Prediction of Malaria Fever Using Long-Short-Term Memory and Big Data

Joseph Bamidele Awotunde(✉) , Rasheed Gbenga Jimoh ,
Idowu Dauda Oladipo , and Muyideen Abdulraheem

Department of Computer Science, University of Ilorin, Ilorin, Nigeria
{awotunde.jb,jimoh_rasheed,odidowu,muyideen}@unilorin.edu.ng

Abstract. Malaria has been identified to be one of the most common diseases with a great public health problem globally and it is caused by mosquitos' parasites. This prevails in developing nations where healthcare facilities are not enough for the patients. The technological advancement in medicine has resulted in the collection of huge volumes of data from various sources in different formats. A reliable and early parasite-based diagnosis, identification of symptoms, disease monitoring, and prescription are crucial to decreasing malaria occurrence in Nigeria. Hence, the use of deep and machine learning models is essentials to reduce the effect of malaria-endemic and for better predictive models. Therefore, this paper proposes a framework to predict malaria-endemic in Nigeria. To predict the malaria-endemic well, both environmental and clinical data were used using Kwara State as a case study. The study used a deep learning algorithm as a classifier for the proposed system. Three locations were selected from Irepodun Local Government Areas of Kwara State with 34 months periodic pattern. Each location reacted differently based on environmental factors. The findings indicate that both factors are significant in malaria prediction and transmission. The LSTM algorithm provides an efficient method for detecting situations of widespread malaria.

Keywords: Big data · Prediction · Diagnosis · Malaria fever · Deep learning · Mosquito parasites

1 Introduction

Global warming has been shown to contribute to the rising abundance of human insect vector diseases, especially in tropical countries [1, 2]. The world has been fighting with mosquito-borne diseases, which have significantly be a threat to health worldwide [3, 4]. To ultimately eliminate and control the widespread of insect-borne diseases like malaria, there is an ongoing effort [5–7]. Apart from Christopher's work of 1911 [8], there have been different models recently for prediction malaria spread and abundance. Malaria prediction and treatment are very difficult and complicated due to the transmission of mosquitoes and complex ecology [9–12].

Malaria is one of the precarious illnesses in Africa especially Nigeria [13, 14]. A reliable and early parasite-based diagnosis, identification of symptoms, prescription, and

further malaria regulation plans, For example, the use of insecticide-medicated bed-net and indoor residual spray is crucial to reducing the incidence of malaria in Nigeria. Many donors have been helpful in this regard by distributing treated bed nets in Nigeria. The Plasmodium falciparum opposition to generally known anti-malarial medications imposes the consumption of potentially and extra cost-effective combination therapy in addition to accurate laboratory diagnosis [15]. To prevent the undiscerned use of anti-malaria medicines, the World Health Organization (WHO) stresses the occurrence of malaria in a human body before the use of or treatment of malaria drugs. But in Africa especially in Nigeria the presumptive treatment of malaria and other related fevers is still popular due to a lack of medical personnel, technical expertise, and effective laboratory structure. The aforementioned problems have cause delays in the diagnosis of other severe feverish diseases and the assumption of treatment sometimes results in misuse or abuse of anti-malarial drugs [16, 17].

Malaria is as old as a human being because it is an ancient disease, which causes social, economic, and health burden amongst people in the world [9]. The disease is common in the warm and humid nations and the ailment happened to be in existence for more than hundreds of centuries now, malaria has remained a prominent community well-being challenges among many countries. WHO declared malaria endemic in 109 countries in the year 2008 with 243 million malaria conditions recorded and over millions of demises of the endemic which are mainly children under 5 years of age (WHO).

Nigeria in recent years experiences a high incidence of malaria. The healthcare system is characterized by an adequate large quality of data, with little or no effort to apply the robust information of data contained for solving life-threatening problems in the medical diagnosis of diverse diseases [10]. The data mining approach remains the most significant method among various techniques for the prediction or diagnosis of several diseases [18]. There have been many proposed prediction models by researchers on environmental factors [19–23]. The use of big data for predictions based on environmental factors and clinical conditions has not been explored [23].

Therefore, this study aims to predict the incidence of malaria using clinical and environmental factors with the help of the LSTM prediction model in the R programming language environment. Finally, the paper conducts a comparative performance evaluation on the prediction algorithms. The LSTM model was used on big data to predict malaria prevalence in the study areas. The study required a dataset and the data collected from the study area were used for the model. The deep learning method was chosen because of its strength, prediction, and forecasting ability. the predictions are made based on past values and a relationship has to be established between the fluctuations occurred in values.

