SUST Journal of Engineering and Computer Sciences (JECS), Vol. 21, No. 2, 2020

A Big-Bang Big-Crunch Type-2 Fuzzy Logic-based System for Malaria Epidemic Prediction in Ethiopia

Belay Enyew Chekol[1] and Hani Hagras[2]

[1]Department Computer Science, Sudan University of Science and Technology, Sudan
[2] School of Computer Science and Electronic Engineering, University of Essex, UK
belayenyew@gmail.com

Received: 12/03/2020
Accepted: 10/04/2020

ABSTRACT- Malaria is a life-threatening disease caused by Plasmodium parasite infection with huge medical, economic, and social impact. Malaria is one of a serious public health problem in Ethiopia since 1959, even if, its morbidity and mortality have been reduced starting from 2001. Various studies were conducted to predict the Malaria epidemic using mathematical and statistical regression approaches, nevertheless, they had no learning capabilities. In this paper, we presented a type-2 fuzzy logic-based system for Malaria epidemic prediction (MEP) in Ethiopia which has been optimized by the Big-Bang Big-Crunch (BBBC) approach to maximizing the model accuracy and interpretability to predict for the future occurrence of Malaria. We compared the proposed BBBC optimized type-2 fuzzy logic-based system against its counterpart T1FLS, non-optimized T2FLS, ANFIS and ANN. The results show that the optimized proposed T2FLS provides a more interpretable model that predicts the future occurrence of Malaria from one up to three months ahead with optimal accuracy. This helps to answer the question of when and where must make preparation to prevent and control the occurrence of Malaria epidemic since the generated rules from our system were able to explain the situations and intensity of input factors which contributed to the Malaria epidemic and outbreak.

Keywords: *Type-2 fuzzy logic system, Big Bang-Big Crunch, Malaria prediction, machine learning.*

المستخلص – الملاريا مرض يهدد الحياة وينتج عن عدوى طفيل البلازميوم وله تأثير طبي واقتصادي واجتماعي ضخم. الملاريا هي واحدة من مشاكل الصحة العامة الخطيرة في إثيوبيا منذ عام 1959، وقد تم تخفيض معدلات الاعتلال والوفيات بدءًا من عام 2001. وقد أجريت دراسات مختلفة للتنبؤ بمرض الملاريا باستخدام مناهج الانحدار الرياضي والإحصائي، ومع ذلك ، لم يكن لديهم قدرات التعلم الذكي. في هذه الورقة ، قدمنا نظام مبني على المنطق من النوع 2 للتنبؤ بمرض الملاريا (MEP) في إثيوبيا تم تحسينه من خلال نهج الانفجار الكبير أزمة كبيرة (BBBC) لزيادة دقة النموذج وإمكانية تفسيره إلى أقصى حد للتنبؤ بحدوث الملاريا في المستقبل. قارنا النظام القائم على المنطق الضبابي المحسن من النوع 2 من BBBC مع نظيره و النوع 1 نظام المنطق الضبابي T1FLS و النوع 2 نظام المنطق الضبابي غير الأمثل T2FLS و تنسيق نظام الاستدلال الضبابي العصبي التكيفي ANFIS والشبكات العصبية ANN. تشير النتائج إلى أن T2FLS المقترح الأمثل يوفر نموذجًا أكثر قابلية للتفسير يتنبأ بحدوث الملاريا في المستقبل من شهر إلى ثلاثة أشهر متواصلة بدقة مثلى. يساعد هذا على الإجابة عن السؤال الخاص بالوقت والمكان الذي يجب فيه التحضير لمنع حدوث وباء الملاريا والسيطرة عليه لأن القواعد التي تم إنشاؤها من نظامنا كانت قادرة على شرح حالات وكثافة عوامل الإدخال التي ساهمت في وباء الملاريا وتفشيها.

Introduction

Malaria is a life-threatening disease caused by Plasmodium parasite infection with huge medical, economic, and social impact [1]. According to WHO 2019 report [2] in 2018, an estimated 228 million cases of Malaria occurred worldwide and most cases were in the WHO African Region (213 million or 93%), followed by the WHO South-East Asia Region with 3.4% of the cases and the WHO Eastern Mediterranean Region with 2.1%.

