

Research Article

Probabilistic Model-Based Malaria Disease Recognition System

Rahila Parveen ^{1,2}, Wei Song ¹, Baozhi Qiu ², Mairaj Nabi Bhatti ³, Tallal Hassan ⁴,
and Ziyi Liu ⁵

¹Henan Academy of Big Data, Zhengzhou University, Henan, Zhengzhou, China

²School of Information Engineering, Zhengzhou University, Henan, Zhengzhou, China

³Department of Information Technology, Shaheed Benazir Bhutto University, Shaheed Benazirabad, Nawabshah, Pakistan

⁴School of Business, Zhengzhou University, Henan, Zhengzhou, China

⁵College of Engineering & Science, University of Detroit Mercy, Detroit, USA

Correspondence should be addressed to Wei Song; iewsong@zzu.edu.cn

Received 14 October 2020; Revised 8 December 2020; Accepted 17 December 2020; Published 6 January 2021

Academic Editor: Dr Shahzad Sarfraz

Copyright © 2021 Rahila Parveen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we present a probabilistic-based method to predict malaria disease at an early stage. Malaria is a very dangerous disease that creates a lot of health problems. Therefore, there is a need for a system that helps us to recognize this disease at early stages through the visual symptoms and from the environmental data. In this paper, we proposed a Bayesian network (BN) model to predict the occurrences of malaria disease. The proposed BN model is built on different attributes of the patient's symptoms and environmental data which are divided into training and testing parts. Our proposed BN model when evaluated on the collected dataset found promising results with an accuracy of 81%. On the other hand, F1 score is also a good evaluation of these probabilistic models because there is a huge variation in class data. The complexity of these models is very high due to the increase of parent nodes in the given influence diagram, and the conditional probability table (CPT) also becomes more complex.

1. Introduction

The name of malaria is given to this disease in back 1740. Malaria is a genuine worldwide sickness and the main source of bleakness and deaths in tropical and subtropical nations. It influences somewhere in the range of 350 and 500 million persons and caused over a million casualties every year. However, malaria is both preventable and treatable. It is very much important for humans to treat themselves quickly when someone is affected by any disease at early stage. It is brought about by parasitic protozoa (a sort of unicellular microorganism) of the family *Plasmodium*. As per the World Health Organization (WHO) Report, [1] the most effected region from this disease in Sub-Saharan Africa and India is at least 85% of the whole world. The WHO shows that malaria cases are increased in the Bolivarian Republic of Venezuela as compared to Sub-Saharan Africa from 2010 to 2018.

This life-threatening disease is spread with the bite of female mosquitoes. These female mosquitos are responsible for spreading the *Plasmodium* parasite from one person to

another, and these for the most infect the humans in between twilight and dawning time bites. Due to these *Plasmodium* infection [2] patients, red blood cells are damaged, which leads to this disease. There are five major categories of parasites by which humans are getting infected with malaria, namely, *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, *Plasmodium ovale*, and *Plasmodium knowlesi* [3]. All of the deadliest parasites are *Plasmodium falciparum* in this disease of malaria. If malaria is detected at an early stage, then it can be remediable and escapable. The main technique for the detection of malaria parasites is the microscopic diagnosing method which is used until today as a gold technique. Slender or thick Giemsa recolored blood spreads are analyzed with an amplification factor of 100x goal and 10x visual focal point under a microscopic method [4]. It is derived from a blood smear screen report that any person might be influenced by more than one malaria parasite at the same time [5]. Besides, there are five unique types of malaria parasites and every species experiences an alternate life cycle. Each phase of the existence cycle

experiences an adjustment in its shape, size, morphology, shading, and so forth. These stages are named as a ring, trophozoite, schizont, and gametocyte [2].

Additionally, there exists an immediate connection between high episodes of malaria malady and the change in climate conditions. The high volume of cases in intestinal sickness scourge territories and deficiency of talented experts frequently prompts delay in revealing the outcome which might be basic for malaria treatment [6]. The deferral in finding forestall brief treatment of sickness further prompts constant stage, at considerable expense to the persons and enormous expense to society. This disease mostly exists in rural or remote areas of underdeveloped countries where technical staff did not skill so that in the current era of the Internet of Things (IoT) technologies for the collection of environmental data, advanced artificial intelligence (AI) can be utilized to detect malaria by analyzing the visual symptoms of infected persons. The visual symptoms of this disease are high temperature, feeling cold, headache, vomiting, and pain in body muscles or feeling too much tired. The doctor's reports regarding the mentioned symptoms can be monitored 24/7 in the affected area of respective countries for malaria disease. Due to this, less human interaction will be involved and automatically all these data are gathered at a central point. As these symptoms are detected in any person, the panel of doctors will recommend them for the clinical test of malaria at its early stage. Due to this, life of patients can be saved by monitoring his health condition not only for malaria but also blood pressure, sugar level, and heartbeat. In this process of gathering patient's data, privacy is also considered as the main part as per the recommended communication method for IoT by Jalbani et al. [7].

To validate the visual symptoms detected by the proposed system in this paper, the clinical test recommended by the doctors for malaria disease will be analyzed. As per the current study, it is suggested that probabilistic models are fit for learning information for the prediction of malaria disease. The Bayesian system (BS) is one such approach and has been broadly applied in healthcare systems because of their capacity to viably deal with variable data. A BN is a graphical portrayal of probability circulation where nodes speak to uncertain factors and connections speak to coordinate probabilistic impact between the uncertain values [8]. The connection between a node and its parents is evaluated by a CPT, determining the likelihood of the irregular variable precast on all assortments of the estimations of the parents. The structure of the system encodes data about probabilistic freedom with the end goal that the CPTs alongside the sovereignty relations give a full particular of the joint likelihood conveyance over the irregular factors spoke to by nodes. By decaying a joint likelihood dispersion into an assortment of littler nearby disseminations (the CPTs), a BN gives an exceptionally minimized interpretation of the total joint appropriation. In BN indicative models, factors that can impact the determination, including clinical signs, side effects, and lab results, are remembered for the model to frame a causal relationship organize [9]. The BN is one such class of probabilistic-based recognition approaches. It

