10.48047/jocaaa.2024.32.02.36

MALARIA DIAGNOSIS USING CNN FEATURE EXTRACTOR, PARASITE INFLATOR AND DOUBLE HIDDEN LAYER EXTREME LEARNING MACHINE ALGORITHM

Mrs. D.SOWMYA[1], RACHAMSHETTY VAISHNAVI[2], SATHWIKA YADAGIRI[3], SAMALA PRIYANKA[4]

[1]Assistant Professor, Department of Computer Science and Engineering, Sridevi Women's Engineering College, Hyderabad, India

Email: sowmyadurga167@gmail.com

[2,3,4,]B.Tech Student, Department of Computer Science and Engineering, Sridevi Women's Engineering College, Hyderabad, India

Abstract: Malaria, a perilous and fatal disease transmitted through the bite of female Anopheles mosquitoes carrying Plasmodium parasites, poses significant threats in high-risk regions like South-East Asia, the Eastern Mediterranean, the Western Pacific, and the Americas. With approximately 400 types of Anopheles mosquitoes, 30 primarily serve as carriers, necessitating their bite on a malaria-infected person to become vectors. The mosquito's life cycle, encompassing egg, larva, pupa, and adult stages, relies on feeding on human blood to advance from eggs to adults, facilitating the spread of malaria. Prompt diagnosis is crucial as malaria can swiftly progress to a severe, life-threatening stage. While not directly transmissible between individuals, it can be transmitted from mother to fetus, through blood transfusions, or shared injections. The disease thrives in hot, humid climates near natural water sources where Anopheles mosquitoes proliferate, posing a heightened risk. Addressing the formidable challenge of eradicating malaria involves diverse strategies, encompassing medical interventions and socio-economic analyses. The World Health Organization's 'Malaria Microscopy Quality Assurance Manual—Version 2' plays a pivotal role in ensuring accurate diagnoses, emphasizing the significance of microscopy in disease detection. This abstract underscores the multifaceted nature of combatting malaria and the critical role played by precise diagnostic methodologies outlined in global health guidelines.

Index Terms - *Malaria, Anopheles mosquitoes, Plasmodium parasites, Mosquito life cycle, Diagnosis, Vector-borne disease, World Health Organization, Microscopy, Eradication strategies, Global health guidelines.*

Keywords: Malaria diagnosis system, CNN feature extractor, Machine Learning, Double hidden layer , Learning machine algorithm, Data Mining, Data Augmentation

1. INTRODUCTION

South-East Asia, the Eastern Mediterranean, the Western Pacific, and the Americas have all been recognized as high-risk regions by the WHO. Malaria is a dangerous and deadly disease that is triggered from the bite of female anopheles mosquitoes that host plasmodium parasites. There are around 400 types of anopheles; among them, 30 types mainly act as parasite carriers. To be a parasite carrier, the female anopheles mosquito has to bite a malaria-infected person. The life cycle of the anopheles mosquito consists of four stages: egg, larva, pupa,

10.48047/jocaaa.2024.32.02.36

and adult. To bring the eggs into the adult stage, they have to feed on human blood. In this feeding process, malaria spreads all around. It takes 10 to 15 days or even more to remain in hibernation after a carrier mosquito bite. Malaria should be diagnosed as soon as possible because it can rapidly turn into a severe stage and is life-threatening. It cannot be passed from one person to another, however malaria can be transmitted from mother to fetus, contracted through blood transfusions or sharing injections. This disease can be spread in hot, humid climates near natural water sources where Anopheles mosquitoes transmit deadly diseases.

In the pursuit of eradicating malaria, a formidable global health challenge, diverse approaches have been employed, ranging from medical interventions to socio-economic analyses. The World Health Organization's 'Malaria Microscopy Quality Assurance Manual—Version 2' [1] serves as a foundational guide in ensuring the accuracy of malaria diagnoses, emphasizing the crucial role of microscopy in disease detection.