2 Big Data in the Healthcare System

The advancement in information technology increases the amount of data used by the modern establishment thereby making data science an essential tool for maintaining data in any organization [24–26]. For instance, an organization with a huge amount of user data needs data science to effectively improve methods to collect, accumulate, and process the data. Various systematic techniques may be used by the company to process and obtain

resourceful outcomes about the user data. Data mining and big data are used interrelated with Data Science. To comprehend and process actual data concepts, data science is of great importance as it incorporates the field of statistics and machine learning with their various procedures [27, 28]. The study of Information Science, Computer Science, Mathematics, and Statistics with their concepts and methods are all embedded in the field of data science [28, 29].

The 21st century is an age of big data affecting every aspect of human life, including biology and medicine [30, 31]. The move from paper medical records to Electronic Health Record (EHR) systems has resulted in an unprecedented increase in data [32, 33]. Big data thus offers a great opportunity for doctors, epidemiologists, and specialists in health policy to make evidence-driven decisions that will eventually enhance patient care [34, 35]. "Big data is not only a modern reality for the biomedical scientist but an imperative that needs to be fully grasped and used in the search for new knowledge" [36–38].

Data in the healthcare sector are usually huge and not easy to handle [38–40]. This results from the enormous way by which data grows in the health care sector, the rate at which data are been produced, and the variety of various data in the healthcare system [41, 42]. The rate at which data are been captured, stored, analyzed, and retrieved in the health care sector has swift rapidly from the aged paper-based storage technique to the use of digital technique and method [42, 43].

On the other hand, the complication of data makes processing and analysis of data by the aged long traditional method very difficult and uneasy to handle. However, the large volume of data, as well as the sophistication of the data, make it difficult for conventional methods and techniques to process and interpret the data. Therefore, the application of advanced technologies which includes virtualization and cloud computing allows for huge and effective data processing in the healthcare system. Thereby rapidly turning the healthcare system into a big data industry. Nevertheless, in these modern days, the improvement in the information and communication technologies (ICT) brings the advancement of varied data from new sources in the healthcare system.

This source includes the Global positioning system (GPS), data from gene sequence, file logs, devices that identify Radio Frequency (RFID), smart meters, and posts from social media. The increasing rate with which data is been produced from various sources brings about an increase in the amount of data in the healthcare system [42, 44]. Thus, their results give tedious means of storing, processing, and analyzing data with the aged long traditional method of data processing applications [45]. Nevertheless, modern methods and techniques, as well as advanced computational technology, have been used to store, control, and analyze values from broad and varied data in the health care system in real-time [42, 44, 45]. As a result, the healthcare sector has now become a big data industry. Big data now provides massive opportunities for the healthcare system [46].

Improvement in information technology and data computing has greatly changed researches on population-based by encouraging easy access to a huge amount of data. Sometimes, such database links are referred to as "big data" [47, 48]. In other to make efficient use of these data for researches in clinical health or public health, the researchers need to widen researches further than the traditional surveillance model, as operating with big data differs from focusing on performing narrow analysis, treatment-oriented

clinical data. Therefore, leveraging on Big data to reflect accurately on the heterogeneous population it represents becomes expedient [42, 45]. This endeavor needs a swift research environment that can adopt a quick advancement in computing technology to at all-time combine data while making use of new methods to reduce their complexity [42, 48].

Big data in healthcare have been established for its timely advancement in disease detection, diagnosis, and enable the best control of any disease outbreaks. Predictive analysis of health care can be achieved easily with the aid of big data [49, 57]. In the United States of America, for example, predictive analytics was used to enhance disease response management and a deeper understanding of diseases [50]. The use of a health-based framework for broad data collection to enhance patient care and prevent insurance fraud has been widely used in Australia [51]. The use of data from social media, sensor data, and air quality data with machine learning models are currently studies to predict asthma [23]. Generally, the use of broad health data in the predictive analysis has created considerable interest in research in recent years [52]. For the management of complex computational data, the modern scalable database has been used examples of such are Hadoop MapReduce and Apache Spark. The latter does not come with a file system other than Hadoop, so you need to combine it with a different file system based on the cloud [53]. Hadoop spends time running computationally complex machine learning algorithms [54, 55], making Apache Spark 100 times faster than ever before. Also, trends and patterns which make it easy to diagnose and treat patients are been revealed by big data.

