AN EFFICIENT APPROACH FOR MALARIAL EPIDEMIC PROGNOSIS USING MACHINE LEARNING CLASSIFIERS

N. Komal Kumar

Department of Computer Science and Engineering, St. Peter's Institute of Higher Education and Research, Chennai, Tamil Nadu, India.
komalkumarnapa@gmail.com

Dr. K. Thirunadana Sikamani

Department of Computer Science and Engineering, St. Peter's Institute of Higher Education and Research, Chennai, Tamil Nadu, India.
ktsikamani@yahoo.com

Abstract

Malaria is an acute infectious disease, which affects nearly two-thirds of the global population. Annually, the deaths due to malaria have crossed millions, the countries with fewer medical facilities are the ones which are most affected, thus the prediction of its outbreak at early stages will reduce the intense diminishing of human lives. This systematic investigation deals with the analysis and prediction of the malarial epidemic outbreak by investigating several factors such as climate, global warming, human activities, mosquitoes, sewage, etc., with malarial incidence. The first stage includes data collection by the passive surveillance system; the second stage includes establishing relationships among the climatic factors with malarial incidence and finally predicting using machine learning classifiers. In the analysis, the adaptive boosted J48 (AB-J48) decision tree machine learning classifier outperformed other classifiers under the study with an accuracy of 95% in establishing a relationship among climatic factors with malarial incidence. The inferred results from the investigation are found to be stupendous which helps the public health authorities and medical practitioners to take precautionary steps to avoid more deaths.

***Keywords*: Malaria; Prediction; Epidemic; Paradigm; Outbreak.**

1. Introduction

Malaria is a grave infectious disease, where nearly half the global population is at risk [1, 2]. In 2015, nearly 212 million people were affected with malarial infectious disease causing a whopping death of 429 000 people around the globe, sub- Saharan Africa continues to be the one which shares the most. There are several factors of malaria, such as environmental conditions, human activities leading to poor sanitation, and sewage. The environmental conditions i.e. climatic factors influence a major role in contributing to malarial infections [3]. Technological advancements can help to analyze several factors causing malaria by understanding several patterns with the help of machine learning algorithms. Pattern identification is a tedious process that involves various methodologies. Probing and prophecy of the malarial epidemic contain three stages, the first stage is the data collection which is the foundation for the study, and the data collection is made possible by the online repositories. The second stage is establishing a relationship with the climatic factors since there always exists a relationship with the climatic factors [4]. Several studies suggest [5] [6] the incidence of malaria with the combination of climatic factors helps in the early prediction of malaria. The third stage uses several machine learning algorithms for pattern finding, where each pattern is tested for its prediction accuracy. The relationship between the malarial incidence and climatic factors was discussed in [7] which uses correlation and regressing to analyze the patterns among them. A malarial outbreak prediction model is proposed in[8] which uses the support vector machines, but this has some flaws such as the dataset used is smaller and there is no relationship mapping among climatic factors which makes it a less accurate prediction model. A hybrid MSO-MLP machine learning algorithm proposed in [9] uses multi swarm optimization with the multilayer perceptron for predicting dengue virus serotypes such as dengue fever, dengue hemorrhagic fever, and dengue shock syndrome, in the prediction of DENV serotypes. Diseases such as cancer [10] and Parkinson's disease [11] analyze and diagnose with the help of machine learning algorithms. A survey proposes certain machine learning algorithms [12] for disease management. Machine learning techniques for breast cancer with computer-aided diagnosis is proposed in [13]. The work in [14] proposes the computational intelligence involved in heart diagnosis. Malarial Incidence Factors (MIF) is the crucial step in the study, which elaborates in finding several factors causing malaria; these factors mainly include mosquito, human activities, climatic factors such as temperature, humidity, precipitation, solar radiation in the study area. Fig. 1 shows the several factors leading to malaria.

The work is divided into the following sections, section 2 contains the research method, section 3 contains results and discussion, and finally concluded in section 4.

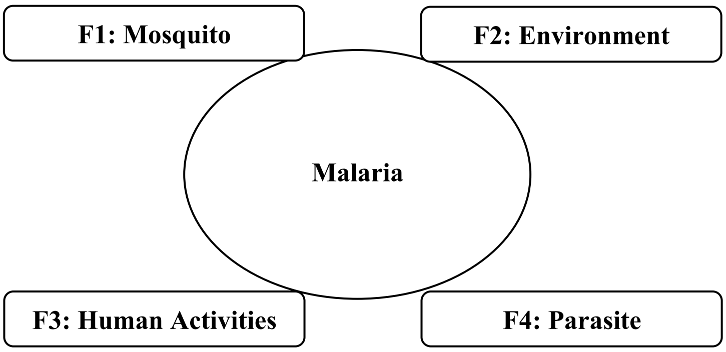

Fig. 1. Factors causing malaria

Climatic factors are dynamic since they change per global warming. The works in [15] expose that; environmental changes could affect human health. The present study was carried out in some parts of Greater Chennai Corporation (GCC) in Tamil Nadu, which has a population of 7,088,000 [16] with an area of 426 sq. km lies in 13°5′N 80°16′E latitude and longitude respectively. The climatic character of Chennai is tropical wet and dry, where the temperature peaks the mercury 35 to 40 degrees between May and June, relative humidity falls between 62 and 72 during May to September, major rainfall begins in October and ends in November. Malaria incidence data has been retrieved from the National Vector Borne Disease Control Programme, Government of India [16]. The climatic factors such as temperature, relative humidity, precipitation, wind, and solar has been retrieved from the National Centers for Environmental Prediction [17], the study area varies in boundary dimension of North-east corner: 13.3549, 80.3818 and South-west corner: 12.4715, 79.5908 as shown in Fig. 2. The climatic factor dataset contains 7 attributes and 3653 observations, Table. 1 shows sample data-Climatic factors causing malaria.