Ethiopia is among the highly affected countries by Malaria endemic, mainly due to its varying topographical and climatic features. Malaria disease has a high Epidemiologic and socio-economic burdens. It is estimated that about 75% of the total area of the country and 65% of the population to be at risk of infection [3]. Figure 1, shows the number of Malaria cases (in million) and trends from 2012 to 2017 in Ethiopia; and thus the incidence is still high [4].

According to WHO 2018 country profile report[4] in 2016 and 2017,more than 65 million of USA dollars allocated in each years for Malaria intervention and control activities in Ethiopia.

Historical in Ethiopia, a major Malaria epidemic in 1958 resulted in an estimated 3 million cases of whom 150 000 died and such kind of large-scale epidemic has been returned at some irregular intervals of years; for example, during the 1980s and 1990s, severe epidemics were recorded in 1981, 1988, 1991, 1992 and 1998 [5].

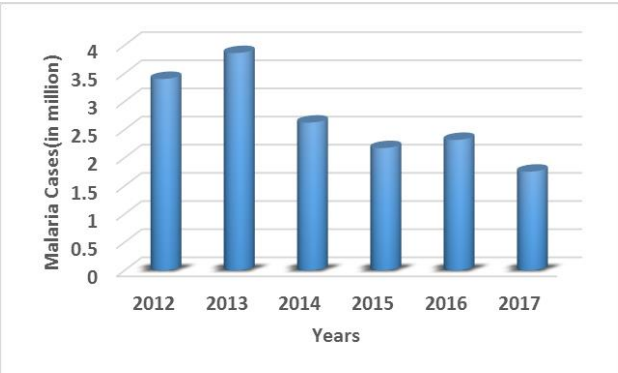

Figure 1: Clinical confirmed Malaria cases in Ethiopia from 2012-2017 [4]

Malaria is not only a major health problem but also has a significant economic impact on the socio-economic development of nations. The full nature of the economic burdens of malaria epidemics remains unclear in Ethiopia. Studies conducted in stable transmission areas have shown that Malaria causes substantial losses to households in the form of foregone income, treatment costs, school abstain , and decreased agricultural production; and in the unstable Malaria affects children and people in the productive age groups, resulting in substantial economic loss because of the compromised capacity and efficiency of the labor force [6].

Various studies showed the importance of climatic variables and morbidity data to predict Malaria transmission and early detection using elevation, temperature, rainfall and humidity as coincided with the increased magnitude and frequency of Malaria epidemics [7].

Malaria epidemic prediction (MEP) is a challenging problem due to its non-stationary nature and many influencing factors. This research aims to present novel techniques which can operate on complicated MEP and identify the main causes behind it and which could be then used in decision making.

Related Works

The main idea of Fuzzy Logic Systems(FLS) was introduced by Zadeh[8] and was first applied to control theory in 1974 by Mamdani[9]. Based on these works, fuzzy controllers have been successfully used in various real-world applications, like in [10][11][12].

Type-1 Fuzzy Logic Systems

FLS that is defined entirely in terms of type-1 fuzzy sets is known as a type-1 fuzzy logic system (T1FLS), and its elements are displayed in the following Figure 2 [13].

A type-1 fuzzy set in the universe X is characterized by a membership function $\mu_A(x)$ taking values on the interval $[0,1]$ and can be represented as a set of ordered pairs of an element and defined by the following Equation(1):

$$A = \{(x, \mu_A(x)) | x \in X\} \quad (1)$$

Where $\mu_A(x) \rightarrow [0,1]$

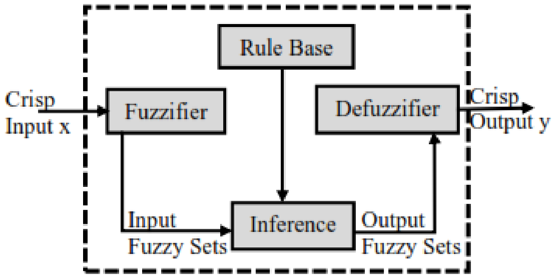

Figure 2: Type-1 FLS Structure

Type-2 Fuzzy Logic Systems

A type-2 fuzzy set is characterized by a fuzzy membership function, i.e., the membership value for each element of this set is a fuzzy set in [0, 1], unlike a type-1 set where a membership grade is a crisp number in [0, 1][13]. An interval type-2 fuzzy set A is characterized by Gaussian membership function shown in Figure 3.