Big data would be useful in today's digital world as the key to controlling malaria outbreaks, but the criteria for successful data collection and global analysis need to be clear. The study states that the data and algorithms available in digital form are used for prediction and monitoring. For example, it is vital to consider both the clinical and environmental factors where the infection has spread and been detected in the battle against endemic malaria. Nonetheless, it is equally important that these data and algorithms are used in a protected manner, following data security laws and with proper regard for privacy and privacy. Inability to do so would undermine public confidence, making people less able to follow advice or recommendations on public health, and more likely to have poor health outcomes [56].

Hence the exploitation of big data in malaria-endemic will bring about improved care with minimum cost and good satisfaction to patients. On Big Data, the LSTM model was used to estimate the prevalence of malaria in the study areas. For the model, the analysis needed a dataset, and the data obtained from the study area was used. Because of its power, prediction, and forecasting ability, the deep learning method was chosen. The estimates are made based on previous values and a correlation must be formed between the variations in values that occur.

2.1 Clinical Data

The clinical data used for this study were collected from three General Hospital in Irepodun Local Government (LGA) of Kwara State (Omu-Aran, Oro, and Agbamu), the health centers (HC) were set up by the Kwara State Government of Nigeria. This HC was used to collect samples of malaria from endemic sites within Irepodun LGA. The three hospitals were chosen because they are government-owned hospitals and were also

allowed to control endemic malaria in Kwara State and are in highly endemic areas. Such hospitals were well equipped with malaria test kits and their personnel were also well trained in the treatment and testing of malaria. Hospital data attributes are the number of successful cases, symptom-free, regular age, sex, per month. The data used for this analysis covered three years, from the beginning of 2017 to the end of 2019.

2.2 Environmental Data

It is very essential to know the environmental risk even if it is at a very local scale for malaria-endemic. The most considered environmental factors are temperature, relative humidity, rainfall, and the vegetative index to estimate malaria abundance. Environmental data were obtained from Minet's daily forecast and the satellites. From the MiNet forecast, the MiNet01AB product predicted a Composite picture of 7 days at 1.23 km * 1.23 km of spatial resolution for day and night temperature. Nigerian Meteorological Agency's data acquisition center reported normal relative modesty at 2 m and 8 m above ground level. The vegetative index was obtained by resolution of 1.5 * 1.5. Rainfall was measured based on the amount of rainfall in each region. The following information was also systematically collected: the vegetation index (NDVI) measured at 1 km around the site with SPOT 5, the number of residents, the existence of abandoned projects, and the existence of abandoned construction with tools or holes that could serve as possible breeding sites for Anopheles mosquitoes.

3   Methodology

Hochreiter [29] introduced Long-Short -Term Memory (LSTM) to tackle the Recurrent Neural Network (ANN) reliability and speed problems. Both are very similar but in LSTM a new concept is introduced with interaction per module or one cell. RNN uses a definition of time and is a neural network that is aided by edges spanning adjacent steps in time. It's mutual parameters over a series of time steps and allows connections to self-loop over time from a node to itself. The model receives the input values at step t and the node's previously hidden value and determines the hidden state. The final output value in the prior state will be determined by the input value. LSTM was used to tackle the issue of vanishing gradients appropriately. Unlike RNN, which uses hidden layer structure, LSTM is a linear self-loop memory block that flows through the gradients with long sequences.