exceeds expectations from clinical report characterization and impersonates the artificial neural network (ANN) system engineering such as human knowledge [10]. The IoT devices extracted data, and clinical tests will be further analyzed by the proposed method of AI, the BN. By the extracted symptom data and clinical reports, the information will be trained into the proposed framework for the accuracy of malaria disease recognition. On the bases of trained data, this system will decide whether the patient is infected with malaria or not. Because of this, the long process of microscopic tests will be reduced for remote areas and human life can be saved with this smart system. With this proposed model, not only will malaria disease be predicted but also other diseases can be monitored. It seems that machine learning (ML) approaches are used to build the different models using different approaches; however, every approach has its own advantages and disadvantages. Rule-based methods are based on informed search strategies which allow the system to construct direct classification connections while using probabilistic methods which are based on casual dependency relationships. Many classifiers such as ANN, support vector machines (SVM), decision tree, and BN have been used in medical diagnosis to get acceptable results; however, complexities of these algorithms are different which depend on data sets used for prediction. A BN seems to be more interpretable compared with other methods such as rule-based methods [11].

As discussed earlier that there are different ML techniques available to use in modeling the disease to recognize them properly, however, BN is very much helpful for predicting the malaria disease with different features such as environmental features. The BN has capabilities to integrate with different features to provide the prediction of the particular disease. The BN also shows the cause and effect relationship efficiently whereas decision tree (DT) is only work on yes and no properties and it can limit the results in certain range.

The main disadvantages of DT is overfitting and underfitting when using the small amount of data. Another main advantage of using BN is that it only connects nodes that are probabilistically linked by some sort of causal dependency, which helps to reduce the computational cost of the system. A second main reason is that the BN nets are very much flexible; therefore, they are so adjustable [12].

The paper is divided into the following sections. In Section 2, the related work is given for this area of research. In Section 3, the proposed methodology is described for the Bayesian model to predict the malaria disease. In Section 4, the results and discussions are explained for the proposed model. In Section 5, the final words of conclusions are given regarding this research paper.

2. Related Work

Vijayalakshmi [13] has proposed a new visual geometry group-support vector machine (VGG-SVM) organizes utilizing move learning approach for perceiving tainted falciparum intestinal sickness. Here, a preprepared VGG is considered as a specialist learning model and it is focused to

characterize 1000 classes. SVM is an area explicit classifier used to group tainted and noncontaminated from malaria microscopic images.

Parveen et al. [14] have utilized neural networks (NNs) by mental undertaking and simulation of the brain. A multilayer feedforward system with a back proliferation learning algorithm is utilized. The effectiveness of the proposed framework is contrasted with other comparative frameworks.

Dong et al. [15] have used a completely automated process with no manual element extraction, and we picked a profound convolutional neural system (CNN) as the classifier. CNN can separate various leveled portrayals of the input information. They have used LeNet-5 to get proficiency with the intrinsic highlights of intestinal malaria contaminated and nontainted cells.

Bari et al. [16] have used identification, and analysis of lung cancer growth can be processed as images on three fundamental stages which are prehandling, division lastly followed by postpreparing. Any sort of diagnosing technique is expected to gather the blood tests to recognize malaria. There are two distinctive blood films are utilized for distinguishing proof of malaria in particular, thick and thin blood films by Elter et al. [17]. Thick blood films acknowledge the blood tests for the location of intestinal malaria parasite thickness. Slender blood films acknowledge the blood tests to recognize or portrayal of malaria parasites.

Maqsood et al. [18] have proposed identifying dark scale images utilizing hidden Markov tree (HMT) which is introduced. Utilizing the above-displaying structure, a joined spatial and fleeting separating procedure can expel Gaussian just as dot commotion from shading images and video successions significantly. By molesting the conditions among wavelet coefficients, better execution has been accomplished.

Cooper et al. [19] have developed a BN with 20 million hubs for *Bacillus anthracenes* outburst identification. The whole populace is displayed by a self-contained system with every individual in the system associated with the remainder of the system through a hub called sickness status. The undertaking is to compute the back prospect of the alarm hub given the construed illness statuses surprisingly in the public. The work does not think about spatial or sophisticated viewpoints.

Jiang and Wallstrom [20] have explored the utilization of BS for cryptosporidium occurrence place and projection. This examination centers essentially around the weaknesses of traditional techniques for time-arrangement investigation and improving them by utilizing BS. The model is tried on a recreated occurrence informational collection.

In Martha et al.'s study [21], an adaptable smart framework has been developed by using fuzzy logic in diagnosing malaria sickness. The framework recognized malaria sickness with high identification exactness. In any case, the framework had gone with issues: the framework intelligent module could not make a bidirectional conclusion and handle issues of impossibility.

In Mehanian et al.'s study [22], a smart framework that investigated malaria using CNN was developed. The

framework decided to *Plasmodium falciparum* malaria sickness to have high acknowledgment exactness. In any case, the framework had the going with insufficiencies to be explicit: the arrangement gotten from the NN is difficult to get a handle on and the learning method of the framework is tedious and capital-escalated, and the framework fails to perceive cerebral malaria and other mosquitoborne disorders due to the covering indications the sickness impacts to neurological and febrile infections.

Junior et al. [23] have assessed the utilization of ANN and BN with regards to the determination of asymptomatic malaria disease. They have figured out how to assemble ANNs and BNs dependent on immunological and epidemiological information gathered from people from an exceptionally endemic area for malaria in the Brazilian Amazon. The results were contrasted with those acquired utilizing light microscopy and a subatomic test (settled PCR). The Naïve-Bayesian has utilized as a probabilistic learning technique, and these classifiers are among the best-known algorithms for figuring out how to arrange text records such as e-mail spam separating [24]. Some exploration shows that it is likewise helpful for heart illness forecast.

Osubor and Chiemeké [25] have proposed a solution of the adaptive neurofuzzy inference system (ANFIS) that examined malaria disease was made. The framework decided malaria sickness with 98% disclosure precision. Notwithstanding the high acknowledgment exactness, the framework had the going with quandaries, for instance, inconvenience in recognizing cerebral malaria sickness and other mosquitoborne infirmities as a result of the covering reactions and the contamination grants to neurological and febrile maladies.

