

# A Patient-doctor portal for Malaria parasite prediction

Mrs.ALEESHA MARIA SIBY  
*Dept of CSE*  
 MBC CET  
 Kerala, India  
 aleeshamariasibi@gmail.com

Mrs.SHIVAPRIYA B  
*Dept of CSE*  
 MBC CET  
 Kerala, India  
 sivapriya4570@gmail.com

Mrs.THABITHA IGNATIOUS  
 FERNANDEZ  
*Dept of CSE*  
 MBC CET  
 Kerala, India  
 fernandezthabitha4597@gmail.com

Mrs. TINA ELSA VARGHESE  
*Dept of CSE*  
 MBC CET  
 Kerala, India  
 tinaelsa97@gmail.com

Ms. JYOLSNA MARY P  
*Asst. Prof of CSE*  
 MBC CET  
 Kerala, India  
 jyolsnamary@gmail.com

**Abstract**—Malaria is one of the significant general medical issues in India. Early expectation of a Malaria is the key for control of disease grimness, mortality just as decreasing the danger of transmission of disease in the network and can help policymakers, well-being suppliers, restorative officers, service of well-being and other well-being associations to all the more likely target therapeutic assets to regions of most prominent need. In this work, we have designed doctor-patient information portal with several parameters. Machine learning algorithms like random forest and support vector machine are used to train the data. Information of has been gathered from various resources. Parameters utilized are cold, rigor, fatigue, headache, bitter\_tounge and etc. A substantial quantities of tests were gathered from various sources. It is seen that execution of the model created utilizing Random Forest is more precise than Support Vector Machine(SVM). The Random Forest model can foresee the episode 15 - 20 days ahead of time. Anyway precision of forecast can be expanded utilizing additionally preparing information. This model can be scaled-up at nation level.

**Index Terms**—Malaria, Machine learning, Random forest

## 1 INTRODUCTION

Ailments brought about by the irresistible small scale living being which is transmitted to individuals by methods for parasitic arthropods are known as vector-borne ailment. Among a few arthropods, mosquitoes are the most widely recognized creepy crawlies which are in charge of causing a few sicknesses like malaria, dengue, and so on. Among the vector conceived infections, intestinal sickness is an extraordinary reason for dread for human wellbeing. Particularly in the creating nations like African nations and India, a large number of individuals experience the ill effects of this malady consistently. Regardless of numerous investigations on the malaria [6 - 8] still there were 243 million intestinal sickness cases detailed in 2008[9]. Intestinal sickness is still a noteworthy general medical issue as 109 nations are announced endemic to the ailment in 2008. Government specialists are causing gigantic expense to control/wipe out the episodes of jungle fever.

Because of a dangerous atmospheric deviation, fast atmosphere changes are happening which result in the expansion or lessening of disease transmission bases on the particular small scale atmosphere of that specific district. Temperature variances influence the existence cycle of vector just as parasite[10]. Further, the procedure engaged with disease transmission is dynamic and is totally site explicit. In this manner, it is incredible challenge for the scientists to anticipate the malarial episodes in advance. Without

information about probabilistic assault of these sicknesses, government neglects to give sufficient treatment office on schedule. In this manner, it is important to figure the event of these sicknesses ahead of time with the goal that its staggering effect on the society can be decreased.

We designed a doctor-patient interacting portal in order to predict the malaria disease based on various symptoms like cold, headache, fatigue and etc. Doctor and patient can login into the portal and enter the details of symptoms thereby depicting the disease. Based on the trained data, machine can be able to predict whether the particular patient is suffering with the disease or not.

In section 2, we presented related work. Various studies exist for specific areas which suffer with malaria. Section 3, we explained the dataset and parameters that we have gathered for training the machine. Section 4 gives a clear cut idea about the machine learning algorithms that have been considered for processing data and predicting the disease. Section 5 discussed about the results followed by conclusion.

## 2 EXISTING METHODS

Vijeta Sharma[1] et.al proposed a malaria outbreak prediction algorithm based on SVM and Artificial neural network(ANN) and shown that SVM outperforms with the given training data. They

have collected data from different areas of Maharashtra with parameters such as average monthly rainfall, humidity, total number of positive cases, total number of pF cases and etc.