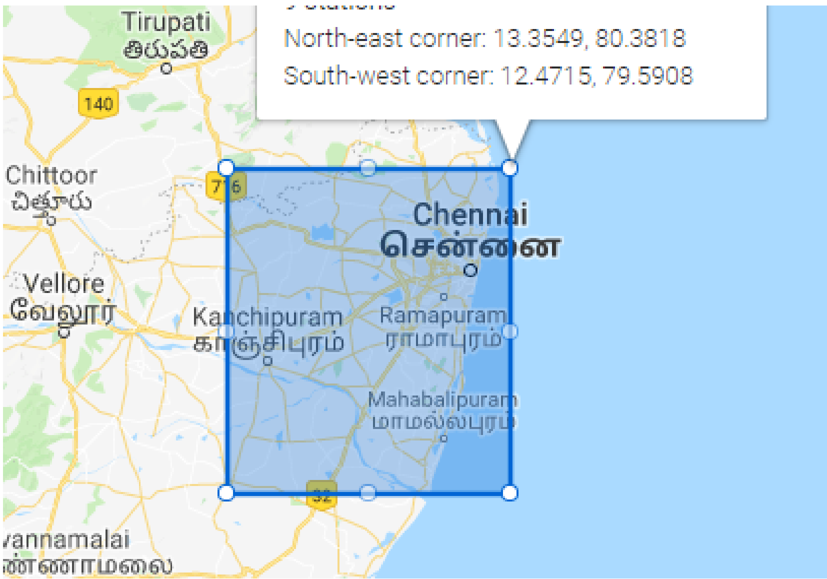

Fig. 2. Study area boundary

MalI	MaxT	MinT	PPT	Wind	RH	S
0	38.85	26.61	6.36	2.32	0.58	16.00
1	36.40	26.32	0.54	2.79	0.52	12.53
1	36.90	27.89	0.05	3.65	0.45	17.74
0	41.73	27.16	0.02	2.73	0.55	12.43
0	38.66	26.45	0.006	2.76	0.58	13.35
0	41.53	26.74	0.06	3.02	0.56	16.86
0	39.31	25.38	2.07	2.51	0.66	21.75
0	37.63	25.72	0	2.27	0.63	13.89
0	39.07	27.00	5.55	2.76	0.64	15.26
0	39.38	24.92	0.90	2.27	0.67	18.77
1	27.77	24.11	21.16	2.33	0.89	18.65
0	29.01	23.66	19.90	1.95	0.87	21.04
0	28.28	23.65	27.23	2.56	0.88	20.89
0	27.98	22.05	0.40	3.92	0.80	9.42
0	27.47	22.75	3.72	3.72	0.83	9.43
0	29.17	23.17	2.83	3.84	0.79	19.32
0	30.82	20.91	0.14	3.83	0.73	20.57
0	26.18	24.44	4.38	2.71	0.81	2.19
0	29.31	24.00	0.73	2.72	0.80	9.88

Table 1. Sample data- Climatic factors causing malaria

Note: MalI- Malarial Incidence(0-No,1- Yes) MaxT- Maximum Temperature(°C), MinT- Minimum Temperature(°C), PPT- Precipitation(kg/m2/s), Wind(in KM/s), RH- Relative Humidity, S- Solar (W/m²)

2. Research Method

The probing and prophecy of the Malarial Epidemic Paradigm (MEP) are shown in Fig. 3, which contains a hierarchy of steps in predicting malarial epidemic, are Factor Assessment (FA), Structural Equation Modeling (SEM), and Machine Learning (ML).

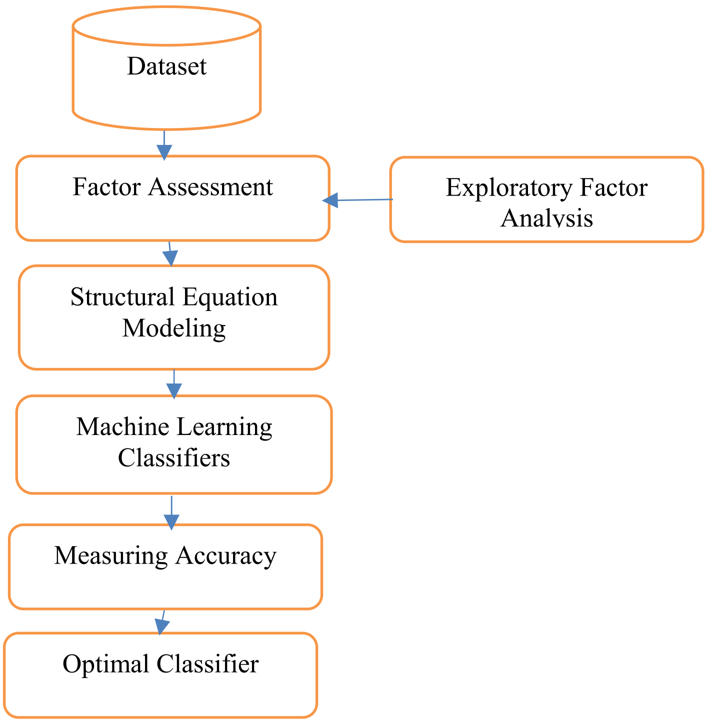

Fig. 3. Malarial epidemic paradigm

Exploratory Factor Analysis (EFA) [18] is a technique used to identify complex relationships among the climatic factors which are grouped under a common name. Factor Analysis (FA) can be explained easily using the mathematical equations. Consider the following variables X_1, X_2, X_3… X_N of size N which has small identified factors F_1, F_2, F_3, … F_M with M<<N, such that, the variables can be expressed as the factors plus leftover terms. The above description has been be sketched to mathematical equations as.

