Novel AI-driven Malaria Prediction for Optimizing Public Health Management

Priya Vij[1], Patil Manisha Prashant[2]

[1]*Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India*

[2]*Research Scholar, Department of CS & IT, Kalinga University, Raipur, India*

KEYWORDS

Malaria, public health, prediction, management, augmented tree with penguin search optimization (AT+PSO), random forest (RF)

ABSTRACT

Malaria is an unmoving real common well-being apprehension in Asian nations like the Republic of India, where the state characterizes approximately 55% of the session events cutting-edge the infirmaries. They are usually distinguished by the absence of appropriate therapeutic care provision and the often late and error-prone diagnosis of the condition. In particular, commonly used devices such as the Rapid Diagnosis Test (RDT) are not completely dependable. They are primarily notable for their failure to provide adequate medical treatment and their tendency to diagnose the disease late and incorrectly. For example, widely used devices like the RDT are not consistent. To improve public health management actions, this work offers a unique augmented tree with penguin search optimization (AT+PSO) methodology for malaria forecasting. The suggested method combines the PSO algorithm with the augmented tree model, also known as random forest (RF). In the preprocessing stage, raw data samples are subjected to data normalization. Then, we applied the PSO to improve the characteristics of the RF model after successfully predicting malaria with the RF. The Python program is used to implement the suggested technique and analyze performance using a range of measures, including accuracy (0.988), sensitivity (0.987), specificity (0.991), F1-score (0.988), and MCC (0.975). In summary, our suggested approach produced the best results in terms of accuracy as opposed to other current strategies for predicting malaria to improve governance of public health.

1. Introduction

The World Health Organization's (WHO) enormous aim was currently being pursued by several different experts and groups. Recent advances in cognitive sciences have enabled complex quantitative and statistical mathematical methods of the transmission of malaria. These methods were effective at forecasting malaria prevalence for places with little knowledge of the neighborhood infection frequency [1]. Fortunately, the disadvantages are because they are complex and require a lot of computer resources to function. Reduced prices of modern therapies and malaria prevention initiatives are facing considerable hurdles due to the absence of investment, increases in medication and pesticide and diagnostic-resistant organisms, as well as distribution and application issues. As a result, the malaria pharmaceutical field has grown rapidly over the past 15 years [2]. Original goods include medicinal and immunology treatments like medications and vaccinations, as well as new mosquito control methods. Several recent researches have revealed indicators of risk for malaria-related deaths. Coma is among the prognostic factors. Despite advances in comprehending the causes and consequences of serious malaria, no clinical trials have been effective in lowering fatality rates. The medical care system has traditionally given priority to infectious illness response and management [3]. Infections caused by parasites rank higher than any other biological threat agents. Parasites, once thought to be a problem for low-income countries, have evolved to adapt to warming temperatures. Present therapies and efforts to control malaria face significant obstacles because of a lack of finance, increased medication and pesticide resistance, diagnostic-resistant parasitic organisms, and supply chain and implementation issues [12]. Antibodies called monoclonal antibodies, which provide malaria protection for several months and are safe to inject while pregnant, are being considered as prospective therapies for periodic malaria and protection for certain risk populations [7]. Given the enormous number of malaria interventions presently in growth, systematic methodologies are required to notify expansion conclusions and arrange new methods to ensure long-term resource commitment and rapid creativity [5]. The study aims to create and evaluate modern artificial intelligence methods able to correctly anticipate malaria epidemics.