The memory block containing self-connecting memory cells is used to memorize the time state, and the three gates from the memory block help control the flow of information within and outside the memory block. It helps the model to manage long-term and short-term correlation in time series. The LSTM registered at time t, instead of a simple RNN. As depicted in Fig. 1, LSTM has a four-layer long-term training process with retrieving information called chain-like structure, thus consisting of memory blocks such as cells. The vital condition called the state of the cell remains unchanged and enables data to flow forward. Nonetheless, useable sigmoid activation gates may be used to add or remove data inside a layer. Applying gates when memorizing the LSTM will prevent long-term dependence. Such gates have a range of matrix operations including individual weights.
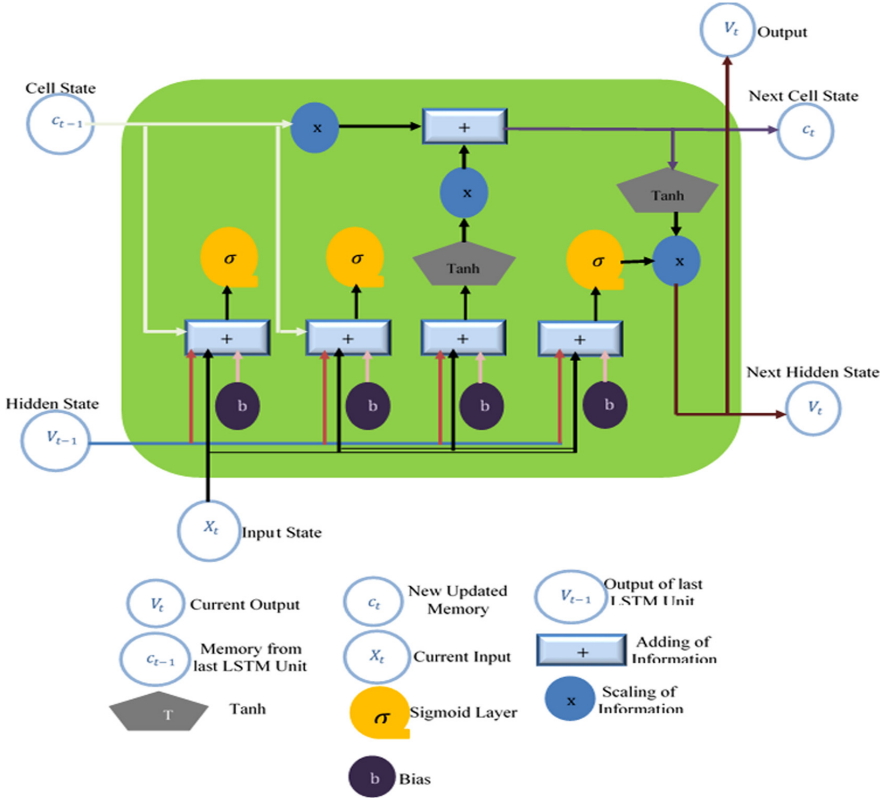
Fig. 1. Development of the Long-Term Memory (LSTM)

3.1 The Staging Procedure of LSTM Is as Follows

The sigmoid function of LSTM identifies data that is not needed for any phase and takes the current input as (X_t) at time $t-1$ and produces output (V_{t-1}) at time $t-1$. The sigmoidal function specifies which part of the output will be separated from the old output. This stage is called the Forgetting Gate (f_t). Forget the f_t gate, the value is between 0 and 1 and is a vector that matches each number in the cell (C_{t-1}).

$$f_t = \sigma\left(w_f\left[V_{t-1}, X_t\right] + b_f\right) \tag{1}$$

Where
σ is the sigmoidal function, w_f = weights and b_f = forget gate.

Both equations have two states: the ignoring condition and the storing condition of the current input and the X_t in the cell state. The two layers are the sigmoidal layer and the tanh layer. To the sigmoidal layer provided to decide if the new information needs to be modified or not using 0 or 1. The *tanh* update weights in the second layer and passes the values between (-1 to 1). Using their level of importance, the values are selected

accordingly. Both the values are updated as presented in Eq. 4 and form the new cell state.

$$m_t = \sigma\left(w_f\left[V_{t-1}, X_t\right] + b_m\right) \tag{2}$$

$$N_t = tanh\left(w_f\left[V_{t-1}, X_t\right] + b_n\right) \tag{3}$$

$$C_t = C_{t-1}f_t + N_t m_t \tag{4}$$

In the final stage, the output V_t is based on the output of sigmoidal gates Q_t and is multiplied by the new $tanh(C_t)$ layer generated.

$$Q_t = \sigma\left(w_f\left[V_{t-1}, X_t\right] + b_q\right) \tag{5}$$

$$V_t = Q_t tanh(C_t) \tag{6}$$

w_q and b_q are the weights and biases for the output gates, respectively

3.2 Model

As input to predict malaria fever was used the clinical factors with asymptomatic cases and environmental data such as temperature, relative humidity, rainfall, vegetative index. The average rainfall recorded, the average temperature of day and night time, and the average enhanced vegetable index at a resolution of 1.5 km, thus combined to be monthly. The problem of missing variables of about 7% was tackled by clustered equations in pre-processing data with multivariate imputation. From 2017–2019 the data were taken from both the clinical and environmental data at each location (Omu-Aran, Oro, and Agbamu) in Irepodun LGA as presented in Fig. 1.

The model was trained from January 2017 to December 2019 using the preprocessed data. Following some iterations, the model was subsequently updated for 2 years to provide a monthly malaria abundance forecast for 24 months, based on historical data on environmental and clinical variables. The proposed model used an open-source online cloud big data to implement the system using the Apache Spark framework. The Apache Kafka was used for the real-time streaming and batch processing of the data on Spark. The output measurements for the model were determined and model results were produced for each geographic location.