$$X_1 = \alpha_{10} + \alpha_{11}F_1 + \alpha_{12}F_2 ... + \alpha_M F_M + r_1$$
$$X_2 = \alpha_{20} + \alpha_{21}F_1 + \alpha_{22}F_2 ... + \alpha_M F_M + r_2$$
$$X_2 = \alpha_{30} + \alpha_{31}F_1 + \alpha_{32}F_2 ... + \alpha_M F_M + r_3$$
$$X_N = \alpha_{N0} + \alpha_{N1}F_1 + \alpha_{N2}F_2 ... + \alpha_M F_M + r_N$$

where r_1, r_2, r_3,…r_N are the leftover terms

Structural Equation Modeling (SEM) [19] is used in social sciences since it has the capacity of identifying complex relationships between the observed and latent variables. The SEM for the malarial incidence study have been represented as,

$$F_1 = \alpha_{10}(MinT) + \alpha_{11}(RH) + \alpha_{12}(W) + r_1$$
$$F_2 = \alpha_{20}(MaxT) + \alpha_{21}(S) + \alpha_{22}(RH) + r_2$$
$$F_3 = \alpha_{30}(MinT) + \alpha_{31}(W) + \alpha_{32}(MalI) + r_3$$
$$F_4 = \alpha_{40}(PPT) + \alpha_{41}(MalI) + \alpha_{42}(W) + r_4$$

	Mean	**Std. Deviation**	**N**
MalI	0.4585	0.49835	3653
MaxT	35.2146	5.32038	3653
MinT	23.7735	3.76556	3653
PPT	4.4354	11.30130	3653
Wind	2.6026	0.75488	3653
RH	0.6556	0.14087	3653
S	18.7512	5.82425	3653

Table 2. Descriptive statistics of the dataset

		MalI	**MaxT**	**MinT**	**PPT**	**Wind**	**RH**	**S**
MalI		1	0.020	0.011	0.020	-0.022	-0.003	-0.017
	N	3653	3653	3653	3653	3653	3653	3653
MaxT		0.020	1	.635**	-.284**	.036*	-.850**	.332**
	N	3653	3653	3653	3653	3653	3653	3653
MinT		0.011	.635**	1	.074**	.185**	-.525**	-.156**
	N	3653	3653	3653	3653	3653	3653	3653
PPT		0.020	-.284**	.074**	1	-.056**	.409**	-.328**
	N	3653	3653	3653	3653	3653	3653	3653
Wind		-0.022	.036*	.185**	-.056**	1	-.236**	.055**
	N	3653	3653	3653	3653	3653	3653	3653

RH		-0.003	-.850**	-.525**	.409**	-.236**	1	-.244**
	N	3653	3653	3653	3653	3653	3653	3653
S		-0.017	.332**	-.156**	-.328**	.055**	-.244**	1
	N	3653	3653	3653	3653	3653	3653	3653

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Table 3. Correlation matrix of Climatic factors with malarial incidence

	Initial	**Extraction**
MalI	1.000	0.679
MaxT	1.000	0.893
MinT	1.000	0.857
PPT	1.000	0.622
Wind	1.000	0.413
RH	1.000	0.653
S	1.000	0.868

Table 4. Communalities

Component	**Initial Eigen values**			**Extraction Sums of Squared Loadings**		
	Total	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
1	2.586	36.936	36.936	2.586	36.936	36.936
2	1.381	19.727	56.663	1.381	19.727	56.663
3	1.019	14.558	71.221	1.019	14.558	71.221
4	0.958	13.685	84.906			
5	0.688	9.834	94.739			
6	0.286	4.082	98.821			
7	0.083	1.179	100.000			

Table 5. Total variance explained with Components, initial Eigen values and Extraction Sums of Squared Loadings (SSL)

	Component		
	1	2	3
MalI	0.005	0.080	0.820
MaxT	0.929	0.079	0.153
MinT	0.653	0.657	0.007
PPT	-0.464	0.637	0.014
Wind	0.265	0.142	-0.568
RH	0.378	-0.714	0.015
S	-0.931	-0.028	-0.001