Complementing this, Taylor et al.'s work on the 'Respiratory Manifestations of Malaria' [2] sheds light on the broader clinical aspects, emphasizing the multi-faceted nature of the disease. Moving beyond the clinical realm, Barat et al. [3] scrutinize the equity dimensions of malaria control interventions, exploring whether these measures effectively reach impoverished populations. Gharakhanlou et al.'s agent-based model [4] delves into the dynamic spread of Plasmodium vivax malaria, providing insights crucial for devising effective prevention strategies. The socio-economic ramifications of malaria find resonance in Singh et al.'s research [5], unveiling the economic burden borne by households in high and low transmission areas of central India. On a different front, foundational works in database systems [6], software engineering principles [7], and software testing techniques [8] and tools [8] highlight the pivotal role of information technology in managing and combating malaria.

Watts S. Humphrey's 'An Introduction to the Team Software Process' [9] and Krishnan and Gehrke's 'Database Management Systems' [10] contribute to the broader understanding of systematic processes that underpin comprehensive malaria control efforts. This eclectic collection of literature underscores the interdisciplinary nature of the battle against malaria, weaving together insights from medical, socio-economic, and technological domains. In this intricate tapestry, each strand of knowledge plays a vital role in informing strategies for a more effective and equitable malaria control paradigm.

2. LITERATURE REVIEW

The literature survey encompasses a diverse range of topics related to malaria, database systems, software engineering, and testing techniques. The references provide insights into malaria diagnosis and treatment, the economic impact of malaria, and various aspects of database systems and software engineering. This survey aims to synthesize key findings from the cited works and highlight their contributions to the respective fields. The World Health Organization's Malaria Microscopy Quality Assurance Manual [1] lays the foundation for ensuring the accuracy and reliability of malaria diagnosis through microscopy. This manual is pivotal in guiding healthcare

10.48047/jocaaa.2024.32.02.36

professionals and laboratories to maintain high standards in malaria diagnosis, crucial for effective disease management.

Taylor et al.'s study on the respiratory manifestations of malaria [2] sheds light on the diverse clinical presentations of the disease. Understanding these manifestations is crucial for accurate and timely diagnosis, ensuring appropriate treatment and reducing mortality rates associated with severe cases. Barat et al. [3] present an equity lens on malaria control interventions, emphasizing the importance of reaching vulnerable populations. The study underscores the need for targeted efforts to ensure that malaria control strategies effectively reach the poor, contributing to more inclusive public health policies.

Gharakhanlou et al.'s agent-based model for simulating the dynamic spread of Plasmodium vivax malaria [4] provides a computational approach to understanding the transmission dynamics of the disease. Such models are valuable tools for predicting and planning interventions to control malaria spread, especially in regions where multiple malaria species coexist. Singh et al.'s research on the economic cost of malaria at the household level in central India [5] quantifies the financial burden of the disease. This information is crucial for policymakers and healthcare providers to assess the economic impact of malaria and design targeted interventions to alleviate the financial strain on affected households. Shifting focus to the realm of information technology, "Database Systems design, Implementation, and Management" [6] by Rob and Coronel provides a comprehensive guide to designing and managing database systems. The book serves as a foundational resource for individuals involved in creating robust and efficient database solutions.

Jawadekar's "Software Engineering principles and practice" [7] and Beizer's "Software Testing techniques" [8] contribute to the software engineering domain. These texts offer valuable insights into the principles, methodologies, and techniques employed in software development and testing, ensuring the delivery of high-quality software products.

Prasad's "Software Testing Tools" [8] further extends the literature on software testing by delving into specific tools that aid in the testing process. Understanding and employing these tools are essential for improving the efficiency and effectiveness of software testing practices. Finally, Humphrey's "An Introduction to the Team Software Process" [9] and Krishnan and Gehrke's "Database Management Systems" [10] provide additional perspectives on software development processes and advanced concepts in database management, respectively.

In conclusion, the literature survey presents a multidimensional view, covering critical aspects of malaria diagnosis, treatment, and economic impact, alongside insights into database systems and software engineering. The integration of these diverse topics offers a comprehensive understanding of the challenges and advancements in these fields, providing a solid foundation for future research and practical applications.

3. METHODOLOGY

10.48047/jocaaa.2024.32.02.36

In literature they proposed an ML technique based on a CNN to automatically categorize single cells in thin films. Based on 27,578 singles, ten-fold cross-validation was performed using a new 16-layer CNN model on cell images. Another researcher utilized the global threshold approach on a green channel color image to extract RBCs from the background. Then, a morphological filter and related component labeling were used, noise and holes in the RBCs were removed. Following that, the geometrical features of the RBCs are used to extract information from them. They utilized an Artificial Neural Network (ANN) to detect malaria from this derived information.