Djam et al. [26] have an expert framework called the fuzzy expert system for the management of malaria (FESMM) which relied upon fluffy principles to investigate malaria sickness called was made. The framework decided malaria to have high disclosure accuracy. Despite the framework's high distinguishing proof accuracy, it had the going with detriments: the framework could not recognize cerebral malaria and other mosquitoborne sicknesses.

Jiang and Wallstrom [20] have analyzed the utilization of BNs for cryptosporidium spreading location and forecast. This examination concentrates predominantly on the weaknesses of old-style strategies for time-arrangement investigation and improving them by utilizing BNs. The model is tried on a reproduced spreading informational index.

Sebastiani et al. [27] have developed an automatic BN for anticipating flu-like disease and the quantity of pneumonia and flu passing's dependent on past pediatric and grown-up instances of respiratory disorder. No natural factors are utilized.

Warfield [28] has proposed interpretive structural modeling (ISM) which was utilized to build up a visual progressive structure of complex frameworks. The method was utilized in overseeing decision making for complex issues. The contribution of the ISM strategy was unstructured, and indistinct data about the framework factors and

their interdependencies were utilized. The yield of ISM investigation was a very much characterized, ordered, and instructive model, which is valuable for some different purposes.

Colin et al. [29] have executed ISM to consider the interdependencies in gracefully chain hazards. The inter-relationship among flexibly hazard factors, which they created, depended on the reliance and driving intensity of separate components. Their investigation is very much important for future researchers in this area. They contemplated 21 danger factors, and beginning interrelationships were created utilizing bunch conversations among the creators and individual specialists.

Singh and Kant [30] have organized nine boundaries in knowledge management (KM) in the business procedure. The KM hindrances were those which antagonistically influence the execution of KM in a business association. They executed the ISM strategy for their investigations. The shared connections among various boundaries were grouped in various levels and incorporated a visual relationship also. Be that as it may, the examination was deficient in clarifying how frail or solid relations were among different interconnected elements.

Hussain et al. [31] have proposed a novel methodology for identification of hand developments during impediment, and they reported consequences of isolating hand (s) from the head (face) locale during explicit signal advancement, utilizing shading and surface invariant recognition plot. Moreover, broadening their work for the improvement of a keen mentoring framework that requires distinguishing motions and foreseeing the understudy's emotional (mental) state progressively, they proposed and assessed a coordinated methodology of identification and understanding of mental states from hand signals utilizing information detailed before by Abbasi et al. [32]. The outcomes are promising in further turn of events.

The BN has played a vital role in healthcare systems. BN has got great attention in all over the world and maintains to obtain momentous importance for the facility to mingle the available evidence and perfect reasoning under uncertain situation by McLachlan et al. [33].

3. Proposed Methodology

The data have been collected from various sources regarding malaria-affected peoples. For the collection of information regarding the symptoms of malaria disease, different types of IoT devices have been utilized. For the environmental information, the research target was tropical and nontropical area weather conditions. Due to these, more effected areas from malaria can be identified on the basis of weather conditions changed, for example, in rain, weather humidity will be increased. In this condition, the number of malaria cases may increase due to the growth of mosquitoes in these conditions. This all gathered information will be processed in BN for getting perfect results of probability regarding malaria disease. The proposed methodology is described in Figure 1.

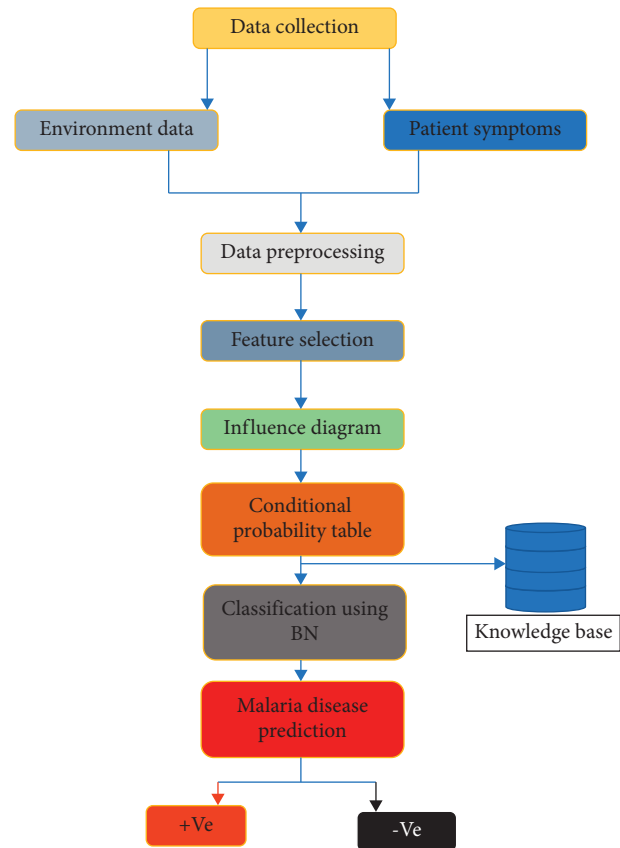


FIGURE 1: A proposed framework for malaria disease prediction.

3.1. Visual Symptoms of Malaria Patients. The symptoms of malaria disease may be visible from 8–25 days in infected peoples from this parasite. However, at the same time, these symptoms can be visible for some days later as mentioned above if those patients have used antibiotic medicine already in proactive measures from malaria disease. These symptoms can vary for each type of malaria disease parasite infection. Likewise, for the *Plasmodium vivax* and *Plasmodium ovale*, the patients feel skin trembling with high temperature and wilting after every two-day cycle. On the other hand, for *Plasmodium malariae* and *Plasmodium falciparum*, parasite-infected patients feel these symptoms after every three-day cycle in which high temperature may remain also for the 36–48 hours. And other main categories of malaria symptoms are headache along with high temperature and feeling too much fatigued. The doctors working on malaria recommend that at least two types of symptoms will be visible simultaneously in the patients infected with parasites. Alongside these four main types of symptoms as mentioned earlier, there are four other low-level types of symptoms also observed in malaria patients, such as spewing, amplification in irritation, cough, and pain in body muscles. In all of the parasite infections, plasmodium falciparum is more dangerous and life-threatening for infected patients. In this type of infection, the patients' symptoms can be visible after 9–30 days. The headache is the more prominent type of symptoms in this infection. This will help to increase the chance of more accurate prediction of malaria disease.

