feature extraction and visualisation. It is an artificial neural network of some sort. CNN is mostly used for object detection, image classification, face and image recognition, and object recognition. The input layer, output layer, and hidden layer are the three types of layers that make up the CNN architecture. Because VGG-19 is regarded as the greatest visual computer model and is most efficient in their process, CNN uses it for transfer learning. One of the deep CNN models, VGG 19, is thought to be particularly useful for the classification of images. We are able to solve the CNN-era challenges with the aid of the VGG-19 design. TheVGG-19 Architecture is represented below in form of, fig (3)
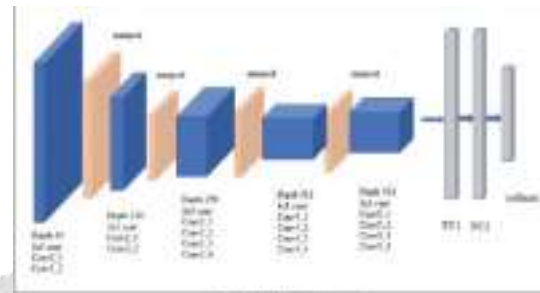
Figure 3. VGG-19 Network Architecture.

In VGG-19 architecture it has 19 deep layers that helps us to classify the images or objects under certain categories. As the result of this pretrained model we can achieve the accuracy with high effective manners along with the decrease in the amount of the time it takes to train the model. Python was employed in this case for coding purposes. One of the predefined models for transfer learning approaches is the VGG-19. Transfer learning which is nothing but it has the feature used to transfer the knowledge learned by the previously learned model and it is imported as the input for the upcoming or further developing model to reduce the time consuming to detect. This method helps us to build the model from the previous perspective despite of building the model from the scratch. Here we recognized that CNN trained dataset in order to extract the feature for the computer vision. The VGG-19 has 16 convo layers which are pre trained available in it. Each layer in the convolutional network will use batch normalization technique and it also uses mini batch per variable which are helpful in the activation of node from the layer previous to next node, then we have to optimize the model by using optimization techniques.

In convolutional neural networks we have a deep model which is integrated by transfer learning. It has 19 substantial layers. One of the varieties of VGG models is the VGG-19.There also many other types like VGG-11,VGG16. The latest version of the VGG model type is VGG-19 it is one of the trending and advanced level models with pretrained layers. VGG named from Visual Geometry Group of the oxford so that it is named as VGG. It mainly uses the convolutional layer deep to improve accuracy and result. This model consists of the 3 fully connected layer with max of five max pooling layer and 16 convolutional layers. Here we have used adam as the optimizer and relu in the activation function for the neural network, after building the model we have to integrate it with the application in order for the detection of malaria as the easiest process.

3.5 INTEGRATION OF MODEL WITH WEB APPLICATION

Here, we've used Python Flask to integrate our model with the web application. The Python framework for creating lightweight apps is called Flask. After the model has been created, it must be integrated with the web application in order to determine whether or not the person has malaria. In the front end we will have the upload option and predict button it will help us to upload and detect. After the input passed over the front-end part it will directly move on to the pretrained model for the analysis purpose. After the integration is completed we will have the upload option in the web application for uploading the blood smear images as the input for the detection purpose whether the person is suffering from malaria or not. As we have used the transfer learning the result would be good and greater accuracy, not only that it will also take very less time for the analysis of the disease. The necessary things required for this proposed work is given below in the form of Table. 1.

Table 1: List of environments for development

REQUIREMENT	ENVIRONMENT/ RESOURCE
DATASET	Kaggle
LANGUAGE	Python

ALGORITHM	CNN+VGG-19
METHOD	DeepLearning,Transfer Learning
FRONTEND LANGUAGE	HTML,CSS, JAVASCRIPT
MIDDLEWARE	FLASK
CODE	VISUAL STUDIO CODE

4. RESULT AND DISCUSSION

Over 50 epochs were used to train this model.The dataset which has thousands of images used for the training and testing purpose.After this we have saved our model. Trained model is tested with the images of infected and uninfected smear images as the input and we got the output in the correct manner.The result of the given input image was good due the proposed work.We have also tested with the other blood smear images which was collected from the laboratory as the result of this is also results with great.The VGG-19 model achieves the accuracy of 95% so that the detection of the disease would become easier. This research shows that the deep learning along with the transfer learning is the good and great technique inorder to get the result very promising.CNN alone will require huge amount of the data and the time for the predicting when compared with the VGG-19 because it has the pretrained model and the result would be very faster than CNN. The VGG19 model's Structure is given in below fig (4),

Fig(4):The structure of VGG-19 model

After building the model while training and Testing the dataset we have obtained the accuracy which are given below in Fig.5 along with Table. 2.