Contemporary sickness mapping endeavors have grasped measurable demonstrating ways to deal with appropriately recognize vulnerabilities in both the accessible estimations and their spatial introduction. The most widely recognized such approach is Gaussian procedure relapse, a numerical structure com- presented of two segments: a mean capacity tackling the prescient intensity of different free factors, and a covariance work yielding spatio-worldly shrinkage against lingering variety from the mean. Here, Samir Bhatt[2] et. al present a troupe approach dependent on stacked generalization that takes into account different nonlinear algorithmic mean capacities to be mutually implanted inside the Gaussian procedure structure. They apply this technique to mapping Plasmodium falciparum pervasiveness information in sub-Saharan Africa and demonstrate that the summed up outfit approach notably beats any individual technique.

Dhanya Bibin[3] et al. propose a novel strategy to recognize the nearness of Malaria parasites in human fringe blood smear pictures utilizing a Deep Belief Networks(DBN). This work presents a prepared model dependent on a DBN to characterize 4100 fringe blood smear pictures into the parasite or non-parasite class. The proposed DBN is pre-prepared by stacking limited Boltzmann machines utilizing the contrastive difference technique for pre-preparing. To prepare the DBN, they extract highlights from the pictures and instate the noticeable factors of the DBN. A linked component of shading and surface is utilized as an element vector in this investigation. At last, the DBN is discriminatively fine- tuned utilizing a back propagation calculation that processes the likelihood of class names. The proposed strategy has performed altogether superior to the other best in class strategies with a F-score of 89.66%, an affectability of 97.60% and particularity of 95.92%. This work is the principal use of a DBN for intestinal sickness parasite location in human fringe blood smear pictures.

A novel strategy has been proposed by Sudheer Ch[4] et. al that depends on coupling the Firefly Algorithm (FFA) and Support Vector Machines (SVM) to predict Malaria disease frequencies. The execution of SVM models relies on the proper decision of SVM parameters. In this investigation FFA has been utilized for deciding the parameters of SVM. The proposed SVM-FFA display has been received in foreseeing the malarial rates in Jodhpur and Bikaner zone where the jungle fever transmission is unsteady. Month to month midpoints of precipitation, temperature, relative moistness also, malarial rates have been considered as info factors. The outcomes show that the proposed SVM-FFA display gives progressively precise estimates contrasted with the other customary methods.

Leila Malihi[5] et.al proposed an approach in which stained components of blood for example, red platelets, parasites and white platelets are extricated. At the following stage, red platelet veil is situated on the extricated recolored components to isolate the conceivable parasites. At long last, shading histogram, granulometry, inclination and level surface highlights are separated and utilized as classifier inputs. Here, five classifiers were utilized: Support vector machines (SVM), nearest mean (NM), K closest neighbors (KNN), 1-NN, and Fisher. In this exploration

K closest neighbors classifier had the best precision, which was 91%.

### 3 DATA COLLECTION

Data from various resources has been gathered to train the machine initially. Parameters such as, as shown in below figure, fever, cold, rigor, fatigue, headache, bitter\_tongue, vomiting, diarrhea, convulsion, anemia, jaundice, cocacola\_urine, hypoglycemia, prostration, hyperpyrexia are extracted to predict the disease.