Table 6. Component matrix

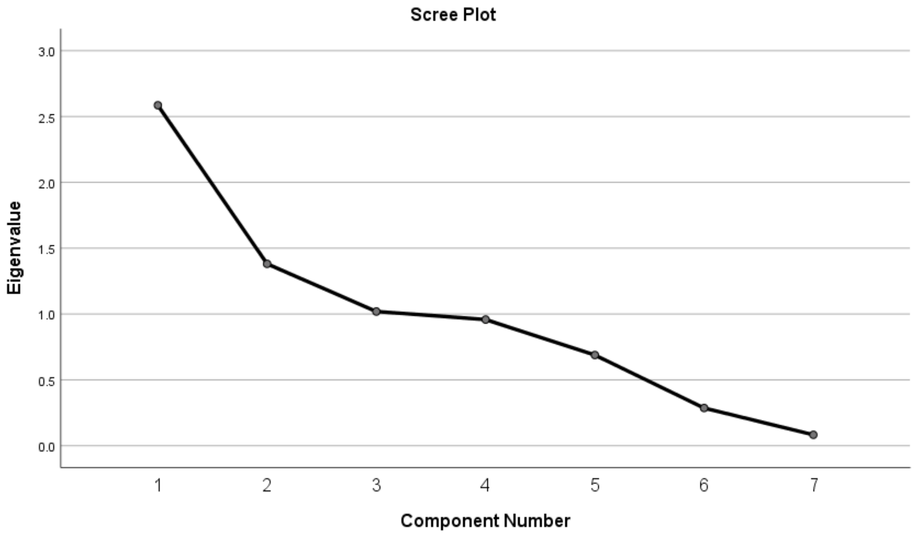

Fig. 4. Scree plot

Table. 2 represents the descriptive statistics such as mean and standard deviation of the dataset containing 3653 observations, Table. 3 entails the Pearson correlation [20] between the climatic factors with malarial incidence where the correlation is significant at the 0.01 level (2-tailed) for some factors and 0.05 level (2-tailed) for other factors. Table. 4 shows the communalities of the climatic factors with malarial incidence, they define how one variable correlates with other variables, and Maximum Temperature (MaxT) has high communality with the malarial incidence (0.893), whereas Wind (W) significantly has low communality with the malarial incidence (0.413), so the temperature has a major effect in the incidence of malaria. Table. 5 depicts the total variance with components, initial Eigen values, and Extractions of Sums of Squared Loadings (SSL). Table. 6 shows the component matrix which estimates the correlation between each variable with the components, in this analysis three components, were extracted, the Maximum Temperature, Minimum Temperature, Malarial Incidence having the positive values whereas the Precipitation, Wind, Solar have negative values, this is due to the bi-annual rainfall in the regions of Chennai. With Table. 1, the Scree plot [21] is obtained, which depicts that two components namely Malarial Incidence (MalI) and Maximum Temperature (MaxT) are greater than the unity and other factors such as Precipitation (PPT), Wind (W), Relative Humidity (RH) and Solar (S) are lesser than the unity which is shown in Fig. 4.

3. Results and Discussion

The next step is to identify the patterns which will provide the accuracy for malarial prediction. Machine learning classifiers learn complex patterns from the climatic factors such as Maximum Temperature, Minimum Temperature, Precipitation, Solar, Wind, and Relative Humidity with the malarial incidence parameter and predict the forthcoming. Machine learning classifiers like Support Vector Machine (SVM) [22], Decision Tree (DT) [23], K- Nearest Neighbor (K-NN) [24], and Multi-Layer Perceptron (MLP) [25] are applied to the structurally modeled data from the previous stage. A dataset containing 3653 observations with malarial incidence data are split into training and test dataset with split criteria of 70% and 30% training and test data respectively. To evaluate the classifier model, 10-fold cross-validation technique has been used. Table 7 represents the MLP network information, containing factors, hidden layers, and dependent variables. Hyperbolic tangent is used as an activation function for the construction of hidden layers. The following tables represent the initial settings and results obtained for machine learning classifiers.

Input Layer	**Factors**	1	MaxTemperature
		2	MinTemperature
		3	MalarialIncidence
	Number of Units[a]		4726
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		8
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Precipitation
		2	RelativeHumidity
		3	Solar
		4	Wind
	Number of Units		4
	Rescaling Method for Scale Dependents		Standardized
	Activation Function		Identity
	Error Function		Sum of Squares

a. Excluding the bias unit

Table 7. Multi-Layer Perceptron network information

Training	**Sum of Squares Error**		5187.198
	Average Overall Relative Error		1.016
	Relative Error for Scale Dependents	Precipitation	1.015
		RelativeHumidity	1.008
		Solar	1.017
		Wind	1.022
	Stopping Rule Used		1 consecutive step(s) with no decrease in error[a]
	Training Time		0:02:25.89
Testing	Sum of Squares Error		23.338
	Average Overall Relative Error		0.908
	Relative Error for Scale Dependents	Precipitation	1.088
		RelativeHumidity	0.908
		Solar	0.980
		Wind	0.757

Table 8. MLP model summary

Table 8. shows the Multi-Layer Perceptron model summary containing sum of the squares error, average overall relative error, relative error for scale dependents, training time to build the model.

Specifications	Growing Method	CHAID
	Dependent Variable	MaxTemperature
	Independent Variables	MalarialIncidence, MinTemperature, Precipitation, Wind, RelativeHumidity, Solar
	Validation	None
	Maximum Tree Depth	3
	Minimum Cases in Parent Node	100
	Minimum Cases in Child Node	50
Results	Independent Variables Included	RelativeHumidity, Solar, MinTemperature, Wind, Precipitation
	Number of Nodes	62
	Number of Terminal Nodes	41
	Depth	3

Table 9. Model Summary for DT

Node	N	Percent	Mean
13	99	2.7%	44.9731
12	132	3.6%	42.7951
43	74	2.0%	41.6982
16	95	2.6%	41.6699
49	147	4.0%	40.7170
44	114	3.1%	40.6569
47	63	1.7%	39.8069
42	60	1.6%	39.4775
21	87	2.4%	38.9078
45	69	1.9%	38.7321
48	61	1.7%	38.2742
50	133	3.6%	38.1382
14	87	2.4%	38.1157
19	168	4.6%	37.5490
46	95	2.6%	37.1383
20	110	3.0%	36.9827
51	93	2.5%	36.8323

Table 10. Gain Summary for DT

Table. 9 shows the model summary of decision tree, the growing method is chosen as CHAID [26] with MaxTepmerature as dependent variable and Malarial Incidence, MinTemperature, Precipitation, Wind, RelativeHumidity, Solar as independent variables, the maximum depth is chosen as 3 with minimum case in parent node as 100 and minimum case in child node as 50. Table. 10 shows the various gain of Decision Tree with node number, number of observations with its percentage and mean. Table 10 shows the gain summary for decision tree.