The experiment was split into two phases: i) offline batch processing ii) streaming online. During the first stage, the data is pre-processed and used for processing. In the second stage, the real-time analysis was carried out by gathering the batches of message brokers. Initially, preprocessing was optimized by applying multivariate variable imputation using a clustered equation to manage missing data points for each region's ambient and clinical time-series data. Then the LSTM classifier was used to make the preparation. In the second stage, the results were expected in real-time analysis, and the results over a given duration were estimated. The findings were ultimately processed for on-line analysis on the distributed Hadoop file system.

4   Results and Discussion

The experiment has been divided into two phases: i) offline batch processing ii) streaming on-line. The data are pre-processed in the first stage and used for processing. In the second stage, the real-time analysis was performed by collecting the message brokers in batches. Preprocessing was initially optimized by applying multivariate imputation using a clustered equation to manage missing data points on the environmental and clinical time series of each venue. Later, the LSTM classifier was used for the preparation. The results were projected for a certain amount of time during the second phase of the real-time analysis, and the results were estimated. The findings were eventually stored on the online reporting file system provided by Hadoop.

Throughout the analysis, the variables depend seasonally on environmental and clinical factors. For example, the number of asymptomatic cases and the vegetative index are time series, following [32]. Therefore, the research included the number of asymptomatic patients and their diagnoses. Between January 2017 until December 2019, the study expected cases of malaria is for 24 months. The Omu-Aran site reported the highest number of incidents, which may be attributed to the large number of people residing in the region and the high population density compared with other sites. The maximum rainfall was 2.3 mm, and the maximum day and night temperatures differed significantly from the other two locations by 2.5 °C. The results for the vegetative index were also significantly different from those locations. General Hospital, Agbamu has reported the smallest number of malaria cases. It was because i) the area's population was very small relative to the other two sites and ii) we also wanted to assume that the areas were using local herbs to manage their malaria index, which could minimize the index reported in the hospital as in two other areas.

Precise health prediction models are very important for improved and more effective care and are useful for planning and decision-making. The results of the study have shown that clinical and environmental factors are resourceful and also have played a crucial role in the forecasting of malaria. Individuals with malaria without symptoms were present in the studied region, and this was verified with laboratory tests with high predictive capacity. Environmental factors play a significant role in the prediction of malaria since rainfall normally started around July to October of each year, the incidence and incidence of malaria was higher than in other months of the year. It was attributed to the fact Anopheles gambiae mosquito breeding aid of the season during this period. Factors like the presence of a watercourse near any residential area and a higher index of vegetation correlated with higher vector density support the eminences of malaria. Table 1 displays the statistical findings at the site selected for the malaria cases.

Table 1. Results of malaria incidences in the selected areas

Location	Minimum	Maximum	Mean	Standard deviation	Mode
Omu–Aran	27	56	34.45	3.57	43
Oro	25	45	31.23	3.03	43
Agbamu	16	21	15.05	5.29	10

4.1 Performance Evaluation Metrics

The basic evaluation criteria and assessment are based on the classification accuracy level using True Negative (TN), True Positive (TP), False Negative (FN), and False Positive (FP) (Table 2).

Table 2. Definition of performance evaluation metrics

Measures	Definition	Formula
Sensitivity test	The sensitivity test is the ability to correctly detect diabetes disease	$TP/(TP + FN)$
Specificity test	Specificity is the ability to avoid calling normal things as diabetes disease in medical tests	$TN/(TN + FP)$
PPV	Positively predicated value	$TP/(TP + FP)$
NPV	Negatively predicated value	$TN/(TN + FN)$
Accuracy	Measure the percentage of correctly detect diabetes disease	$\frac{TN+TP}{TN+FP+FN+TP} \times 100$

Where:
True Positive = TP, True Negative = TN, False Positive = FP, and False Negative = FN

Table 3 shows the performance evaluation of the three locations in the study, Agbamu has the highest accuracy of 98.34%, 99.37% sensitivity 93.78% specificity, and 98.14% precision. Oro has an accuracy of 97.94%, 98.45% sensitivity 94.05% specificity, and 97.56% precision and the least of them which is Omu-Aran has an accuracy of 97.47%, 98.05% sensitivity 94.67% specificity, and 97.56% precision.

Table 3. Performance evaluation for the metrics

Location	Accuracy	Sensitivity	Specificity	Precision
Omu-Aran	97.47%	98.05%	94.67%	97.75%
Agbanu	98.34%	99.37%	93.78%	98.14%
Oro	97.94%	98.45%	94.05%	97.56%

Table 4 shows a comparison of the proposed method with other machine learning. The methods outperformed other machines learning this maybe because of the introduction of big data into the LSTM deep learning and result bring about better accuracy compared with other machine learning. Random Forest follows with an accuracy of 93.94% and ANN is the least of all the methods with an accuracy of 25.83%.