Related works

Ezeand Asogwa [6] compared four deep learning algorithms for quick malaria case identification, with an emphasis on improving the precision of detection while minimizing computer resources and energy usage [10]. The results reveal that quantized versions of Binarized Convolutional Neural Network (B-CNN) and MobileNetV2 have improved detection accuracy, less memory utilization, and a quicker inference time. Sherrard-Smith et al. [13] created a methodology for deciding the communal well-being influence and cost-effectiveness of malaria preclusion interventions, as well as selecting the best vector control intervention packages for different finances and entomological parameters across recurrent malaria [4]. Nkiruka et al. [8] establish how policymakers can evaluate interventions impact and implement targeted strategies for high and low-burden regions. They offer a technique for fitting a generalized linear mixture method with a conditional autoregressive structure that employs the Bayesian hierarchical approaches and (Markov Chain Monte Carlo) MCMC implication. Ribeiro et al. [9] offer contraption knowledge approaches for the prediction of the incidence of malaria over 28 years using climate inconsistency in six nations. The characteristics exposed climate parameters that influence malaria incidence, with non-seasonal fluctuations in temperatures, rainfall, and energy from the surface all causal knowingly to outbreaks. Onuche-Ojo et al. [14] investigate the consequences of ecological alteration on malaria mosquitoes in the environment, arguing that improved water utilization might guide malaria control efforts, with field measurements finding a 40% chance of adolescent habitat presence.

2. Methodology

This section presents data gathering, a novel Augmented Tree model, and a Penguin Search Optimization (AT+PSO) strategy for estimating malaria forecasting.

Dataset description

The data set was collected from Kaggle. This dataset delivers an impression of malaria occurrence and humanity in India, representing equally the projected figures and their varieties (https://www.kaggle.com/datasets/imdevskp/malaria-dataset). Malaria, while being treatable and reversible, continues to be an important worldwide health issue. Children below the tender age of five are especially sensitive. This knowledge highlights the crucial need to speed up malaria preventive and therapy efforts, particularly in places most afflicted by the parasite.

Malaria forecasting using an augmented tree with penguin search optimization (AT+PSO)

The AT+PSO technique for plasmodium prediction combines Augmented Tree (AT) models and Penguin Search Optimization (PSO). The AT element incorporates extra information or patterns from malaria datasets, such as comprehensive patient demographics and environmental variables, to improve traditional tree-based models. PSO maximizes the model's variables and structure by emulating penguin browsing conduct, resulting in improved tuning of variables such as patient demographics, environmental conditions, and disease indicators. The method seeks to improve forecasting precision and robustness, providing a novel strategy to anticipate cases of malaria through sophisticated statistical simulation and optimization techniques. Improved random forest: The random forest technique is upgraded to compensate for combined data shortages, resulting in stronger flexibility. The W eliminating aberrant data is interpolation linearly, followed by completing the vector with objective adjustment. The Equation (1) represents a collection of weights or coefficients corresponding to different features such as patient demographics, (e.g., age, gender), environmental factors (e.g., temperature, humidity), and historical malaria incidence (e.g., previous cases, seasonal trends). By applying random forest regression, these weights are adjusted to optimize predictions. The random forest regression method predicts the contents according to the forecasting and development of malaria, persons might be divided into five stages: incubation, acute, chronic, severe, and recovered, and its combinations are utilized to calculate s_a times. The approaches to enhancing random forest design are listed below: In **step 1**, the omitted value W has been interpolated using the conventional linear extrapolation method, resulting for the purpose of predicting malaria, the medical record of an individual contains details on their appearance, biography, getting exposed the past, diagnosis, past medical interventions, and associated sequelae vector Q, represented as:

$$Q = [w_1, w_2, w_3, \dots, w_m] \tag{1}$$

In Equation (2), the average size of Q and W is identical ($n \times m$ multidimensional vector), where m is the number of restrictions during the mosquito bites to illnesses, from bites by mosquitoes to repeated indications, due to lapses with implications such as cerebral plasmodium and n, where every node represents a particular group of patients according to infections stage of life. In **step 2**, use row j in W to simply target the fang column and the other-1 columns in Q using the appropriate flexible support to generate the satisfying direction R_{fill}, where $J = 1, 2, 3, .. m$, signified below:

$$D(v) = 2(In(v-1) + \xi) - \frac{2(v-1)}{v} \tag{2}$$

Step 3 involves randomly selecting a training matrix C_{yu} from C to create an arbitrary decision tree in the random forest model. Popular **step 4**, use the appropriate adjustable support in R_{fill} as an effort and the board heavy post as the **Step 5** $s < s_a$, then $s = s + 1$, return to step 4; then, take the regular cost ofs_a calculations as the concluding recompense cost z_y, and change the misplaced value in W_j with z_y to comprehensively inform of environment W.