		K=2		K=3		K=4	
		N	Percent	N	Percent	N	Percent
Sample	**Training**	2548	69.8%	2591	70.9%	2555	69.9%
	Holdout	1105	30.2%	1062	29.1%	1098	30.1%
Valid		3653	100.0%	3653	100.0%		3653
Excluded		779			779		779
Total		4432			4432		4432

Table 11. Case processing for K-NN

Table. 11 represents the case processing of K- Nearest Neighbor with K=2, 3, 4 with its training and test percentage and validity.

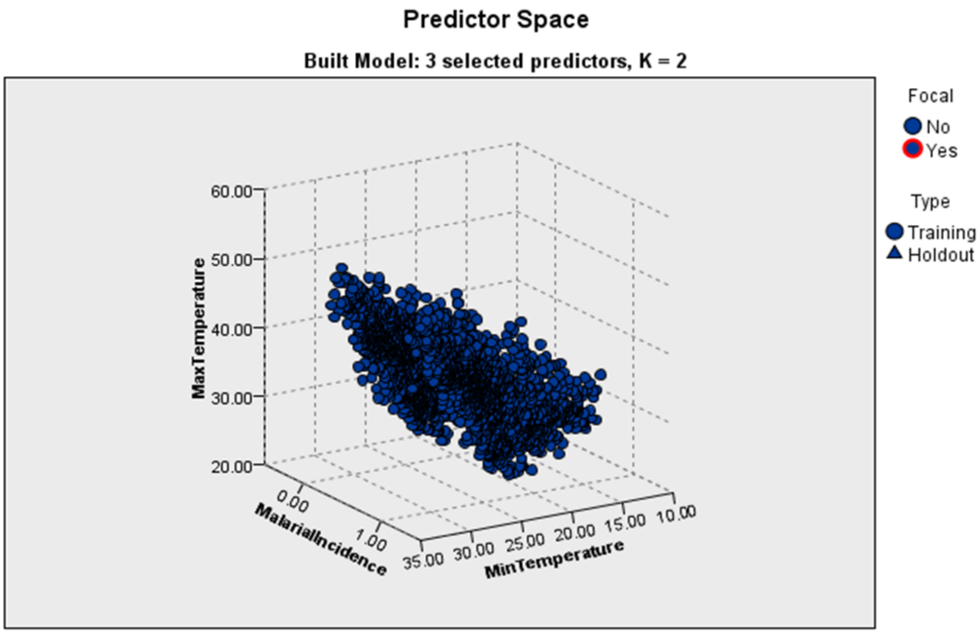

Fig. 5. Predictor Space for KNN (K=2)

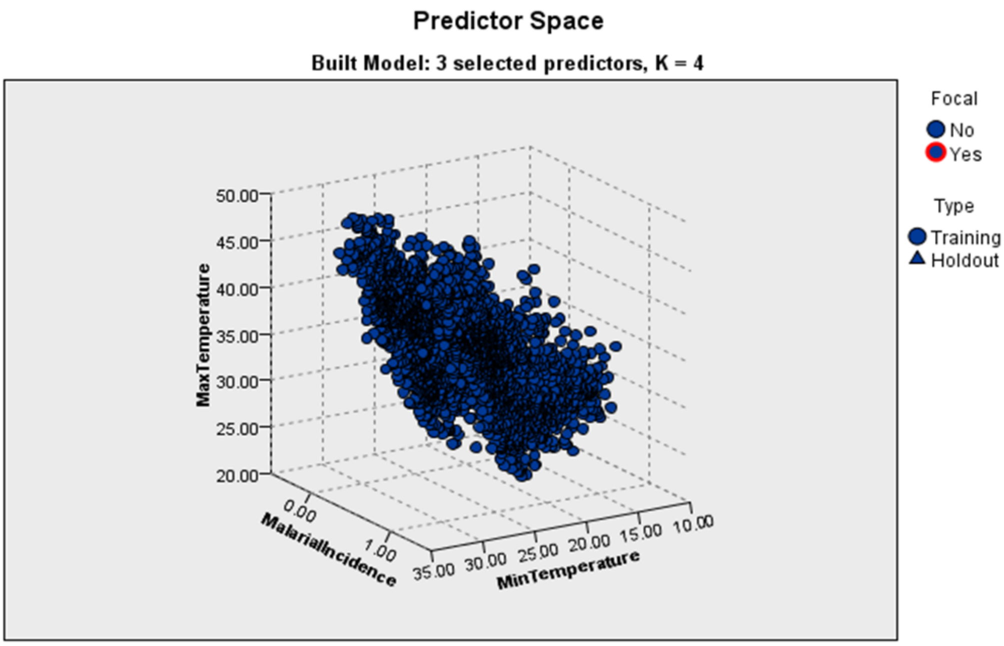

Fig. 6. Predictor Space for KNN (K=4)

Fig. 5 and 6 shows the Predictor Space of K-NN when K=2, with selected predictors as 3 and Predictor Space of K-NN when K=4, with selected predictors as 3.