Drawbacks:

1. In one of the existing work uses global threshold for extraction of geometrical features from the RBCs, which may not capture all the relevant information for accurate malaria detection.

2. The power of CNNs to automatically extract more comprehensive and representative features from the images, potentially leading to improved classification performance.

3. It does not mention any data preprocessing techniques, which may result in noise or uncertainties in the training data and potentially affect the performance of the malaria detection system.

4. The existing study relies on traditional image processing techniques combined with a basic ANN which leads to decrease the performance of model.

The proposed work in the paper is an unconventional approach for the automated diagnosis of malaria using an extreme learning machine (ELM) algorithm with double hidden layers (DELM) and Convolutional Neural Networks (CNN) as classifiers. The CNN model is used as a feature extractor and classifier, and the derived features are used to train ELM and DELM. Two versions of the malaria image dataset have been used: one is the original dataset, and the other is a modified dataset where ambiguous samples have been removed. The parasite inflator acts as the shape increaser of the small, darker malaria parasites in the RBC images in order to detect malaria easily.

Benefits:

1. Small darker parasite spot has been increased in shape using proposed ‘‘Parasite Inflator’’ to detect malaria easily.

2. The original and the modified malaria dataset, has been used in this study.

3. The power of CNNs to automatically extract more comprehensive and representative features from the images, potentially leading to improved classification performance.

4. It utilizes more advanced techniques such as CNNs and DELM as a classifiers, which have demonstrated exceptional performance in various image-related tasks

10.48047/jocaaa.2024.32.02.36

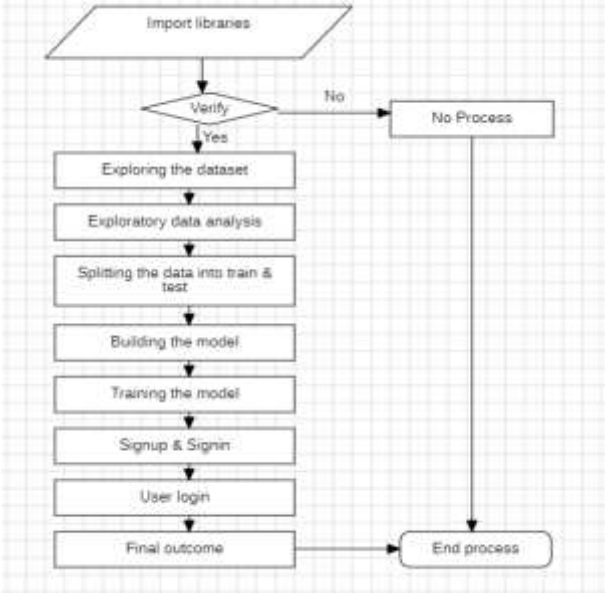

Fig 1 System Architecture

Modules:

The modules are:

- Dataexploration:usingthismodulewewillload data intosystem
- Processing:Usingthe modulewewillreaddatafor processing
- Splitting data into train &test: using this module data will be divided into train & test
- Model generation: Building the model – CNN - Feature Extraction using CNN - ELM - one layer - ELM – Two layer - Voting Classifier (RF + DT) -ResNet18 - XceptionAlgorithms accuracy calculated
- Usersignup&login:Usingthismodulewillgetr egistrationandlogin
- Userinput:Usingthismodulewillgive inputfor prediction
- Prediction:Finalpredicteddisplayed

4. IMPLEMENTATION

AlexNet -AlexNet is a convolutional neural network that is 8 layers deep. You can load a pretrained version of the network trained on more than a million images from the ImageNet database [1]. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals.

VGG16 - VGG16 is a convolutional neural network model that's used for image recognition. It's unique in that it has only 16 layers that have weights, as opposed to relying on a large number of hyper-parameters. It's considered one of the best vision model architectures.

Xception - Xception is a convolutional neural network that is 71 layers deep. You can load a pretrained version of the network trained on more than a million images from the ImageNet database [1]. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals

ResNet50 -ResNet-50 is a 50-layer convolutional neural network (48 convolutional layers, one MaxPool layer, and one average pool layer). Residual neural networks are a type of artificial neural network (ANN) that forms networks by stacking residual blocks.