$$z_y = \frac{1}{s_a} \sum_{s=1}^{sa} z(s) \tag{3}$$

Step 6: If$j < m$, continue stages $j = j + 1$; alternatively, cease forecast and wide-ranging information return for$W$. The material provided above is planned to enhance the general phase movement of the system for random forests. Penguins Search Optimization Method: An optimization technique inspired by nature and based on penguin foraging behavior is called the Penguins Search Optimization Algorithm (PeSOA). Penguins are creatures that can live and scavenge in pairs. This method uses the penguin population to represent several feature subsets from a malaria dataset, including environmental factors, past malaria incidence, and patient demographics. Like divers, each group of penguins looks for pertinent information by exploring particular areas of the feature space. As well as communicating food quantities, penguins also share information about how successful their feature sets are, and they modify their search tactics appropriately. Penguins assemble and fine-tune their feature subsets following each cycle to increase the accuracy of their malaria risk prediction**.** In every cycle, the penguin's location with each new solution changes in the following manner: The method is carried out until the desired numbers of fish are obtained or the maximum number of repetitions is achieved. To improve prediction accuracy, every new solution optimizes variables such as patient demographics, environmental factors, and past malaria data.

$$C_{new} = C_{LastLast} + rand \times (W_{LocalBest} - W_{LocalLast}) \tag{4}$$

In equation (4), C_{new} represents the penguin's new location, C_{Last}represents the penguin's last position, and $W_{LocalBest}$and $W_{LocalLast}$represent in the context of PeSOA-based malaria forecasting, the local best concentrates on a particularly effective method under a particular patient subgroup or disease stage, whereas the global best finds the best model for projection worldwide to improve total accuracy of predictions. Rand () generates a randomized dispersed value. The method continues either a certain number of fish becomes available or the smallest amount of the current iteration limit has been achieved.

3. Results and discussion

For developing an AI-driven malaria prediction system, use a 16 GB RAM, Intel i7 setup with a 512 GB SSD. Comparing our proposed augmented tree with penguin search optimization (AT+PSO) with the existing techniques such as Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and CNNs with 2-level segmentation [11] and the performance was evaluated in various metrics such as accuracy, sensitivity, specificity, F1 score and MCC. **Accuracy:** This metric evaluates the general precision in malaria estimation; those with a positive result are those who have been successfully recognized as having malaria, while the negative results in patients have been correctly classified as not suffering from the disease. Accurate forecasts contribute to appropriate evaluation and therapy of the identification strategy by calculating the proportion of successfully identified coincidences in every instance such as positive and negative groups are there in malaria prediction. The research discovered that DNN with 0.986, pre-trained CNN with 0.986 and CNN with 2-level segmentation with 0.977 and the proposed system achieves 0.988, which is higher than the existing systems. **Sensitivity:** The standard is the fraction of actual positives accurately recognized by an examination, indicating its capacity to identify affirmative. The best in parasitic infections forecasting, our proposed method (0.987) was the highest in the range when comparing the existing methods based on the DNN (0.985), pre-trained CNN (0.981), and CNN with 2-level segmentation (0.971). **Specificity:** This percentage of genuine criticism correctly recognized by an examination indicates how effectively it detects non-positives. Artificial intelligence methods have been examined for malaria forecasting precision: (0.992) a pre-trained CNN, (0.988) DNN, and (0.972) two-level segmentation. The pre-trained CNN was the most precise, while the 32-layer DNN and 2-level segmentation CNNs can improve malaria forecasting and management of public health. The proposed method (0.991) has efficient specificity. **F1-score:** Average of precision and sensitivity, giving a balance result between two metrics. The following chart examines multiple malaria models for forecasting, with DNN 0.987, pre-trained CNN 0.987, and CNN with two-level segmentation (0.959). The suggested AT+PSO model achieves 0.988, indicating higher accuracy for prediction. Figure 1 shows the outcomes of the existing and proposed systems.