3.2. Environmental Data Collection. The environmental effects are also major sources for spreading the malaria infection in tropical and nontropical areas around the globe. The major carrier of these malaria parasites is female mosquitoes, and these types of environments can be too friendly for their growth. The environment conditions are also considered in this research paper as a data component for the prediction of malaria. The different sources have been utilized for the collection of environmental data the online resources such as weather.com and local metrological department data are the main contributor to this research, such as maximum temperature, minimum average temperature, humidity, rainfall months, and how many malaria cases are reported during that period. The humidity will increase during the rainfall seasons, and that is the period in which mosquitoes will grow in more numbers. After that these cases will be categorized as per the type of parasite infections. This will help near about 85% accurate prediction of malaria disease.

3.3. Influence Diagram. Howard [34] has developed an influence diagram which is known as relevance diagrams. These diagrams are based on acyclic directed graphs for the solution of decision issues. The motive behind this influence diagrams was to get a higher rate profit by selecting alternative decisions from given values. The influence diagrams are the same as BNs for the complete overview of domain architecture such as the structure of the decision problems. The influence diagrams are further divided into four nodes which are decision, chance, deterministic, and values and two kinds of arcs which are influences and informational. Typically, an arc in an influence outline indicates an influence, such as the way that the node at the tail of the circular segment impacts the worth or the probability dissemination over the potential estimations of the node at the top of the arc. A few arcs in influence graphs have causal importance. Specifically, a guided way from a choice node to a possibility node implies that the choice (for example, control of the diagram) will affect that opportunity node in the feeling of changing the probability appropriation over its results.

In Figure 2, the sample influence diagram is shown and can be obtained from the factorized form and where directed edges represent direct dependencies and the absence of edges shows the conditional independence. These types of the model are also known as belief network (BN) or graphical models and causal network as well. BN or BN is a directed acyclic graph represented with pairs $G(V, E)$. V stands for vertices or nodes which represents random variables (events) in our case symptoms and environmental variables, and E stands for edges or links between nodes which shows a causal dependency relationship. A direct link from variable X - Z indicates that X can cause Z , or, in BN terminology, X is a parent of Z , and Z is a child of X . P is a probability distribution over vertices V . Discrete random variables are assigned to the node variables representing a finite set of mutually exclusive states and interpreted with a CPT that represents the conditional probability of the

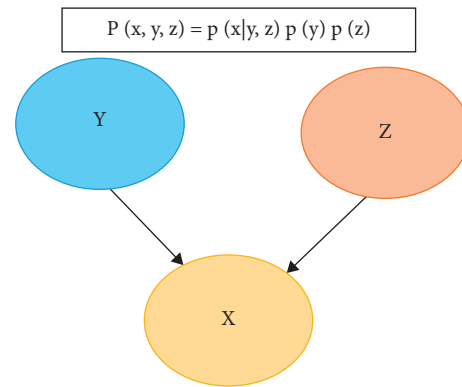


FIGURE 2: Sample influence diagram.

variable given the values of its parents in the given graph. In our proposed BN model, nodes are symptoms of malaria disease and environmental variables.

3.4. Conditional Probability Tables. It is a common observation that the effects of direct or indirect factors on the number of events are very difficult to model. However, BN is considered a very easy and powerful framework for graphical modeling of cause-effect relationships between different variables using influence diagrams. The sample CPT is shown in Figure 3.

The BN is divided into two parts: one is the qualitative part and the other is the quantitative part. The qualitative part is based on structural modeling and learning, which is used to make the structure of a directed acyclic graph (DAG) which is made of a set of vertices (nodes) and directed arrows (edges) which show the parent-child relationship between nodes. When two nodes are connected directly by an edge, the node with a link directed to a successive node is called a “parent node” of the succeeding node. The subsequent node is called a “child node.” Child nodes are also known as conditionally dependent nodes on their parent nodes. For instance, nodes Z and Y are the parent nodes of node X (Figure 2). The quantitative part of a BN is also known as parameter learning, which is used to find out the conditional probability distribution of each node, according to the established BN structure of the proposed scenario which is based on Bayes’ theorem. The same scenario has been used in our proposed BN prediction model where symptoms are given to the BN model and the CPT has been calculated using Bayes theorem.

4. Results and Discussion

For the prediction of malaria disease based on environmental information and the patient’s symptoms by using the probabilistic model, for a better prediction with less time, the BN model is used in this paper. If the patient is not diagnosed at the early stage of this parasite virus, then this will be life-threatening for them. The malaria disease exists in the underdeveloped countries where a shortage of experts as a laborer’s technician to diagnose it. As technology evolved such as IoT networks and AI in the field of healthcare systems, by

Variable Y		Variable Z	
Yes	0.4	Yes	0.2
No	0.6	No	0.8

Variable X			
Y	Z	Yes	No
Yes	Yes	0.55	0.45
Yes	No	0.4	0.6
No	Yes	0.55	0.45
No	No	0.9	0.1

FIGURE 3: Sample CPT.

utilizing these two technologies, the malaria disease can be predicted at an early stage around 80–85%. The current research has been done in the diagnosis methods which are processed by microscopic tests. There are chances that the medical staff can be infected from this virus. The proposed model in the research paper will predict malaria disease remotely with the help of IoT devices and sensors. The environmental information has been collected from various sources regarding weather conditions of those tropical and nontropical areas where the malaria parasites exist. This disease spreads due to the bites of female mosquitoes with the infection of parasite virus so that there is also the season of rising these mosquitoes in different weather conditions. The humans may be infected more during more rain falling and humid conditions. Or they may be at the start of winter or the end of winter. However, the study suggests that this infection rate increases during the few last months of summer because in this period the mosquitoes are rising too much due to heavy rain falls and humidity. In this paper, at least 16 types of environmental and patient's symptoms have been considered for the probabilistic.