Table 4. Comparison of the proposed method with other machine learning

Location	Accuracy	Specificity	Precision
KNN	86.21%	98.66%	92.90%
SVM	92.69%	97.09%	91.82%
Random Forest	93.94%	94.51%	85.42%
Naïve Bayes	91.69%	96.42%	88.36%
ANN	25.83%	0.0	25.83%
Big Data with LSTM	**98.34%**	**94.67%**	**98.14%**

5 Conclusions

Malaria fever is a social problem in that it can widespread harm and cause personal damage. It is also a worldwide problem face by many developing nations. Hence, many types of research have been conducted to minimize the social, economic, and physical losses by predicting the spread of malaria-endemic. The purpose of this study was to design a prediction model using deep learning techniques and big data that is more appropriate than existing models by using environmental and clinical variables. Environmental and clinical factors play a significant role in malaria-endemic prediction. Though malaria prediction using environmental and clinical variables with big data analysis is a new approach but has proved important and very useful in predicting malaria-endemic. The application of correct predictive models are noteworthy for the allocation of medical resources for prevention and to estimate the impact of the disease. The use of a more accurate exploration of malaria prediction and diagnosis of malaria is necessary to improve treatment and its value in practice. The proposed method outperformed other machine learnings compare with an overall accuracy of 98.34%, 94.67%, and 98.14% precision. Future work can still work on other areas within the study area to be able to conclude accurately about the features that cause malaria within Kwara State. Also, features selection algorithms like genetic algorithms can be used for feature selection.

References

1. Kiang, R., et al.: Meteorological, environmental remote sensing, and neural network analysis of the epidemiology of malaria transmission in Thailand. Geospatial Health, 71–84 (2006)
2. Wilke, A.B., Beier, J.C., Benelli, G.: The complexity of the relationship between global warming and urbanization–an obscure future for predicting increases in vector-borne infectious diseases. Current Opinion Insect Sci. **35**, 1–9 (2019)
3. Benelli, G.: Green synthesized nanoparticles in the fight against mosquito-borne diseases and cancer—a brief review. Enzyme Microb. Technol. **95**, 58–68 (2016)
4. Wilder-Smith, A., Gubler, D.J., Weaver, S.C., Monath, T.P., Heymann, D.L., Scott, T.W.: Epidemic arboviral diseases: priorities for research and public health. Lancet. Infect. Dis **17**(3), e101–e106 (2017)
5. Rabinovich, R.N., et al.: malERA: an updated research agenda for malaria elimination and eradication. PLoS Med. **14**(11), e1002456 (2017)

6. Zolnikov, T.R.: Vector-borne disease. In: Autoethnographies on the Environment and Human Health, pp. 113–126. Palgrave Macmillan, Cham (2018)

7. Sougoufara, S., Ottih, E.C., Tripet, F.: The need for new vector control approaches targeting outdoor biting Anopheline malaria vector communities. Parasit. Vectors **13**(1), 1–15 (2020)

8. Christophers, S.R.: Epidemic malaria of the Punjab: with a note of a method of predicting epidemic years. Trans. Committee Stud. Malaria India **2**, 17–26 (1911)

9. Awotunde, J.B., Matiluko, O.E., Fatai, O.W.: Medical diagnosis system using fuzzy logic. Afr. J. Comp. ICT **7**(2), 99–106 (2014)

10. Ayo, F.E., Awotunde, J.B., Ogundokun, R.O., Folorunso, S.O., Adekunle, A.O.: A decision support system for multi-target disease diagnosis: a bioinformatics approach. Heliyon **6**(3), e03657 (2020)

11. Zinszer, K., et al.: Forecasting malaria in a highly endemic country using environmental and clinical predictors. Malaria J. **14**(1), 245 (2015)

12. Rochlin, I., Ninivaggi, D.V., Benach, J.L.: Malaria and Lyme disease-the the largest vector-borne US epidemics in the last 100 years: success and failure of public health. BMC Public Health **19**(1), 804 (2019)

13. WHO: World malaria report 2013. World Health Organization, Geneva (2013)

14. Adebiyi, M., et al.: Computational investigation of consistency and performance of the biochemical network of the malaria parasite, *Plasmodium falciparum*. In: Misra, S., et al. (eds.) ICCSA 2019. LNCS, vol. 11623, pp. 231–241. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24308-1_19