The uncertainty is represented by a gray region called the foot-print of uncertainty (FOU). The uniform shading for the FOU represents the entire interval type-2 fuzzy set, and it can be described in term of an upper membership function $\bar{\mu}_{\tilde{A}}(x)$ and a lower membership function $\underline{\mu}_{\tilde{A}}(x)$ [14].

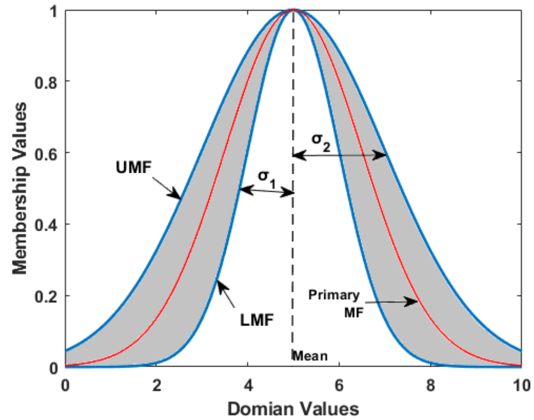

Figure 3: a Type-2 membership function and a red line in the middle is a type-1 membership function

An interval type-2 fuzzy set, $\tilde{A}$, is characterized by Equation (2):

$$\tilde{A} = \{((x,u), \mu_{\tilde{A}}(x,u)) | x \in X, u \in [0,1]\} \quad (2)$$

where $0 \le \mu_{\tilde{A}}(x,u) \le 1$.

In the case of an interval type-2 membership function, the secondary membership value is $\mu_{\tilde{A}}(x,u) = 1, \forall_x \in X, \forall_u \subseteq [0,1]$.

The architecture of type-2 fuzzy logic system (T2FLS) shown in Figure 4. The output processor includes a type-reducer and defuzzifier that generates a type-1 fuzzy set output (from the type reducer) or a crisp number (from the defuzzifier)[14].

An interval type-2 FLS is also characterized by IF–THEN rules, but their fuzzy sets are now of interval type-2 form. The type-2 fuzzy set can be used when circumstances are too uncertain to determine exact membership degrees, as is the case when the membership functions in a fuzzy controller can take different values and we want to find the distribution of membership functions to show better results in the stability of fuzzy control[15].

In Malaria prone areas, different studies conducted in the past to predict Malaria epidemic and transmission by applying Statistical and mathematical approaches using climate and Malaria cases data, like in seasonal autoregressive integrated moving average (SARIMA) [16][17], Poisson regression [18], autoregressive integrated moving average (ARIMA)[19][20].

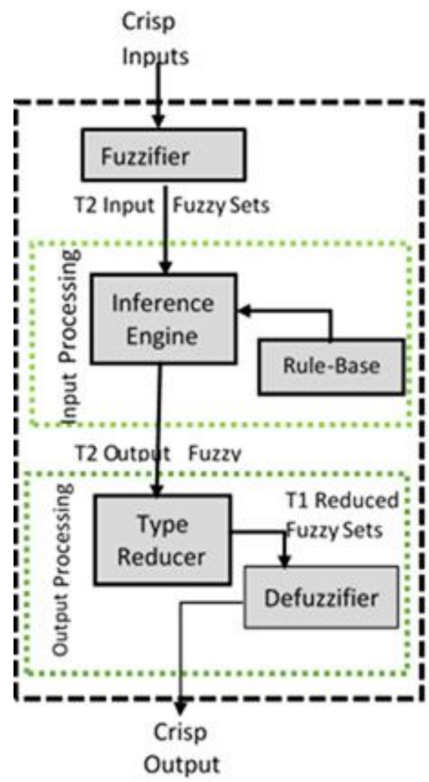

Figure 4: Architecture of type-2 fuzzy logic system

However, the traditional statistical and mathematical techniques did not learn the situation from historical data to capture trends properly and they are not accurate enough for predicting the abnormal incidence of Malaria for use in an epidemic early warning system. In recent years, there are attempts to apply Artificial Intelligence techniques to develop predictive models for Malaria incidence and epidemic. For example, Support Vector Regression (SVR) and Adaptive Neuro-Fuzzy inference System (ANFIS)[21], Bayesian networks[22], Artificial and Evolving Neural Network[23] were proposed to predict the Malaria epidemic.