For data description, two types of data are collected from the patient's symptoms and environmental data. The patient's symptoms are collected from the Peoples Medical Hospital (PMH) Shaheed Benazirabad, Nawabshah, Pakistan, regarding infected patients. And the environmental data are collected from online resources. The collected data are shown in Table 1.

4.1. Bayesian Network Model. The BN architecture is a graphical representation, which demonstrates it as a qualitative base for the communication between variable sets within a model. The architecture of the directed graph could be cloned into the fundamental construction for the designed domain, or it may not be necessary. At the point when the structure is causal, it gives a helpful, measured knowledge into the associations among the factors and considers the prediction of impacts of outside control. In light of the restrictive conditions, a BN factorizes the joint dissemination of factors. The BN registers the circulation probabilities in a given arrangement of factors by utilizing earlier data of different factors by Jensen [8]. The arrangement of nodes and coordinated circular segments are the attributes of a BN, where nodes speak to the framework factors and the bend speaks to

TABLE 1: Data description table.

Description	Value
Total record	250
Training	150
Testing	100
<i>Patient's symptoms</i>	
Fever	
Chills	
Sweats	
Shaking	
Jaundice	
Headache	
Fatigue	13
Low energy	
Nausea	
Vomiting	
Myalgia	
Stomach upset	
Diarrhea	
<i>Environmental data</i>	
Rainfall	
Humidity	03
Temperature	

the reason impact relationship of conditions among the factors. Every node has its likelihood of an event. On account of a root node, such likelihood is from the earlier one and is resolved for the others by surmising. The nodes which are not coordinated towards some other nodes are the parent nodes. A youngster node is a node where a node gets any edge/coordinated curves. The probabilities of parent nodes and a restrictive likelihood (CPT) were the bases for BN calculations. The model has been created using GeNIe/SMILE [35]. The CPT contained the data for contingent probabilities. Figure 4 shows the sample BN model.

BS is noncyclic coordinated diagrams that represent to factorizations of joint probability appropriations. Each joint probability dissemination over n arbitrary factors can be factorized in n times and composed as a result of probability disseminations of every one of the factors contingent on different factors. The basic properties of Bayesian networks, alongside the CPT related to their nodes, take into account probabilistic thinking inside the model. Probabilistic thinking inside a BN is prompted by watching proof. A node that has been watched is called a proof node. Watched nodes become launched, which implies, in the least difficult case, that their result is known with sureness. The effect of the proof can be spread through the system, changing the probability conveyance of different hubs that are probabilistically identified with the proof.

4.2. BN Model Trained Data. This cycle sums at the establishments to a dull use of Bayes's hypothesis to update the probability appropriations of all nodes in the system. Various methods of applying Bayes' hypothesis and distinctive requests for updating lead to various calculations. The current calculations for thinking in BS can be partitioned into three gatherings: message passing, chart decrease, and

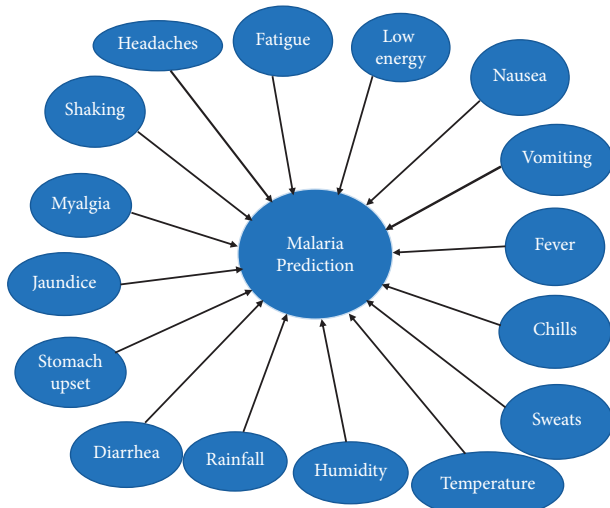


FIGURE 4: Proposed BN model.

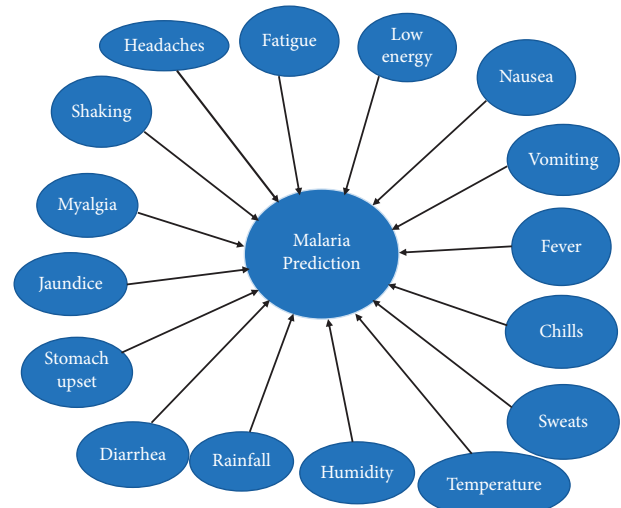


FIGURE 5: Proposed trained BN model.

stochastic reenactment. The express portrayal of freedoms takes into account the expanded computational manageability of probabilistic thinking. Probabilistic surmising in separately associated BN is proficient. Sadly, careful calculations for duplicate associated systems are obligated to exponential multifaceted nature in the number of nodes in the system. The environmental information and patient's symptoms are trained in the BN model as shown in Figure 5.

4.3. Testing of Proposed BN Model. To test the samples, the proposed BN model has been used to predict the occurrence of malaria in a particular patient. Figure 6 shows the result of the proposed model applied to the data of a patient. Different symptoms and environmental data have been given to the model, and the model reported the accuracy 81%. 100 samples are tested on the proposed BN model.

Figure 7 shows the negative test result of a patient.

4.3.1. Conditional Probability Table Generation. In the probabilistic domain, the graphs or diagrams will show a much more detailed view of information regarding architecture but not about the numerical informational. Those values are encrypted into conditional probabilistic delivery matrices which are equal to the factorization form and these are known as CPT connected with the nodes. Significantly, there will consistently be nodes in the system without any forerunners. These nodes are described by their earlier negligible probability dissemination. Any probability in the joint probability dispersion can be resolved from these expressly spoken of earlier and restrictive probabilities. After data collection, there is a need of preprocessing those data for training and testing of the BN model with respect to CPT for the prediction of malaria disease. And after that, this information will be preprocessed into the BN model along with an influence diagram and CPT for the prediction of malaria disease. With this, malaria disease can be predicted at an early stage, and due to this, the life of the patient can be saved.