15. Mutabingwa, T.K.: Artemisinin-based combination therapies (ACTs): best hope for malaria treatment but inaccessible to the needy! Acta Trop. **95**, 305–315 (2005)

16. Leslie, T., et al.: Overdiagnosis and mistreatment of malaria among febrile patients at primary healthcare level in Afghanistan: an observational study. BMJ **345**, e4389 (2012)

17. Bastiaens, G.J.H., Bousema, T., Leslie, T.: Scale-up of malaria rapid diagnostic tests and artemisinin-based combination therapy: challenges and perspectives in sub-Saharan Africa. PLoS Med. **11**, e1001590 (2014)

18. Abisoye, O.A., Jimoh, R.G.: A hybrid intelligent forecasting model to determine malaria transmission. AIT **2015**, 5 (2015)

19. Oladele, T.O., Ogundokun, R.O., Awotunde, J.B., Adebiyi, M.O., Adeniyi, J.K.: Diagmal: a malaria coactive neuro-fuzzy expert system. In: Gervasi, O., et al. (eds.) ICCSA 2020. LNCS, vol. 12254, pp. 428–441. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58817-5_32

20. Ayo, F.E., Ogundokun, R.O., Awotunde, J.B., Adebiyi, M.O., Adeniyi, A.E.: Severe acne skin disease: a fuzzy-based method for diagnosis. In: Gervasi, O., et al. (eds.) ICCSA 2020. LNCS, vol. 12254, pp. 320–334. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58817-5_25

21. Davis, J.K., et al.: A genetic algorithm for identifying spatially-varying environmental drivers in a malaria time series model. Environ. Model Softw. **119**, 275–284 (2019)

22. Gomez-Elipe, A., Otero, A., Van Herp, M., Aguirre-Jaime, A.: Forecasting malaria incidence based on monthly case reports and environmental factors in Karuzi, Burundi, 1997–2003. Malaria J. **6**(1), 129 (2007)

23. Santosh, T., Ramesh, D., Reddy, D.: LSTM based prediction of malaria abundances using big data. Comput. Biol. Med. **124**, 103859 (2020)

24. Wang, Y., Kung, L., Byrd, T.A.: Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. Technol. Forecast. Soc. Chang. **126**, 3–13 (2018)

25. Choi, T.M., Chan, H.K., Yue, X.: Recent development in big data analytics for business operations and risk management. IEEE Trans. Cybern. **47**(1), 81–92 (2016)

26. Donoho, D.: 50 years of data science. J. Comput. Graph. Stat. **26**(4), 745–766 (2017)

27. Shi, B., Iyengar, S.S.: General framework of mathematics. Mathematical Theories of Machine Learning - Theory and Applications, pp. 13–16. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-17076-9_2

28. Okewu, E., Misra, S., Lius, F.-S.: Parameter tuning using adaptive moment estimation in deep learning neural networks. In: Gervasi, O., et al. (eds.) ICCSA 2020. LNCS, vol. 12254, pp. 261–272. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58817-5_20

29. Blei, D.M., Smyth, P.: Science and data science. Proc. Natl. Acad. Sci. **114**(33), 8689–8692 (2017)

30. Hulsen, T., et al.: From big data to precision medicine. Front. Med. **6**, 34 (2019)

31. Baro, E., Degoul, S., Beuscart, R., Chazard, E.: Toward a literature-driven definition of big data in healthcare. BioMed Res. Int. **2015** (2015)

32. Saweros, E., Song, Y.-T.: Connecting heterogeneous electronic health record systems using tangle. In: Lee, S., Ismail, R., Choo, H. (eds.) IMCOM 2019. AISC, vol. 935, pp. 858–869. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-19063-7_68

33. Austin, C., Kusumoto, F.: The application of Big Data in medicine: current implications and future directions. J. Interv. Card. Electrophysiol. **47**(1), 51–59 (2016)

34. Fiske, A., Buyx, A., Prainsack, B.: Health information counselors: a new profession for the age of big data. Acad. Med. **94**(1), 37 (2019)

35. Galetsi, P., Katsaliaki, K., Kumar, S.: Values, challenges, and future directions of big data analytics in healthcare: a systematic review. Soc. Sci. Med., 112533 (2019)

36. Williamson, B.: Big Data in Education: The Digital Future of Learning, Policy, and Practice. Sage, London (2017)

37. Krumholz, H.M.: Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. Health Aff. **33**(7), 1163–1170 (2014)

38. Lacroix, P.: Big data privacy and ethical challenges. In: Househ, M., Kushniruk, Andre W., Borycki, Elizabeth M. (eds.) Big Data, Big Challenges: A Healthcare Perspective. LNB, pp. 101–111. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-06109-8_9