Even though such machine learning approaches provided a learning capability and good prediction accuracies, they have lacked to handle uncertainties and noise in Malaria prediction. In addition, such machine learning techniques are black-box and they have limitations to represent in more interpretable and understandable form the results for decision makers.

One of the commonly used white-box approach is Fuzzy Logic Systems (FLS) which could be easily analyzed and understood by the layman user[10]. For example, in the previous study of this paper's authors showed that the T2FLS gave promising results with better interpretability for MEP [12]. To maximize the performance and interpretability of a designed Type-2 Fuzzy Logic-based System, the use of some optimizing methods becomes a valuable choice when problem complexity grows and modeling becomes difficult. The ability of fuzzy logic systems to be hybridized with other methods extended the usage of fuzzy logic systems in many application domains[24].

FLS with different optimization techniques became more interpretable with reduced number of rules and has good accuracy by adjusting membership functions in the real-world applications like in genetic algorithms [10][25][26][27][28], big bang-big crunch [29][30] [31] [32] and bio-inspired (bee colony, ant colony and particle swarm) optimization [33][34].

The main objective of this study is to develop an intelligent fuzzy logic-based system which is BBBC optimized type-2 Fuzzy Logic Systems to get an optimal prediction of the MEP ahead of one up to three months. Further, the results of optimized T2FLS compared with against its counterpart optimized T1FLS, ANFIS, Artificial Neural Network and non-optimized T2FLS.

The membership functions of the proposed system extracted using the Fuzzy C-Means (FCM) Clustering technique which is available in MATLAB R2019a built-in Toolbox; and the rules of type-2 fuzzy logic system (T2FLS) and type-1 fuzzy logic system (T1FLS) have generated automatically from training data by specifying parameters of the membership of antecedents and consequents of a membership functions using Java Programming. It is, firstly, a new fuzzy logic-based MEP system. How far BBBC can assist designing T2FLS to improve its performance in Malaria epidemic prediction?

The Proposed BBBC Optimized Type-2 Fuzzy Logic-based System for Malaria Epidemic Prediction

The real dataset we used for the experiment is monthly Malaria cases (Morbidity data) and climate data which was collected from 2013 to 2017 throughout the country in Malaria prone areas by the Ministry of Health and National Meteorology Agency of Ethiopia respectively. The number of instances in a dataset is 11282 which was divided 70% for training and the remaining for testing.

The climatic inputs data includes monthly average of temperature (TEMP), rainfall (RF), relative humidity (RH) and Elevation (ELV) collected from ground Meteorology Stations. The other input is a number of monthly lag outpatient Malaria cases (LAGCASE). The output is a number of Malaria cases (CASE). Since these variables have different units, it would be difficult to establish a correlation between them. Thus, the min-max normalization method was applied to limit the value of these variables between 0 and 1. The formula for the normalization is as follows [35]:

$$Normalized(z) = \frac{z_i - z_{\min}}{z_{\max} - z_{\min}} \tag{3}$$

Where z_{min} is the minimum value and z_{max} is the maximum value from the data set Z, Z represents each input and output values.