sex	fever	cold	rigor	fatigue	headache	bitter_tongue	vomiting	diarrhea	Convulsion	Anemia	jaundice	cocacola_urine	hypoglycemia	prostration	hyperpyrexia	severe_malaria
Male	yes	no	no	no	yes	no	no	no	no	yes	no	yes	yes	yes	no	no
Male	no	no	yes	yes	yes	no	no	no	no	no	yes	no	no	no	no	no
Female	yes	no	yes	no	yes	no	yes	yes	no	no	yes	yes	yes	no	no	no
Female	yes	yes	yes	no	yes	no	no	no	no	yes	yes	yes	yes	no	yes	no
Female	yes	yes	no	yes	no	no	yes	yes	yes	yes	yes	yes	no	no	no	no
Female	yes	yes	no	yes	no	yes	yes	no	no	yes	yes	yes	no	no	no	no
Male	yes	yes	yes	yes	yes	no	no	no	no	no	yes	yes	yes	no	no	yes
Male	yes	no	no	no	yes	yes	no	no	no	no	yes	no	yes	no	no	no
Female	yes	yes	no	yes	yes	yes	no	no	yes	no	no	yes	yes	yes	no	no
Male	yes	no	yes	no	yes	no	no	yes	yes	no	no	yes	no	no	no	yes
Female	no	yes	no	no	yes	no	yes	no	no	no	no	yes	no	no	no	no
Female	yes	no	no	yes	no	no	yes	yes	no	yes	no	yes	yes	yes	no	no
Male	yes	no	yes	no	yes	no	no	no	no	no	yes	yes	yes	no	no	no
Male	yes	no	no	yes	yes	yes	no	no	no	yes	no	no	no	yes	no	no
Male	yes	no	no	yes	yes	yes	no	no	yes	no	yes	yes	no	no	no	no
Female	yes	yes	no	yes	no	yes	yes	yes	no	yes	no	yes	yes	no	no	yes
Male	yes	yes	no	yes	no	no	no	yes	no	yes	no	yes	yes	no	no	yes
Male	yes	no	yes	yes	yes	no	no	no	yes	no	no	no	yes	no	no	no
Male	yes	no	no	yes	yes	yes	no	no	yes	no	yes	yes	no	no	no	no
Female	yes	yes	no	yes	no	yes	yes	yes	no	yes	no	yes	yes	no	no	yes
Male	yes	yes	no	yes	yes	no	no	yes	no	yes	no	yes	yes	no	no	no

Figure: Screenshot of dataset

Data is preprocessed by data mining tools and undergoes all the stages of mining such as collection, understanding, feature extraction. Last stage of processing extracts the necessary features for training the machine and removes unnecessary ones.

**Collection:** It is the main phase of the preprocessing, and is exceptionally pivotal, since the nature of information gathered will affect vigorously on the yield. The accumulation procedure needs to guarantee that the information assembled are both characterized and precise, with the goal that ensuing choices dependent on the discoveries are legitimate. This stage gives both the pattern from which to gauge, and an objective on what to improve.

**Preparation:** is the control of information into a structure appropriate for further examination and handling. Crude information can't be prepared and should be checked for precision. Arrangement is tied in with developing an informational index from at least one information sources to be utilized for further investigation and handling. Examining information that has not been painstakingly screened for issues can create very deceptive outcomes that are intensely reliant on the nature of information arranged.

**Input:** It is where confirmed information is coded or changed over into machine discernible structure with the goal that it very well may be handled through an application. Information section is done using a console, scanner, or information passage from a current source. This tedious procedure requires speed and exactness. Most information need to pursue a formal and severe grammar since a lot of handling power is required to breakdown the perplexing information at this stage. Because of the costs, numerous organizations are turning to re-appropriate this stage.

**Processing:** is the point at which the information is exposed to different methods and strategies for amazing specialized controls utilizing Machine Learning and Artificial Intelligence calculations to produce a yield or elucidation about the information. The procedure might be comprised of different strings of execution

that all the while execute guidelines, contingent upon the sort of information.

**Output:** is where prepared data is currently transmitted and shown to the client. Output is introduced to clients in different report positions like graphical reports, sound, video, or record watchers. Output should be deciphered with the goal that it can give important data that will control future choices of the organization.

**Storage:** is the last stage in the information handling cycle, where information, and metadata (data about information) are held for sometime later. The significance of this cycle is that it permits snappy access and recovery of the prepared data, enabling it to be passed on to the following stage legitimately, when required.

## 4 PROPOSED APPROACH

This section explains about machine learning algorithms used for training and testing the data, doctor-patient interaction portal, programming language and framework used for designing the portal.

### 4.1 Random Forest algorithm:

Random Forest has almost a similar hyperparameters as a decision tree or a bagging classifier. We don't need to consolidate a decision tree with a sacking classifier and can just effectively utilize the classifier-class of Random Forest. With Random Forest, you can manage Regression undertakings by utilizing the Random Forest regressor.