Classifier	MLP	DT	SVM	KNN2	KNN3	KNN4	AB-J48
Accuracy	83%	78%	94%	71%	67%	65%	95%
Execution Time (s)	0.24	0.64	0.12	0.23	0.29	0.25	0.68

Table 12. Comparison of machine learning classifier accuracy and execution time, MLP: Multi- Layer Perceptron, DT: Decision Tree, SVM: Support Vector Machine, KNN: K- Nearest Neighbor

Multi-Layer Perceptron predicts the malarial incidents with an accuracy of 83% with an execution time of 0.24s due to its supervised learning capability with three layers, Decision Tree has achieved an accuracy of 78% with an execution time of 0.64s due to the fuzzy nature of the dataset, K-NN achieved an accuracy of 71% with 0.23s when K=2, 67% with 0.29s when K=3, 65% with 0.25s when K=4 failed to predict, this may be due to the noise in data, further preprocessing has to be done for improving its accuracy. Support Vector Machine (SVM) has achieved an accuracy of 94% with an execution time of 0.12s. The AB-J48 has outperformed all classifiers under analysis by producing an accuracy of 95% with an execution time of 0.68s in predicting malarial incidence as shown in Fig. 7 and Fig. 8. Fig. 9 shows the Receiver Operating Characteristics (ROC) for AB-J48 which is calculated to be 0.9538.
Several studies have utilized machine learning classifiers in their work such as network anomaly detection [27], breast cancer [28], aerobic granular sludge [29], acute kidney disease [30], student's performance [31], melanoma detection [32], recognizing human activity [33], regression-based model [34], Sentiment Analysis [35], predicting chronic kidney disease [36], Soybean leaf disease detection [37], investigating dengue outbreak [38], sentiment of mobile unboxing [39, 40].