4.4. Proposed BN Model Evaluation. A confusion matrix is a matrix that can be used to measure the performance of the classifier in supervised ML methods. It is used to evaluate the performance of the classification model on the given test data set whereas truth values are given. The confusion matrix has been used to visualize the accuracy of a classifier by comparing the actual data and predicted data classes. In this research work, 100 patient's records are used to test the performance of the proposed BN model. Our model out of 100 records correctly classified 81 patients as malaria positive and 19 correctly classified as negative. Therefore, the overall accuracy of the classifier is 81% shown in Table 2. After accuracy measurements, the F1 score is also very important to understand the performance of the classifier which shows the weighted average of Precision and Recall. In some problems, the F1 score is considered more useful than accuracy when there is an uneven class distribution of the features.

The statistical significant test is used to measure the probability of different observed relationships of the data and provide the important information regarding the proposed research findings. However, statistical significance may report confusing results due to less numbers and does not have any relationship with practical findings of the particular research. Nonstatistical significance does not mean that there is no significant difference between the different groups or shows no effect of the findings by Amrhein et al. [36].

4.5. Computational Complexity of Probabilistic BN Model. It has been remembered that the complexity of probabilistic models comes in the category of NP-hard in worst-case scenarios because the complexity of these probabilistic models is based on the exponential growth of the CPT in the number of parents and structural modeling problem of connectivity of DAGs. If the number of nodes increased, CPT becomes more complex; for example, if there are 10 nodes, then 2048 parameters are there, so care must be taken in mind when using probabilistic inference.

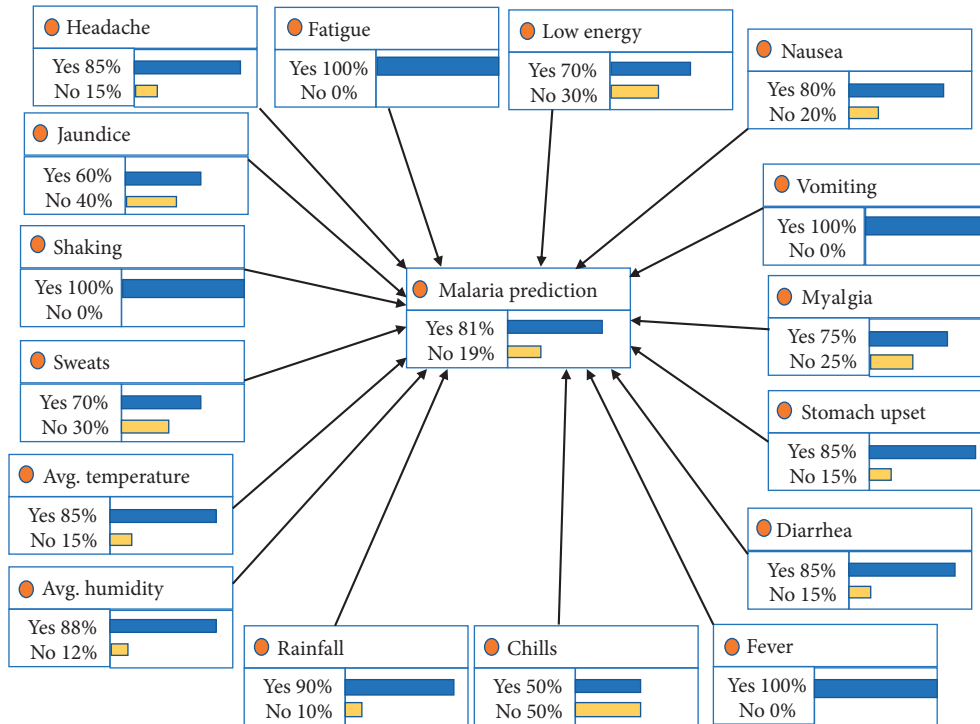


FIGURE 6: Proposed BN model predicted positive malaria disease.

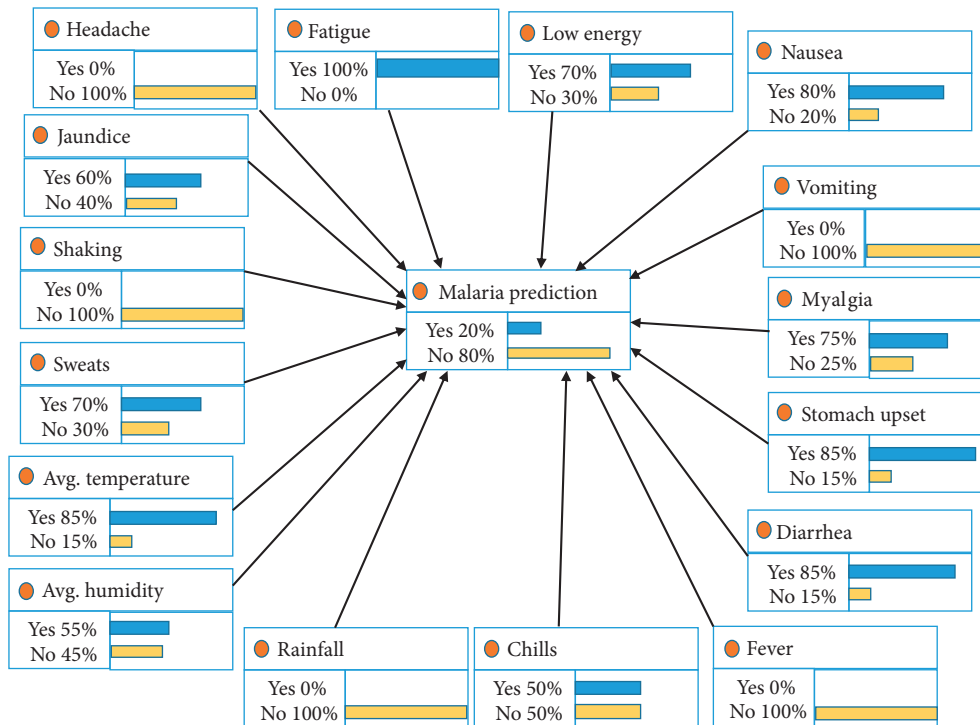


FIGURE 7: Proposed BN model predicted negative malaria disease.