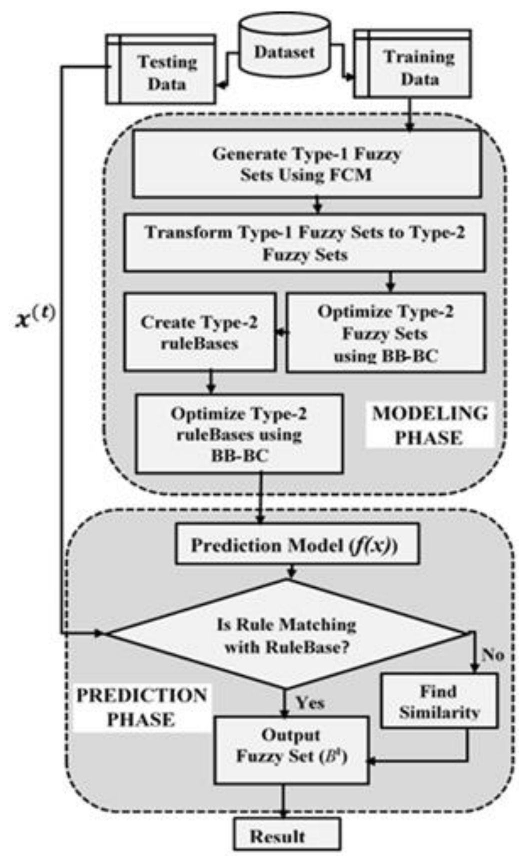

Figure 5: The proposed BBBC optimized type-2 fuzzy logic-based system for MEP

The proposed BBBC optimized type-2 fuzzy logic-based system for MEP works into two phases as shown in Figure 5, a modeling and a prediction phase which are detailed in the following subsections.

Modeling Phase

In this phase, Type-1 Fuzzy Sets extracted using FCM and transformed to Type-2 Fuzzy Sets. Finally, T2FLS optimized using BBBC algorithms.

Step 1-Extracting Fuzzy Sets

The membership function of T1FLS extracted automatically from data using FCM method. FCM has a capability to cluster data into specified number of overlapping groups and each groups considered as membership functions. The FCM clustering algorithm generalized by Bezdek [36]. Fuzzy clustering methods allow the objects to belong to several clusters simultaneously, with different degrees of membership [36].

In this study, each input represented by five Gaussian membership functions such as Very Low, Low, Medium, High and very High by determined Gaussian membership parameters (i.e. mean and standard deviation) dynamically from data using FCM clustering method. The generated fuzzy sets of antecedents and consequents have overlapping domain interval ranges. The extracted samples of type-1 Gaussian membership function for temperature and relative humidity represented in Figures 6 and 7 respectively which could be written as[14]:

$$N(m,\sigma;x) = exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right] \qquad (4)$$

Where m is mean value and σ is standard deviation of the type-1 fuzzy set and x is each input values. A linguistically word was given to each of the resulting fuzzy sets like VLOW and LOW.

Step 2- Transforming Type-1 Fuzzy Sets to Type-2 Fuzzy Sets

The uncertain standard deviation helps to capture the non-stationary behavior of the inputs and output variables. The Gaussian type-2 MFs are used to describe the type-2 fuzzy sets $\tilde{A}_s^q$ (where q=1, 2,…, V and V represents the number of type-2 fuzzy sets for a variable s) with the mathematical definition of equation (5)[14]:

$$\mu_{\tilde{A}_s^q}(x) = exp\left\{-\frac{1}{2}\left(\frac{x-m_s^q}{\tilde{\sigma}_s^q}\right)^2\right\}, \tilde{\sigma}_s^q \in \left[\tilde{\sigma}_{s1}^q, \tilde{\sigma}_{s2}^q\right] \quad (5)$$

Where m_s^q is the value of the center (average) and $\tilde{\sigma}_s^q$ are the values of the standard deviation for each Gaussian interval type-2 MF, for the s^{th} input/output variable.