Random Forest adds extra haphazardness to the model, while developing the trees. Rather than looking for the most vital component while part a hub, it scans for the best element among an irregular subset of highlights. This outcomes in a wide assorted variety that for the most part results in a superior model.

In this manner, in Random Forest, just an irregular subset of the highlights is thought about by the calculation for part a hub. We can even make trees increasingly arbitrary, by also utilizing irregular limits for each element as opposed to looking for the most ideal edges (like a typical decision tree does).

#### Pseudo code:

1. Randomly select "k" features from total "m" features.

Where  $k \ll m$

2. Among the "k" features, calculate the node "d" using the best split point.
3. Split the node into daughter nodes using the best split.
4. Repeat 1 to 3 steps until "l" number of nodes has been reached.
5. Build forest by repeating steps 1 to 4 for "n" number times to create "n" number of trees.

### 4.2 Python:

As opposed to having the majority of its usefulness incorporated with its core, Python was intended to be profoundly extensible. This smaller seclusion has made it especially well known as a methods for adding programmable interfaces to existing applications. Van Rossum's vision of a small core language with a substantial standard library and effectively extensible mediator. While offering decision in coding system, the Python logic rejects overflowing linguistic structure, (for example, that of Perl) for a less complex, less-jumbled syntax.

### 4.3 Django

Notwithstanding having its very own classification, for example, naming the callable articles creating the HTTP reactions "views", [6] the center Django structure can be viewed as a MVC architecture. [7] It comprises of an item social mapper (ORM) that intervenes between information models (characterized as Python classes) and a social database ("Model"), a framework for handling HTTP demands with a web templating framework ("View"), and a standard articulation based URL dispatcher ("Controller").

Likewise incorporated into the framework are:

1. a lightweight and independent web server for designing and testing
2. a structure serialization and approval framework that can decipher between HTML structures and qualities appropriate for capacity in the database
3. a layout framework that uses the idea of legacy acquired from article situated programming
4. a reserving system that can utilize any of a few store strategies
5. support for middleware classes that can mediate at different phases of solicitation preparing and complete custom capacities
6. an inside dispatcher framework that enables parts of an application to impart occasions to one another by means of pre-characterized signals
7. an internationalization framework, including interpretations of Django's own parts into an assortment of dialects
8. a serialization framework that can create and peruse XML or potentially JSON portrayals of Django show occasions
9. a framework for expanding the capacities of the format motor
10. an interface to Python's worked in unit test structure

**4.4 Doctor-Patient Portal:** By making use of above explained technologies, we developed an information portal in which we incorporated all the features such as login by the doctor, selecting the symptoms of the disease, prediction result and diagnose for the disease based on the prediction.

#### Doctor Portal:

A doctor can login into his account by submitting

the valid credentials. Main page has links to the following pages: Add Patient, Patient’s history, Diagnose patient, training data. He/she can add a patient by filling the details(name, gender and date of birth) provided in the **add patient** page. In **patient’s history** window, doctor can see the details of the particular patient details such as whether the patient is already diagnosed or not. If he wants to diagnose a patient, he can add the disease symptoms in the **diagnosis** page. Random forest algorithm reads all these features and predicts whether these details lead to disease or not. After predicting the disease, it gives the **result** whether is true or not. If the result is true, **prescribed drug** is shown along with the result. Then the **perform test** provides few questions regarding the medication to test the adherence. With these details, it will perform the test and decides whether the adherence true or false.

Figure: Login page

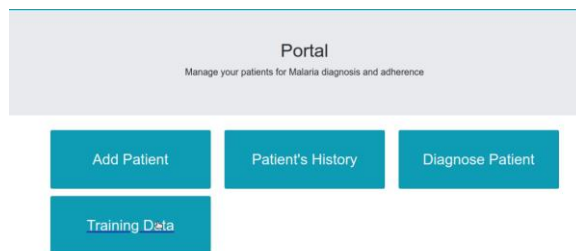


Figure: Links to the various pages



### 5 RESULTS



Figure: Doctor Portal

 A screenshot of a web form titled "Add Patient". It contains three input fields: "Full name:" with the value "john", "Gender:" with a dropdown menu showing "Male", and "Date of birth:" with a date picker showing "31/08/1999". There is a "Create" button at the bottom.