TABLE 2: Confusion matrix.

Actual	Predicted		No
	Yes	No	
	Yes	46	14
	No	5	35
Measure	Value	Derivations	
Sensitivity	0.9020	TPR = TP/(TP + FN)	
Specificity	0.7143	SPC = TN/(FP + TN)	
Precision	0.7667	PPV = TP/(TP + FP)	
Negative predictive value	0.8750	NPV = TN/(TN + FN)	
False positive rate	0.2857	FPR = FP/(FP + TN)	
False discovery rate	0.2333	FDR = FP/(FP + TP)	
False negative rate	0.0980	FNR = FN/(FN + TP)	
Accuracy	0.8100	ACC = (TP + TN)/(P + N)	
F1 score	0.8288	F1 = 2TP/(2TP + FP + FN)	
Mathews correlation coefficient	0.6288	$\frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$	

5. Conclusion

Malaria is a life-threatening disease to the patients if it is not detected timely. For quick detection of this disease and without infecting medical staff from this virus, in this research paper, the AI method of BN is used for the prediction of malaria disease. The two kinds of parameters, the environmental information and patient's symptoms, are considered to predict it. In this process, the environmental information is collected from various types of online resources and local metrological departments of the respective areas regarding weather condition predictions. For this, the weather conditions are collected from tropical and nontropical countries of the world because, as per old or current research, these countries are still more affected by malaria. However, in these countries, the mosquitoes with the virus of parasites are at rising in the season especially in summer and at the start or end of winter. The weather conditions are included in the proposed system which are rain falling, humidity rate, and weather is hot or cold, taken as input into the proposed framework. In this paper, we have used the BN model to predict malaria disease using different symptoms and environmental data. We used BN-based inference from the GeNIe/SMILE tool to train and test our proposed framework. We built our own BN model and trained on our data set.

The proposed model could correctly predict malaria as positive or negative. There are 13 different symptoms of malaria, and 3 environmental factors are used in our proposed system. In this paper, we present a framework based on the BN model for the prediction of malaria disease which provides 81% accuracy of results on a given data set.

The main contribution of this paper is to use clinical diagnosis of symptoms of any disease is based on visual features that showed by patients. This method seems to be useful where there is no any urgent facility of testing the blood samples are available at early stages of the disease, and this may be helpful to control this disease to prevent the human life. In this research paper, additional feature of environmental conditions is also considered which may help to diagnose the patients properly.

The BN inference model is considered an NP-hard problem due to the increasing number of parent nodes in the influence diagram. The CPT also becomes complex which creates a problem for the machine. Therefore, care must be taken into account when selecting the nodes for designing any probabilistic model.

5.1. Future Work. In future work, there may be a possibility to increase the data set to get a more accurate result. Even there is a huge possibility of using a different variation of the BN model for the prediction of malaria and many other diseases also. For the collection of data regarding patient's symptoms and environmental information, the IoT devices and sensors will play a vital role. Due to the combination of AI and IoT technologies, the prediction of malaria or other diseases can be detected at an early stage.

Data Availability

Two types of data are collected from the patient's symptoms and environmental data. The patient's symptoms are collected from the hospitals regarding infected patients. And the environmental data are collected from online resources.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the International Cooperation Project, Department of Science & Technology, Henan Province, China, under grant no. 172102410065; Basic Research Project of the Education Department of Henan Province under grant no. 17A520057; Frontier Interdisciplinary Project of Zhengzhou University under grant no. XKZDQY202010; and the ZZU-Kingduns Joint Laboratory of Cyber Security.