10.48047/jocaaa.2024.32.02.36

DenseNet -A DenseNet is a type of convolutional neural network that utilises dense connections between layers, through Dense Blocks, where we connect all layers (with matching feature-map sizes) directly with each other.

CNN - A CNN is a kind of network architecture for deep learning algorithms and is specifically used for image recognition and tasks that involve the processing of pixel data. There are other types of neural networks in deep learning, but for identifying and recognizing objects, CNNs are the network architecture of choice.

ELM - one layer -ELM - Two Layer -Extreme learning machines (ELMs) have proven to be efficient and effective learning mechanisms for pattern classification and regression. However, ELMs are primarily applied to supervised learning problems. Only a few existing research papers have used ELMs to explore unlabeled data.

LR -Logistic regression is a supervised machine learning algorithm that accomplishes binary classification tasks by predicting the probability of an outcome, event, or observation. The model delivers a binary or dichotomous outcome limited to two possible outcomes: yes/no, 0/1, or true/false.

DT -Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, the decision tree algorithm can be used for solving regression and classification problems too.

RF -Random forest is a commonly-used machine learning algorithm trademarked by Leo Breiman and Adele Cutler, which combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fueled its adoption, as it handles both classification and regression problems.

SVM -SVM is a powerful supervised algorithm that works best on smaller datasets but on complex ones. Support Vector Machine, abbreviated as SVM can be used for both regression and classification tasks, but generally, they work best in classification problems.

KNN -K-Nearest Neighbours Algorithm. The k-nearest neighbours algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point.

Voting Classifier - A Voting Classifier is a machine learning model that trains on an ensemble of numerous models and predicts an output (class) based on their highest probability of chosen class as the output.

5. EXPERIMENTAL RESULTS

Dataset Description:

Dataset URL:

https://www.kaggle.com/datasets/iarunava/cell-images-for-detecting-malaria

The dataset we're going to use can be downloaded from Kaggle. It contains data about 27,558 images of RBC. The dataset contains equal amount of malaria infected and uninfected samples hence it is called the balanced dataset.