Figure: Adding a patient

 A screenshot of a web form titled "Log in". It contains two input fields: "Username:" with the value "doctor" and "Password:" with a masked password "\*\*\*\*\*". There is a "Log in" button at the bottom.

A screenshot of a web page showing patient details. At the top, it says "Patient Created Successfully". Below that, there is a breadcrumb "Home / Portal / Patients / Patient". The patient details are: "Patient: john", "Card No: 9", "Gender: m", "Date of Birth: Aug. 31, 1994", and "Registered On: March 31, 2019, 3:14 p.m.". Under "Malaria History", it says "Patient has never been diagnosed!".

Figure: Patient History

Figure: Diagnosis

Figure: Adherence Test

Diagnosis Date	Has Malaria	Prescribed Drugs	Adhered To Treatment	Take Adherence Test
March 31, 2019, 3:15 p.m.	True	Artemether & Lumefantrine	False	

Figure: Adherence Test results

Diagnosis Date	Has Malaria	Prescribed Drugs	Adhered To Treatment	Take Adherence Test
March 31, 2019, 3:15 p.m.	True	Artemether & Lumefantrine	None	<a href="#">Perform Test</a>

Figure: Diagnosis results

### 6 REFERENCES

1. Sharma V, Kumar A, Lakshmi Panat D, Karajkhede G. Malaria outbreak prediction model using machine learning. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*. 2015 Dec;4(12).
2. Bhatt, S., Cameron, E., Flaxman, S.R., Weiss, D.J., Smith, D.L. and Gething, P.W., 2017. Improved prediction accuracy for disease risk mapping using Gaussian process stacked generalization. *Journal of The Royal Society Interface*, 14(134), p.20170520.
3. Bibin D, Nair MS, Punitha P. Malaria parasite detection from peripheral blood smear images using deep belief networks. *IEEE Access*. 2017;5:9099-108.
4. Ch S, Sohani SK, Kumar D, Malik A, Chahar BR, Nema AK, Panigrahi BK, Dhiman RC. A support vector machine-firefly algorithm based forecasting model to determine malaria transmission. *Neurocomputing*. 2014 Apr 10;129:279-88.
5. Malihi L, Ansari-Asl K, Behbahani A. Malaria parasite detection in giemsa-stained blood cell images. In 2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP) 2013 Sep 10 (pp. 360-365). IEEE.
6. A. Midekisa, G. Senay, G.M. Henebry, P. Semuniguse, M.C. Wimberly, et al., Remote sensing-based time series models for malaria early warning in the highlands of Ethiopia, *Malaria J.* 11 (1) (2012) 1–10.
7. A. Bhadra, E.L. Ionides, K. Laneri, M. Pascual, M. Bouma, R.C. Dhiman, Malaria in northwest India: data analysis via partially observed stochastic differential equation models driven by Lévy noise, *J. Am. Stat. Assoc.* 106 (494) (2011) 440–451.
8. A. Das, A.R. Anvikar, L.J. Cator, R.C. Dhiman, A. Eapen, N. Mishra, B.N. Nagpal, N. Nanda, K. Raghavendra, A.F. Read, et al., Malaria in India: the center for the study of complex malaria in India, *Acta Tropica* 121 (3) (2012) 267–273.
9. Bouma, M.J. and Dye, C., 1997. Cycles of malaria associated with El Niño in Venezuela. *Jama*, 278(21), pp.1772-1774.
10. Das, A., Anvikar, A.R., Cator, L.J., Dhiman, R.C., Eapen, A., Mishra, N., Nagpal, B.N., Nanda, N., Raghavendra, K., Read, A.F. and Sharma, S.K., 2012. Malaria in India: the center for the study of complex malaria in India. *Acta tropica*, 121(3), pp.267-273.