References

- [1] World Health Organization, *The World Malaria Report 2019 at a Glance*, World Health Organization, Geneva, Switzerland, 2019, <https://www.who.int/news-room/feature-stories/detail/world-malaria-report-2019>.
- [2] Z. May, S. S. A. M. Aziz, and R. Salamat, "Automated quantification and classification of malaria parasites in thin blood smears," in *Proceedings of the IEEE International Conference on Signal and Image Processing Applications*, pp. 369–373, Melaka, Malaysia, October 2013.
- [3] P. Khanna and S. Kumar, "Malaria Parasite Classification Employing Chan–Vese Algorithm and SVM for Healthcare," in *Proceedings of the 1st International Conference on Computing, Communications, and Cyber-Security*, pp. 697–711, Springer, New York; NY; United States, November 2020.
- [4] S. N. Chavan and A. M. Sutkar, "Malaria disease identification and analysis using image processing," *International Journal Latest Trends Engineering Technology*, vol. 3, pp. 218–223, 2014.
- [5] World Malaria Report, 2020, <https://www.who.int/malaria/publications/world-malaria-report-2017/en/>.
- [6] S. Rajaraman, S. K. Antani, M. Poostchi et al., "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images," *Peer Journal*, vol. 6, Article ID e4568, 2018.
- [7] K. B. Jalbani, A. H. Jalbani, and S. S. Soomro, "IoT Security: To Secure IoT Devices with Two-Factor Authentication by Using a Secure Protocol," *Industrial Internet of Things and Cyber-Physical Systems: Transforming the Conventional to Digital*, pp. 98–118, 2020.
- [8] F. V. Jensen, *An Introduction to Bayesian Networks*, vol. 210, pp. 1–178, UCL Press, London, UK, 1996.
- [9] K. B. Korb and A. E. Nicholson, *Bayesian Artificial Intelligence*, CRC Press, Boca Raton, Florida, USA, 2010.
- [10] V. V. Makkapati and R. M. Rao, "Ontology-based malaria parasite stage and species identification from peripheral blood smear images," in *Proceedings of the IEEE Annual International Conference on Engineering in Medicine and Biology Society*, pp. 6138–6141, Boston, Massachusetts, August 2011.
- [11] N. Shafaf and H. Malek, "Applications of machine learning approaches in emergency medicine," *Archives of Academic Emergency Medicine*, vol. 7, no. 1, 2019.
- [12] V. Bhavana and T. Adilakshmi, "Comparison of decision tree classifier and Bayes classifier using WEKA," *International Journal of Computer Applications*, vol. 176, no. 3, 2017.
- [13] A. Vijayalakshmi, "Deep learning approach to detect malaria from microscopic images," *Multimedia Tools and Applications*, vol. 79, no. 21, pp. 15297–15317, 2020.
- [14] R. Parveen, A. H. Jalbani, M. Shaikh et al., "Prediction of malaria using artificial neural network," *IJCSNS*, vol. 17, no. 12, p. 79, 2017.
- [15] Y. Dong, Z. Jiang, H. Shen, and W. D. Pan, "Classification accuracies of malaria-infected cells using deep convolutional neural networks based on decompressed images," in *Proceedings of the IEEE Southeast Conference*, pp. 1–6, San Jose California, March 2017.
- [16] M. Bari, A. Ahmed, M. Sabir, and S. Naveed, "Lung Cancer Detection Using Digital Image Processing Techniques: A Review," *Mehran University Research Journal of Engineering & Technology*, Jamshoro, vol. 38, no. 2, p. 351, Pakistan, 2019.
- [17] M. Elter, E. Haßmeyer, and T. Zerfaß, "Detection of malaria parasites in thick blood films," in *Proceedings of the IEEE Annual International Conference on Engineering in Medicine and Biology Society*, pp. 5140–5144, Boston, Massachusetts, August 2011.
- [18] A. Maqsood, I. Touqir, A. M. Siddiqui, and M. Haider, *Wavelet-Based Video Denoising Using Probabilistic Models*, *Mehran University Research Journal of Engineering & Technology*, Jamshoro, vol. 38, no. 1, p. 17, Pakistan, 2019.
- [19] G. F. Cooper, D. Dash, J. Levander, W. K. Wong, W. Hogan, and M. Wagner, "Bayesian Biosurveillance of Disease Outbreaks," 2012, https://www.dbmi.pitt.edu/sites/default/files/Cooper_13.pdf.
- [20] X. Jiang and G. L. Wallstrom, "A bayesian network for outbreak detection and prediction," *Proceedings of National Conference on Artificial Intelligence*, AAAI Press, MIT Press, vol. 21, no. 2, p. 1155, Menlo Park, CA, USA, 2006.
- [21] S. Martha, N. Emilian, A. Frederick et al., "Deep Learning Mobile Application towards Malaria Diagnosis," in *Proceedings of the 9th International Conference on Learning Representations*, pp. 1–6, Vienna, Austria, May 2020.
- [22] C. Mehanian, M. Jaiswal, C. Delahunt et al., "Computer-automated malaria diagnosis and quantitation using convolutional neural networks," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 116–125, Venice, Italy, October 2017.
- [23] A. M. B. Junior, A. A. Duarte, M. B. Netto, and B. B. Andrade, "Artificial neural networks and bayesian networks as supporting tools for diagnosis of asymptomatic malaria," in *Proceedings of the IEEE 12th International Conference on E-Health Networking, Applications and Services*, pp. 106–111, Lyon, France, July 2010.
- [24] Naive-Bayes Classification Algorithm, <http://software.ucv.ro/~cmihaescu/ro/teaching/AIR/docs/Lab4-NaiveBayes.pdf>, 2020.
- [25] V. I. Osubor and S. C. Chiemeké, "An adaptive neuro-fuzzy inference system for the diagnosis of malaria," *Nigerian Society for Experimental Biology Journal*, vol. 14, no. 4, 2017.
- [26] X. Y. Djam, G. M. Wajiga, Y. H. Kimbi, and N. V. Blamah, "A fuzzy expert system for the management of malaria," *International Journal of Pure and Applied Sciences and Technology*, vol. 5, no. 2, pp. 84–108, 2011.
- [27] P. Sebastiani, K. D. Mandl, P. Szolovits, I. S. Kohane, and M. F. Ramoni, "A bayesian dynamic model for influenza surveillance," *Statistics in Medicine*, vol. 25, no. 11, pp. 1803–1816, 2006.
- [28] J. N. Warfield, "Developing interconnection matrices in structural modeling," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-4, no. 1, pp. 81–87, 1974.
- [29] J. Colin, D. Estampe, H. C. Pfohl, P. Gallus, and D. Thomas, "Interpretive structural modeling of supply chain risks," *International Journal of Physical Distribution & Logistics Management*, vol. 41, no. 9, pp. 839–859, 2011.
- [30] M. D. Singh and R. Kant, "Knowledge management barriers: an interpretive structural modeling approach," *International Journal of Management Science & Engineering Management*, vol. 3, no. 2, pp. 141–150, 2008.
- [31] A. Hussain, A. R. Abbasi, and N. Afzulpurkar, "Detecting & Interpreting Self-Manipulating Hand Movements for Student's Affect Prediction," *Human-Centric Computing and Information Sciences*, vol. 2, no. 1, p. 14, 2012.
- [32] A. R. Abbasi, M. N. Dailey, N. V. Afzulpurkar, and T. Uno, "Student mental state inference from unintentional body gestures using dynamic bayesian networks," *Journal on Multimodal User Interfaces*, vol. 3, no. 1, pp. 21–31, 2010.
- [33] S. McLachlan, K. Dube, G. A. Hitman, N. Fenton, and E. Kyrimi, "Bayesian Networks in Healthcare: Distribution by

- Medical Condition,” *Artificial Intelligence in Medicine*, vol. 7, 2020.
- [34] R. A. Howard, “Readings on the Principles and Applications of Decision Analysis,” *Professional Collection Strategic Decisions Group*, vol. 2, 1984.
- [35] *GeNIe Modeler or SMILE Engine*, <https://www.bayesfusion.com/>, 2020.
- [36] V. Amrhein, S. Greenland, and B. McShane, “Scientists rise up against statistical significance,” *Nature*, vol. 567, no. 7748, pp. 305–307, 2019.