39. Metaxiotis, K.: Healthcare knowledge management. In: Encyclopedia of Knowledge Management, 2nd edn., pp. 366–375. IGI Global (2011)

40. Halder, P., Pan, I.: Role of Big data analysis in healthcare sector: a survey. In: 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), pp. 221–225. IEEE, November 2018

41. Dai, H.N., Wang, H., Xu, G., Wan, J., Imran, M.: Big data analytics for manufacturing internet of things: opportunities, challenges, and enabling technologies. Enterp. Inf. Syst., 1–25 (2019)

42. Olaronke, I., Oluwaseun, O.: Big data in healthcare: prospects, challenges, and resolutions. In: 2016 Future Technologies Conference (FTC), pp. 1152–1157. IEEE, December 2016

43. Tresp, V., Overhage, J.M., Bundschus, M., Rabizadeh, S., Fasching, P.A., Yu, S.: Going digital: a survey on digitalization and large-scale data analytics in healthcare. Proc. IEEE **104**(11), 2180–2206 (2016)

44. Oussous, A., Benjelloun, F.Z., Lahcen, A.A., Belfkih, S.: Big data technologies: a survey. J. King Saud Univ.-Comput. Inf. Sci. **30**(4), 431–448 (2018)

45. Sagiroglu, S., Sinanc, D.: Big data: a review. In: 2013 International Conference on Collaboration Technologies and Systems (CTS), pp. 42–47. IEEE (2013)

46. Villars, R.L., Olofson, C.W., Eastwood, M.: Big data: what it is and why you should care. White Paper, IDC, 14, 1–14 (2011)

47. Priyanka, K., Kulennavar, N.: A survey on big data analytics in health care. Int. J. Comput. Sci. Inf. Technol. **5**(4), 5865–5868 (2014)

48. Kruse, C.S., Goswamy, R., Raval, Y.J., Marawi, S.: Challenges and opportunities of big data in health care: a systematic review. JMIR Medical Inform. **4**(4), e38 (2016)

49. Abayomi-Alli, A., Abayomi-Alli, O., Vipperman, J., Odusami, M., Misra, S.: Multi-class classification of impulse and non-impulse sounds using deep convolutional neural network (DCNN). In: Misra, S., et al. (eds.) ICCSA 2019. LNCS, vol. 11623, pp. 359–371. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24308-1_30

50. Birkhead, G.S., Klompas, M., Shah, N.R.: Use of electronic health records for public health surveillance to advance public health. Annu. Rev. Public Health **36**, 345–359 (2015)

51. Cohen, I.G., Amarasingham, R., Shah, A., Xie, B., Lo, B.: The legal and ethical concerns that arise from using complex predictive analytics in health care. Health Aff. **33**(7), 1139–1147 (2014)

52. Ram, S., Zhang, W., Williams, M., Pengetnze, Y.: Predicting asthma-related emergency department visits using big data. IEEE J. Biomed. Health Inform. **19**(4), 1216–1223 (2015)

53. Santosh, T., Ramesh, D.: DENCLUE-DE: differential evolution based DENCLUE for scalable clustering in big data analysis. In: Smys, S., Senjyu, T., Lafata, P. (eds.) ICCNCT 2019. LNDECT, vol. 44, pp. 436–445. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-37051-0_50

54. Ayeni, F., Misra, S., Omoregbe, N.: Using big data technology to contain current and future occurrence of ebola viral disease and other epidemic diseases in West Africa. In: Tan, Y., Shi, Y., Buarque, F., Gelbukh, A., Das, S., Engelbrecht, A. (eds.) ICSI 2015. LNCS, vol. 9142, pp. 107–114. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20469-7_13

55. Behera, R.K., Jena, M., Rath, S.K., Misra, S.: Co-LSTM: convolutional LSTM model for sentiment analysis in social big data. Inf. Process. Manage. **58**(1), 102435 (2021)

56. Ward, P.R.: Improving access to, use of, and outcomes from public health programs: the importance of building and maintaining trust with patients/clients. Front. Public Health **5**, 22 (2017)

57. Okewu, E., Misra, S., Fernandez, S.L., Ayeni, F., Mbarika, V., Damaševičius, R.: Deep neural networks for curbing climate change-induced farmers-herdsmen clashes in a sustainable social inclusion initiative. Problemy Ekorozwoju **14**(2) (2019)