10.48047/jocaaa.2024.32.02.36

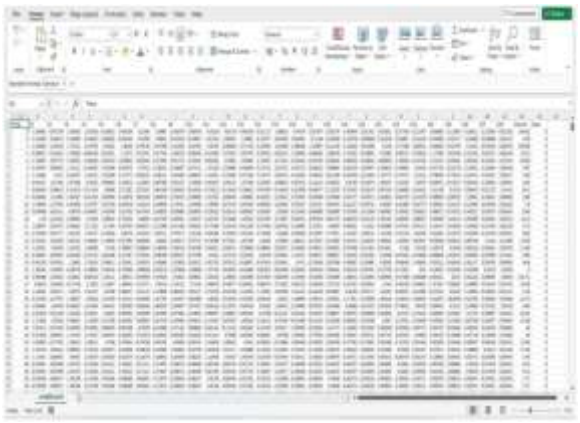

Fig 2 Dataset.csv file

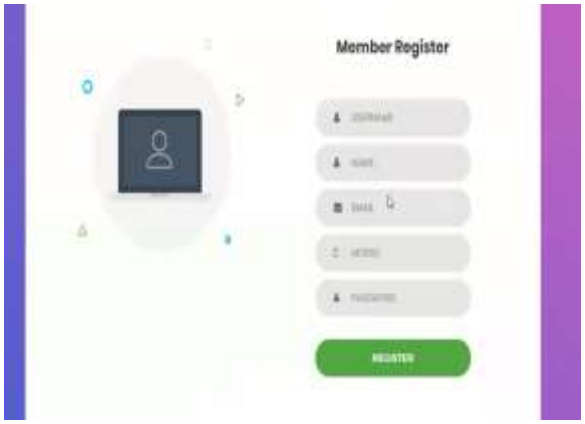

Fig 6Member Registration Window

Fig 3 Anaconda prompt

Fig 4 URL

Fig 7Member login

Fig 5Website dashboard

Fig 8After signup

10.48047/jocaaa.2024.32.02.36

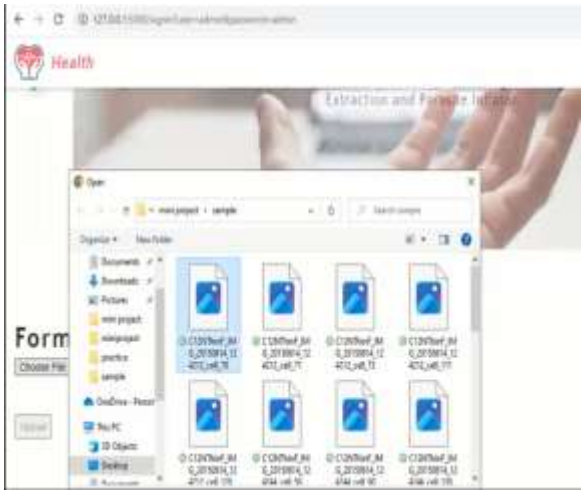

Fig 9Uploading image from the dataset

Fig 10The Output prediction

6. CONCLUSION

The proposed system CNN-DELM has been presented for the automated diagnosis of malaria from tiny blood samples. The result shows that the detection accuracy improves when falsely labeled images are removed. Due to several morphological techniques, the malaria parasites are highlighted. A lightweight CNN has been trained on preprocessed cell images for extracting the most informative features. Eventually, DELM successfully differentiate between infected and uninfected samples from these features. The proposed parasitic inflator, CNN's capability to extract features, and DELM's ability to generalize make the proposed approach more effective. The efficacy of the proposed framework is clear from its magnificent outcomes; whereas, the CNN-DELM has obtained high accuracy and recall for the original dataset. Also, it has acquired a promising score of accuracy and recall for the updated dataset. Since the proposed framework outperformed the state-of-art models, the result of this study is expected to help medical professionals for detecting malaria and diagnose the malaria-affected individuals more quickly and efficiently. We have a plan to work on the multiple hidden layer ELM algorithms with CNN feature extractor and parasite inflator for malaria detection.

FUTURE ENHANCEMENTS

This project mainly focuses on the detection or the diagnosis of the deadly disease malaria. And this project actually uses the elements like ELM model and DELM for classification purpose and the parasite inflator for the image spotting etc. In future this project may also use EDPT is the main strategy of malaria control - radical treatment is necessary for all the cases of malaria to prevent transmission of malaria. And give a detailed description of the disease with the doctor's prescription.

REFERENCES

[1] Malaria Microscopy Quality Assurance Manual—Version 2, World Health Org., Geneva, Switzerland, 2016.

10.48047/jocaaa.2024.32.02.36

[2] W. R. J. Taylor, J. Hanson, G. D. H. Turner, N. J. White, and A. M. Dondorp, ‘‘Respiratory manifestations of malaria,’’ Chest, vol. 142, no. 2, pp. 492–505, Aug. 2012.

[3] L. M. Barat, A. Mills, S. Basu, N. Palmer, K. Hanson, and E. Worrall, ‘‘Do malaria control interventions reach the poor? A view through the equity lens,’’ Amer. J. Tropical Med. Hygiene, vol. 71, no. 2, pp. 174–178, Aug. 2004.

[4] N. M. Gharakhanlou, M. S. Mesgari, and N. Hooshangi, ‘‘Developing an agent-based model for simulating the dynamic spread of plasmodium vivax malaria: A case study of,’’ Ecol. Informat., vol. 54, Nov. 2019, Art. no. 101006.

[5] M. P. Singh, K. B. Saha, S. K. Chand, and L. L. Sabin, ‘‘The economic cost of malaria at the household level in high and low transmission areas of central India,’’ Acta Tropica, vol. 190, pp. 344–349, Feb. 2019.

[6] Database Systems design, Implementation, and Management, Peter Rob & Carlos Coronel 7th

Edition.

[7] Software Engineering principles and practice- Waman S Jawadekar, The Mc Graw-Hill

Companies.

[8] Software Testing techniques - Baris Beizer, Dreamtech, second edition. Software Testing Tools – Dr. K. V. K. K. Prasad, Dreamtech.

[9] An Introduction to the Team Software Process, Watts S. Humphrey, Pearson Education, 2000.

[10] Database Management Systems, Raghurama Krishnan, Johannes Gehrke, Tata Mc Graw Hill3rd Edition.