Table 2: Training and Testing accuracy

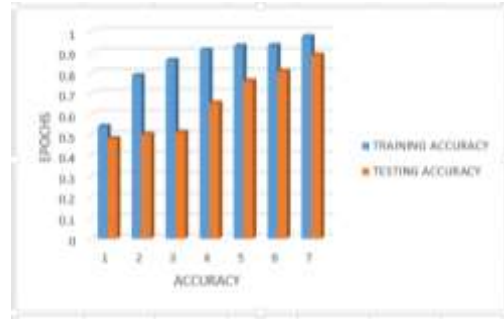| TRAINING ACCURACY	TESTING ACCURACY
0.5434	0.4848
0.7883	0.5041
0.8622	0.5152
0.9107	0.6555
0.9311	0.7629
0.9337	0.8094
0.9764	0.8877

Figure 5. Accuracy Graph

While training the model we will get some loss values which helps us in indicating the state of target and its difference. The loss of testing and training is given below in form of Fig.6 and Table. 3.

Table 3: Training and Testing Loss

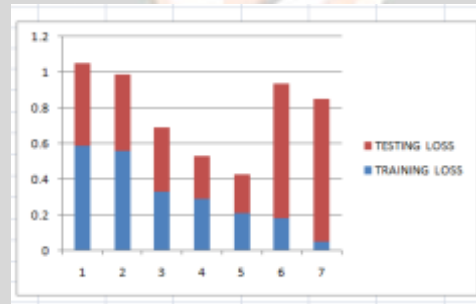| TRAINING LOSS	TESTING LOSS
0.59	0.46
0.56	0.43
0.33	0.36
0.29	0.24
0.21	0.22
0.18	0.76
0.05	0.8

Figure 6. Loss Graph

To evaluate the malaria dataset, we must experiment with a variety of post- and preprocessing methods. The preprocessing method of normalisation and standardisation does not aid in enhancing the model's functionality. Data augmentation helps us to perform the better on the malarial dataset. The performance of the model is increased by using the method of normalization. To see the better version just we have to move post processing process for good accuracy. Since there is no variation in the validation set of data, the VGG-19 model performed well with the vast amount of data. As discussed earlier that we have developed the web application for the detection of malaria by using python flask, it is a framework. We have used visual studio code to run our application. The dataset has two part uninfected and infected cells, the images are shown below Fig.7
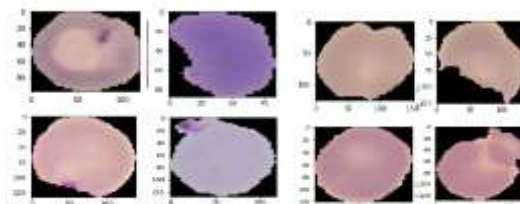

Figure 7. Infected and Uninfected blood cells

After running the flask code we will get the Local host url link to open web application which is running in the web browser, the page of web application which will be opened is displayed below Fig.8.

Figure 8. Frontpage of Web Application

After the web application opens, we will have a button to upload the blood smear image, after uploading we will have another button which is used to predict the output. The user interface is displayed below in Fig.9.

Figure 9. The user inference for prediction

  After the clicking the predict it will show us whether the given blood cell image is infected with malaria or not. The output user interface after the prediction will be displayed as below Fig.10.

Figure 10. The output user interface after prediction

 As the result the web application will give us the output faster and generate high quality output in a less amount of time. This Web application helps us to detect the malaria without any time consuming to analysis.

5. CONCLUSION

   Here we have developed the end-to-end Deep learning neural network to perform malarial diagnosis and detection. We have used preprocessing techniques such as normalization and data augmentation. Our efforts to contribute by raising performance do not help. Transfer learning is a technique which helps us for the great analysis and it provides a lot in scaling features and engineering. Deep learning VGG-19 helpful in image processing as much as it can in the processing. Problems in real world sparsity of data issue and also requires the need of the training hardware, which helps us to automate the process of the detection using VGG-19 model along with the CNN network as per they reduce the epochs needed to run and results in the high accuracy.The proposed model helps us to get the acceptable results with the great accuracy.Since we have

developed the user interface for the upload images and to get output we have connected the model with flask as middleware. Whenever the we are in need to detect the presence and absence of malaria in a person with the limited period of time our app will be very helpful. As the future work, we plan to focus on the architecture of the model for the better and simultaneous response for increasing the performance level of the model. In order to produce a model with improved performance, we have intended to determine the best technique to manipulate the network architecture and feature extraction.