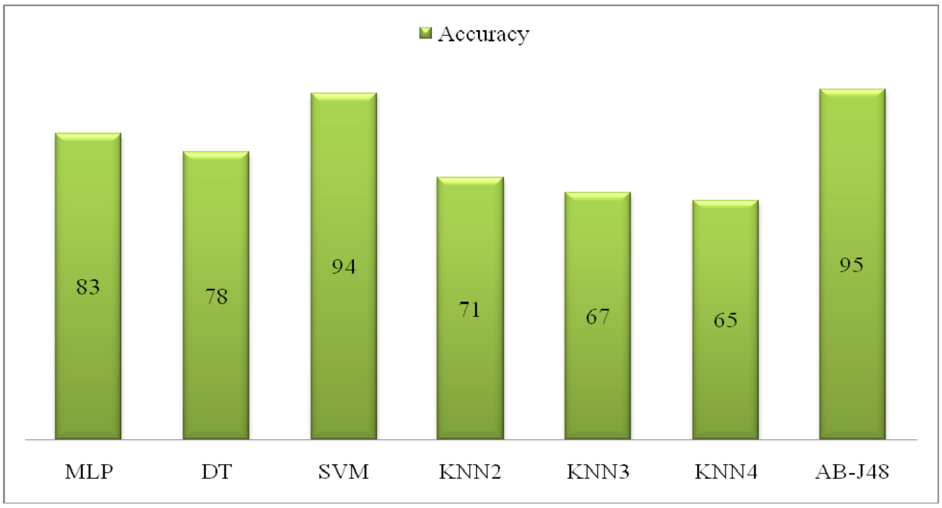

Fig. 7. Accuracy of the classifiers

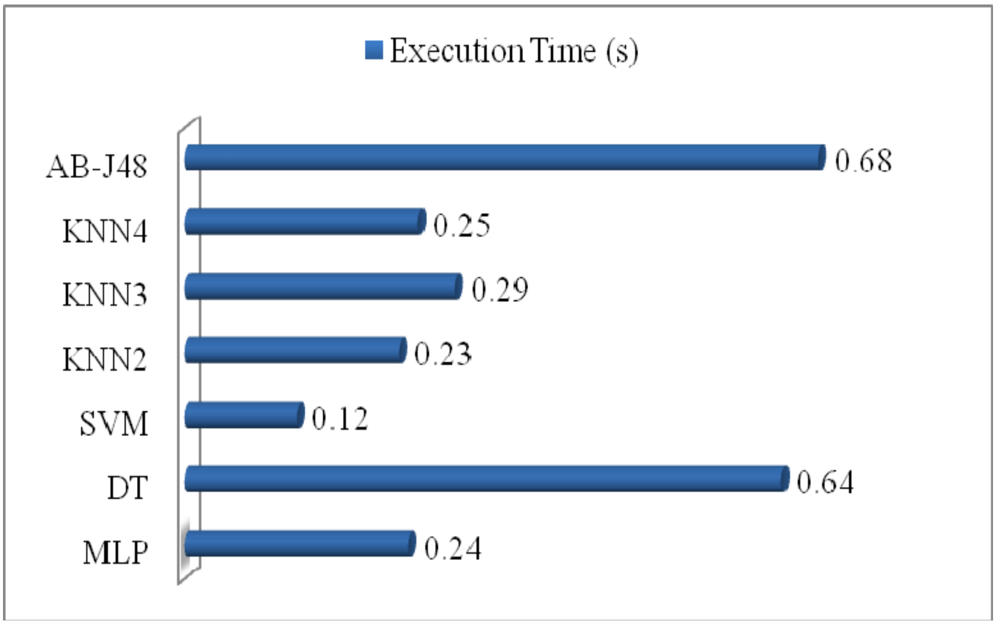

Fig. 8. Execution time of the classifiers

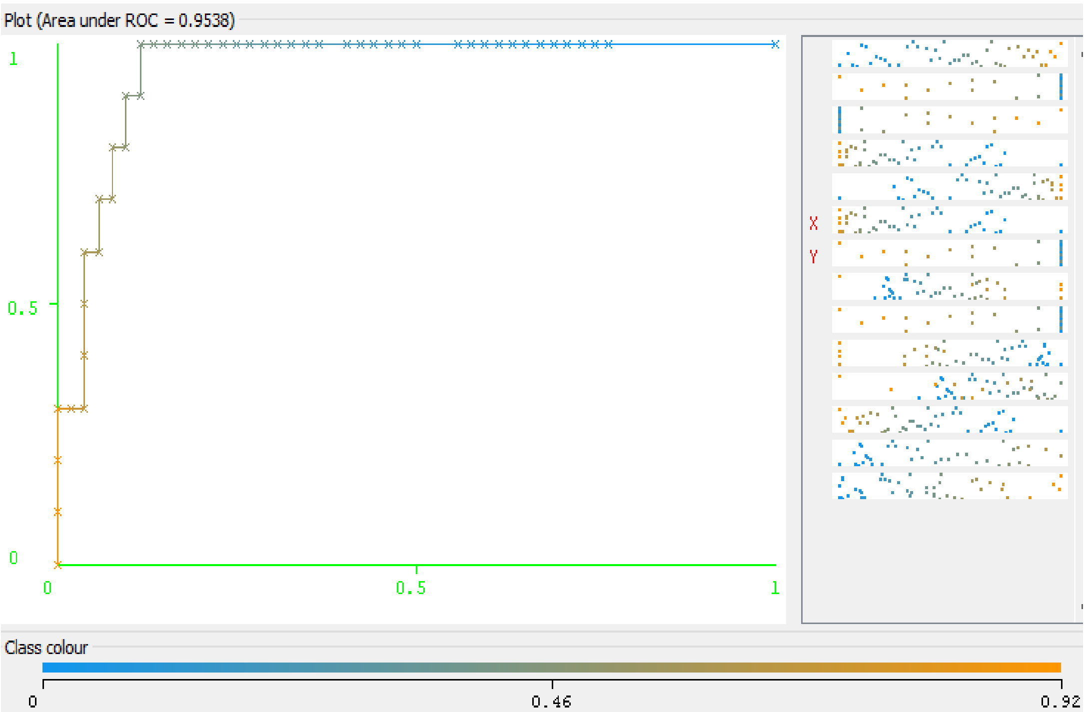

Fig. 9. Receiver Operating Characteristics (ROC) for AB-J48

4. Conclusion

In this study, we have investigated the malarial incidence in the areas of Greater Chennai Corporation (GCC), Tamil Nadu, India. Prediction of malarial incidence is done by analyzing the climatic factors with the help of statistical techniques and machine learning classifiers. The AB-J48 decision tree-based machine learning classifier outperformed all the classifiers under analysis. This investigation paradigm helps the hospitals, clinical centers, and health organizations to take preventive measures before the occurrence of the disease. The climatic factors causing the malarial outbreak in the study region are also discussed. The results inferred from the investigation show that temperature and relative humidity has a greater effect on the growth of mosquitoes causing malaria. Further development of the system can be done through dynamic data collection and processing which saves time and human beings.

Acknowledgements

The authors are grateful to the management of St. Peter's Institute of Higher Education and Research, Chennai, India that greatly assists research and development