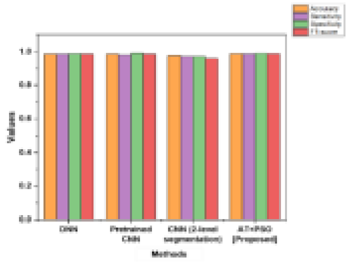

Figure 1. Outcomes of the existing and proposed methodologies

The Matthew's correlation coefficient (MCC): MCC measures the accuracy through discrete categorization through attractive into interpretation together factual and false positives and adverse effects, resulting in a neutral metric even for datasets with inconsistencies. The following Figure 2 illustrates the MCC for malaria prediction approaches, with the AT+PSO model getting the greatest MCC of 0.975, DNNs (0.973), demonstrating excellent forecasting (0.917) capabilities, and Convolutional Neural Networks (0.972) performing effectively. Compared with the existing methods, the proposed system (0.975) has the highest outcomes of the value.

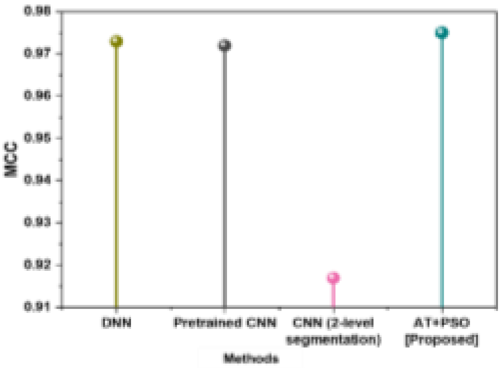

Figure 2. Outcomes of the MCC

Table 1 below examines the outcomes of existing approaches and shows that the suggested AT+PSO method has the highest accuracy (0.988) and F1 score (0.988), while maintaining good sensitivity and specificity.

Table 1. Results

Method	Accuracy	**Sensitivity**	**Specificity**	**F1-** score	**MCC**
DNN	0.986	0.985	0.988	0.987	0.973
Pre-trained CNN	0.986	0.981	0.992	0.987	0.972
CNN (2-level segmentation)	0.977	0.971	0.972	0.959	0.917
AT+PSO [Proposed]	**0.988**	**0.987**	**0.991**	**0.988**	**0.975**

4. Conclusion and future scope

Malaria forecasting is critical for optimizing public health management by providing accurate epidemic prediction using modern statistical techniques and artificial intelligence. This preventive strategy enables quick solutions, such as focused delivery of nets coated with insecticides and public health initiatives, which reduce the prevalence and morbidity. Implementing forecasting techniques within the hospital architecture improves the use of resources by directing medical equipment and workers to high-risk locations. As a result, health conditions improve, while medical organizations shift from responsive to preventative handling of illnesses, making an important boost to worldwide malaria prevention and elimination operations. To address these problems, our suggested AT+PSO method combines the Particle Swarm Optimization (PSO) algorithm with the RF method to improve plasmodium prediction reliability. This method starts with the standardization of raw data samples to achieve consistent and trustworthy results. The RF approach, which is known for its ability to handle complicated datasets, is further optimized using PSO, which fine-tunes the model's parameters to increase its accuracy for prediction.

Reference

[1] P.U. Eze, andC.O. Asogwa, “Deep machine learning model trade-offs for malaria elimination in resource-constrained locations,”*Bioengineering, 8*(11), p.150, 2021. https://doi.org/10.3390/bioengineering8110150

[2] E. Sherrard-Smith, P. Winskill, A.Hamlet, C. Ngufor, R.N’Guessan, M.W.Guelbeogo, A. Sanou, R.K. Nash, A. Hill, E.L. Russell, and M. Woodbridge, “Optimizing the deployment of vector control tools against malaria: a data-informed modeling study,” The Lancet Planetary Health, 6(2), pp.e100-e109, 2022.https://doi.org/10.1016/S2542-5196(21)00296-5

[3] O. Nkiruka, R. Prasad, andO. Clement,“Prediction of malaria incidence using climate variability and machine learning,” Informatics in medicine Unlocked, 22, p.100508, 2021.https://doi.org/10.1016/j.imu.2020.100508

[4] Srinivasa Rao, M., Praveen Kumar, S., & Srinivasa Rao, K. (2023). Classification of Medical Plants Based on Hybridization of Machine Learning Algorithms. Indian Journal of Information Sources and Services, 13(2), 14–21.