REFERENCES

1. Trampuz, A., Jereb, M., Muzlovic, I., & Prabhu, R. M. (2003). Clinical review: Severe malaria. Critical care, 7(4), 1-9.
2. Tangpukdee, N., Duangdee, C., Wilairatana, P., & Krudsood, S. (2009). Malaria diagnosis: a brief review. The Korean journal of parasitology, 47(2), 93.
3. Hänscheid, T. (2003). Current strategies to avoid misdiagnosis of malaria. Clinical Microbiology and Infection, 9(6), 497-504.
4. Quinn, J. A., Nakasi, R., Mugagga, P. K., Byanyima, P., Lubega, W., & Andama, A. (2016, December). Deep convolutional neural networks for microscopy-based point of care diagnostics. In Machine learning for healthcare conference (pp. 271-281). PMLR.
5. Mustafa, W. A., Santiagoo, R., Jamaluddin, I., Othman, N. S., Khairunizam, W., & Rohani, M. N. K. H. (2018, August). Comparison of detection method on malaria cell images. In 2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA) (pp. 1-6). IEEE.
6. Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. Computers & geosciences, 10(2-3), 191-203.
7. Liu, F., & Picard, R. W. (1996). Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. IEEE transactions on pattern analysis and machine intelligence, 18(7), 722-733.
8. Kapoor, R. (2020). Malaria Detection Using Deep Convolution Neural Network (Doctoral dissertation, University of Cincinnati).
9. Bartolo, A., Camilleri, K. P., Farrugia, P. J., & Borg, J. C. (2004). Adaptation of brensen's thresholding algorithm for sketched line drawings.
10. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.
11. Var, E., & Tek, F. B. (2018, September). Malaria parasite detection with deep transfer learning. In 2018 3rd international conference on computer science and engineering (UBMK) (pp. 298-302). IEEE.
12. Rajaraman, S., Antani, S. K., Poostchi, M., Silamut, K., Hossain, M. A., Maude, R. J., ... & Thoma, G. R. (2018). Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. PeerJ, 6, e4568.
13. Dong, Y., Jiang, Z., Shen, H., Pan, W. D., Williams, L. A., Reddy, V. V., ... & Bryan, A. W. (2017, February). Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. In 2017 IEEE EMBS international conference on biomedical & health informatics (BHI) (pp. 101-104). IEEE.
14. Liang, Z., Powell, A., Ersoy, I., Poostchi, M., Silamut, K., Palaniappan, K., ... & Thoma, G. (2016, December). CNN-based image analysis for malaria diagnosis. In 2016 IEEE international conference on bioinformatics and biomedicine (BIBM) (pp. 493-496). IEEE.
15. Bibin, D., Nair, M. S., & Punitha, P. (2017). Malaria parasite detection from peripheral blood smear images using deep belief networks. IEEE Access, 5, 9099-9108.
16. Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., ... & Campilho, A. (2017). Classification of breast cancer histology images using convolutional neural networks. PloS one, 12(6), e0177544.
17. Sountharrajan, S., Karthiga, M., Suganya, E., & Rajan, C. (2017). Automatic classification on bio medical prognosisof invasive breast cancer. Asian Pacific Journal of Cancer Prevention: APJCP, 18(9), 2541.
18. Sountharrajan, S., Nivashini, M., Shandilya, S. K., Suganya, E., Bazila Banu, A., & Karthiga, M. (2020). Dynamic recognition of phishing URLs using deep learning techniques. In Advances in cyber security analytics and decision systems (pp. 27-56). Springer, Cham.
19. Karthiga, M., Sountharrajan, S., Nandhini, S. S., & Sathis Kumar, B. (2020, May). Machine Learning Based Diagnosis of Alzheimer's Disease. In International Conference on Image Processing and Capsule Networks (pp. 607-619). Springer, Cham.
20. Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018, April). Synthetic data augmentation using GAN for improved liver lesion classification. In 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018) (pp. 289-293). IEEE.

21. Haque, U., Ahmed, S. M., Hossain, S., Huda, M., Hossain, A., Alam, M. S., ... & Haque, R. (2009). Malaria prevalence in endemic districts of Bangladesh. PloS one, 4(8), e6737.
22. Khan, A. M., Rajpoot, N., Treanor, D., & Magee, D. (2014). A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. IEEE transactions on Biomedical Engineering, 61(6), 1729-1738.
23. Anghel, A., Stanisavljevic, M., Andani, S., Papandreou, N., Rüschoff, J. H., Wild, P., ... & Pozidis, H. (2019). A high-performance system for robust stain normalization of whole-slide images in histopathology. Frontiers in medicine, 6, 193.
24. Tharsanee, R. M., Soundariya, R. S., Kumar, A. S., Karthiga, M., & Sountharrajan, S. (2021). Deep convolutional neural network–based image classification for COVID-19 diagnosis. In Data Science for COVID-19 (pp. 117-145). Academic Press.