References

[1] Hartman TK, Rogerson SJ, Fischer PR. The impact of maternal malaria on newborns. Annals of Tropical Paediatrics. 30 (4): 271–82, 2010.
[2] http://www.who.int/features/factfiles/malaria/en/
[3] Haque, U, et al., The role of climate variability in the spread of malaria in Bangladeshi highlands. PLoS ONE, 5, e14341, 2010.
[4] Bonan, G.B, Shugart, H.H,. Environmental factors and ecological processes in boreal forests. Annu. Rev.Ecol. Syst, 20, 1–28, 1989.
[5] Kumar, V, et al., Forecasting malaria cases using climatic factors in Delhi, India: A time series analysis. Malar. Res. Treat, 2014.
[6] Ngarakana-Gwasira, E.T.; Bhunu. Assessing the Role of Climate Change in Malaria Transmission in Africa. Malar. Res. Treat, 2016.
[7] Nath, D.C.; Mwchahary, D.D,. Association between Climatic Variables and Malaria Incidence: A Study in Kokrajhar District of Assam, India: Climatic Variables and Malaria Incidence in Kokrajhar District. Glob. J.Health Sci, 5, 90, 2013.
[8] Sharma, V, et al., A. Malaria outbreak prediction model using machine learning. Int. J. Adv. Res. Comput. Eng. Technol, 4, 4415–4419, 2015.
[9] Komal Kumar.N, Vigneswari.D, B A S Roopa Devi. MSO – MLP diagnostic approach for detecting DENV serotypes, International Journal of Pure and Applied Mathematics, 118, 5, 1-6, 2018.
[10] Ganesan, N.; Venkatesh, K.; Rama, M.A. Application of Neural Networks in diagnosing cancer disease using demographic data. Int. J. Comput. Appl, 1, 76–85, 2010.
[11] Aditya, M.; Prince, K.; Himanshu, A.; Pankaj, K., Early heart disease prediction using data mining techniques. Comput. Sci. Inf. Technol, 53–59, 2014.
[12] Enas M.F. El Houby,. A survey on applying machine learning techniques for management of diseases. Journal of Applied Biomedicine. Volume 16, Issue 3, Pages 165-174, 2018.
[13] Nisreen I.R. Yassin et al., Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. Computer Methods and Programs in Biomedicine, Volume 156, Pages 25-45, 2018.
[14] JesminNahar, et al.,Computational intelligence for heart disease diagnosis: A medical knowledge driven approach. Expert Systems with Applications, Volume 40, Issue 1, Pages 96-104, 2013.
[15] Bayles, B.R, et al., Ecosystem Services Connect Environmental Change to Human Health Outcomes. EcoHealth, 13, 443–449, 2016.
[16] http://nvbdcp.gov.in/index1.php?lang=1&level=1&sublinkid=5784&lid=3689
[17] https://globalweather.tamu.edu/
[18] Polit DF Beck CT, Nursing Research: Generating and Assessing Evidence for Nursing Practice, 9th ed. Philadelphia, USA: Wolters Klower Health, Lippincott Williams & Wilkins, 2012.
[19] Tarka, Piotr, An overview of structural equation modeling: Its beginnings, historical development, usefulness and controversies in the social sciences. Quality & Quantity, 52 (1): 313–54, 2017.
[20] Karl Pearson, Notes on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London, 58: 240–242, 1895.
[21] Cattell, Raymond, The scree test for the number of factors. Multivariate Behavioral Research. 1 (2): 245–76, 1966.
[22] Cortes, Corinna; Vapnik, Vladimir N., Support-vector networks. Machine Learning. 20 (3): 273–297, 1995.
[23] Quinlan, J. R. Simplifying decision trees. International Journal of Man-Machine Studies. 27 (3): 221, 1987.
[24] Altman, N. S., An introduction to kernel and nearest-neighbor nonparametric regression, The American Statistician, 46 (3): 175–185, 1992.
[25] Artificial Neural Networks as Models of Neural Information Processing | Frontiers Research Topic
[26] McKenzie, Dean P, Constructing a Minimal Diagnostic Decision Tree. Methods of Information in Medicine, Vol. 32, pp. 161–166, 1993.
[27] Mukrimah Nawir et al., Effective and efficient network anomaly detection system using machine learning algorithm. Bulletin of Electrical Engineering and Informatics, Vol. 8, No.1 pp. 46-51, 2019.
[28] Wan Nor Liyana Wan Hassan Ibeni et al., Comparative analysis on bayesian classification for breast cancer problem. Bulletin of Electrical Engineering and Informatics, Vol. 8, No. 4, pp. 1303-1311, 2019.
[29] Nur Sakinah Ahmad Yasmin et al., "Performance comparison of SVM and ANN for aerobic granular sludge. Bulletin of Electrical Engineering and Informatics, Vol. 8, No. 4, pp. 1392-1401, 2019.
[30] Zainuri Saringat et al., Comparative analysis of classification algorithms for chronic kidney disease diagnosis, Bulletin of Electrical Engineering and Informatics, Vol. 8, No. 4, pp. 1496-1501, 2019.
[31] Ahmad Firdaus Zainal Abidin et al., Adaboost-multilayer perceptron to predict the student's performance in software engineering. Bulletin of Electrical Engineering and Informatics, Vol. 8, No. 4, pp. 1556-1562, 2019.
[32] Maen Takruri et al., PSO-SVM hybrid system for melanoma detection from histo-pathological images. International Journal of Electrical and Computer Engineering, Vol.9, No.4, pp. 2941-2949, 2019.
[33] Abdul Lateef Haroon P.S,U. Eranna, A simplified machine learning approach for recognizing human activity. International Journal of Electrical and Computer Engineering, Vol.9, No.5, pp. 3465-3473, 2019.
[34] Abdulkadir Özdemir, Uğur Yavuz, Fares Abdulhafidh Dael,. Performance evaluation of different classification techniques using different datasets. International Journal of Electrical and Computer Engineering, Vol.9, No.5, pp. 3584-3590, 2019.
[35] Chanintorn Jittawiriyanukoon, Estimation of regression-based model withbulk noisy data. International Journal of Electrical and Computer Engineering, Vol.9, No.5, pp. 3649-3656, 2019.
[36] M. Ali Fauzi, Random Forest Approach for Sentiment Analysis in Indonesian Language, Indonesian Journal of Electrical Engineering and Computer Science, Vol. 12, No. 1, pp. 46~50, 2018.
[37] Komal KumarN, R. Lakshmi Tulasi, VigneswariD, An ensemble multi-model technique for predicting chronic kidney disease, International Journal of Electrical and Computer Engineering, Vol.9, No.2, pp. 1321-1326, 2019.
[38] Sachin B. Jadhav, Vishwanath R. Udupi, Sanjay B. Patil, Soybean leaf disease detection and severity measurement using multiclass SVM and KNN classifier. International Journal of Electrical and Computer Engineering, Vol.9, No.5, pp. 4077~4091, 2019.

[39] N. Komal Kumar , R. Lakshmi Tulasi , D. Vigneswari, Investigating dengue outbreak in Tamil Nadu, India. Indonesian Journal of Electrical Engineering and Computer Science, Vol. 18, No. 1, pp. 502~507, 2020.
[40] Chaithra V D, Hybrid approach: naive bayes and sentiment VADER for analyzing sentiment of mobile unboxing video comments. International Journal of Electrical and Computer Engineering, Vol.9, No.5, pp. 4452~4459, 2019.

Author's Profile

	**N. Komal Kumar** is currently pursuing his PhD in Department of Computer Science and Engineering, St. Peter's Institute of Higher Education and Research. His research interests include machine learning, data mining, and cloud computing.
	**Dr. K.Thirunadana Sikamani** is currently working as Professor in the Department of Computer Science and Engineering, St. Peter's Institute of Higher Education and Research. His research interests include image processing, machine learning, and data mining.