[5] E.S. Onuche-Ojo, B. Eseyin Joseph, D. Dako Apaleokhai, A. IzuafaBraimah, andJ. Nigeria, "Advancing Malaria Detection: A Comparative Study and Proposal for Web-Based Predictive Application Utilizing Convolutional Neural Network and TensorFlow," International Journal of Research and Innovation in Applied Science, 9(6), pp.222-232, 2024..https://doi.org/10.51584/IJRIAS.2024.906020

[6] J. Lubinda, Y. Bi, U. Haque, M. Lubinda,B.Hamainza, and A.J. Moore, "Spatio-temporal monitoring of health facility-level malaria trends in Zambia and adaptive scaling for operational intervention," Communications medicine, 2(1), p.79, 2022.https://doi.org/10.1038/s43856-022-00144-1

[7] Alamer, L., Alqahtani, I. M., & Shadadi, E. (2023). Intelligent Health Risk and Disease Prediction Using Optimized Naive Bayes Classifier. Journal of Internet Services and Information Security, 13(1), 01-10.

[8] M.D.B. Dieng, A.M. Tompkins, J. Arnault, A. Sié, B. Fersch, P.Laux, M. Schwarz, P.Zabré,S. Munga, S. Khagayi, and L. Diouf, "Process-based atmosphere-hydrology-malaria modeling: Performance for spatio-temporal malaria transmission dynamics in Sub-Saharan Africa," Water Resources Research, 60(6), p.e2023WR034975, 2024.https://doi.org/10.1029/2023WR034975

[9] M. Minakshi, T. Bhuiyan, S. Kariev, M. Kaddumukasa, D. Loum, N.B. Stanley, S. Chellappan, P. Habomugisha, D.W.Oguttu, and B.G. Jacob, "High-accuracy detection of malaria mosquito habitats using drone-based multispectral imagery and Artificial Intelligence (AI) algorithms in an agro-village peri-urban pastureland intervention site (Akonyibedo) in UnyamaSubCounty," Gulu District, Northern Uganda. Journal of Public Health and Epidemiology, 12(3), pp.202-217, 2020.https://doi.org/10.5897/JPHE2020.1213

[10] Rika, R., Bob, S.R., and Suparni, S. (2023). Comparative Analysis of Support Vector Machine and Convolutional Neural Network for Malaria Parasite Classification and Feature Extraction. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 14(3), 194-217.

[11] N, Ravi, and Swanand Kulkarni. 2023. "Smart Ways to Catch the Abutment DRCs at IP Level". Journal of VLSI Circuits and Systems 6 (1):51-54. https://doi.org/10.31838/jvcs/06.01.08.

[12] Suriya, M. Chandran, V. and M.G. Sumithra," Enhanced deep convolutional neural network for malarial parasite classification," International Journal of Computers and Applications, 44(12), pp.1113-1122, 2022.https://doi.org/10.1080/1206212X.2019.1672277

[13] M.H.D.M. Ribeiro, V.C. Mariani, andL. dos Santos Coelho,"Multi-step ahead meningitis case forecasting based on decomposition and multi-objective optimization methods," Journal of Biomedical Informatics, 111, p.103575, 2020.https://doi.org/10.1016/j.jbi.2020.103575

[14] T. Awine, and S.P. Silal, "Accounting for regional transmission variability and the impact of malaria control interventions in Ghana: a population level mathematical modelling approach," Malaria Journal, 19, pp.1-21, 2020 .https://doi.org/10.1186/s12936-020-03496-y

[15] M.E. Sinka, S.Pironon, N.C. Massey,J.Longbottom, J.Hemingway, C.L. Moyes, and K.L. Willis, "A new malaria vector in Africa: predicting the expansion range of Anopheles stephensi and identifying the urban populations at risk," Proceedings of the National Academy of Sciences, 117(40), pp.24900-24908, 2020.https://doi.org/10.1073/pnas.2003976117