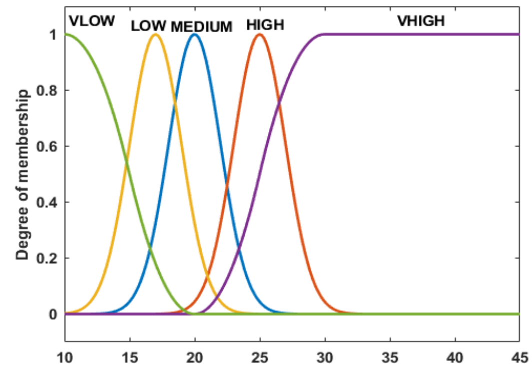

Figure 6: Type-1 membership function for Temperature

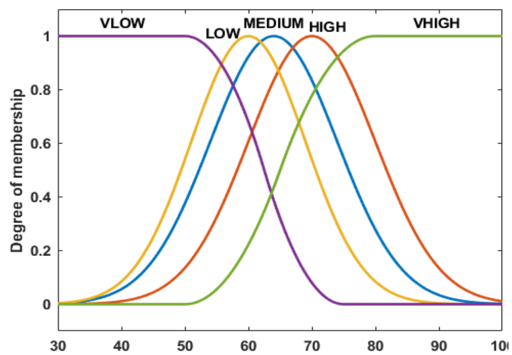

Figure 7: Type-1 membership function for Relative Humidity

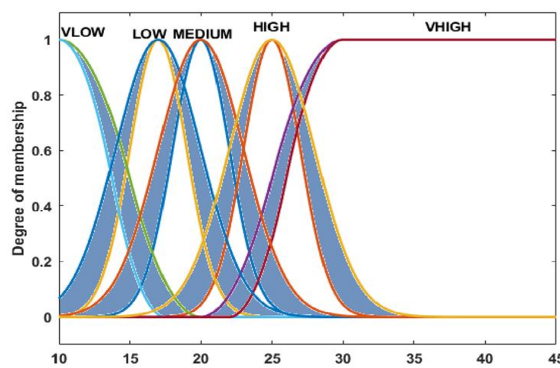

Figure 8: Type-2 Fuzzy Membership for Temperature

In order to construct the initial type-2 MF in our T2FLS, we use certain mean (m) and uncertain standard deviation (σ).

The standard deviation of the given type-1 fuzzy set (extracted by FCM clustering) ,$\tilde{\sigma}_{z1}^j$, used to obtain $\tilde{\sigma}_{z2}^j$ by blurring(including FOU) $\tilde{\sigma}_{z1}^j$ with a certain $\alpha\%$ (α = 10, 20, 30 and 40) factors [37] such that:

$$\tilde{\sigma}_{z2}^j = \tilde{\sigma}_{z1}^j(1+\alpha\%) \qquad (6)$$

Figures 8 and 9 shown the Type-2 fuzzy set for temperature and relative humidity respectively after transformed with 10% FOU.

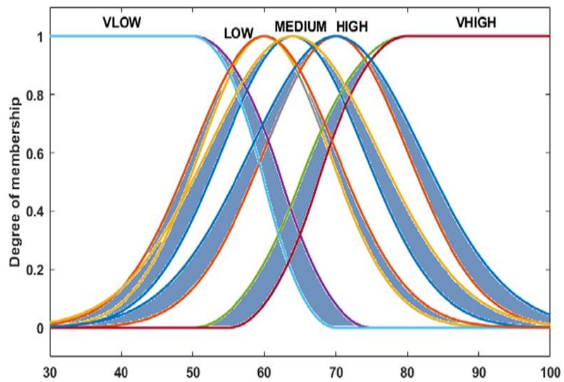

Figure 9: Type-2 Fuzzy Membership for Relative Humidity

Step 3: Optimizing the Type-2 Fuzzy Logic System with BBBC

This section provides how the BBBC was applied to optimize the type-2 fuzzy sets parameters of the type-2 fuzzy logic-based MEP system and how to evaluate the quality and fitness of the achieved set of parameters.

The main purpose of using FCM to generate the membership functions and using the Wang-Mendel method to construct the initial rule base before our BBBC optimization is to obtain a good starting point in the search space, since the BBBC quality of the optimization highly relies on the starting state to converge fast to the optimal position. If we started from random fuzzy sets and rules, the BBBC will take very long time to converge to optimal values[37].

Big Bang-Big Crunch (BBBC) Optimization

The BBBC optimization method developed by Erol and Eksin[31] derived from the theory of the birth of the universe in astrophysics, namely the Big-Bang Big-Crunch Theory. It has two main phases: In the first Big Bang phase, the candidate solutions are randomly distributed over the search space and the next Big Crunch phase where a contraction procedure calculates a center of mass for the population. The key advantages of BBBC are its low computational cost, ease of implementation, and fast convergence[31]. The working principle of the Big Bang phase is explained as energy transformation from an ordered state (a convergent solution) to a disordered or chaotic state (a new set of candidate solutions). After the Big Bang phase, a contraction procedure is applied during the Big Crunch. In this phase, the contraction operator takes the current positions of each candidate solution in the population and its associated cost function value and computes a center of mass according to[31]:

$$x_c = \frac{\sum_{i=1}^{N} \frac{1}{f^i} x_i}{\sum_{i=1}^{N} \frac{1}{f^i}} \tag{7}$$

Where $\boldsymbol{x_c}$ the position of the center of mass, $\boldsymbol{x_i}$ is the position of the candidate, $\boldsymbol{f_i}$ is the cost function value of the $\boldsymbol{i^{th}}$ candidate and $\boldsymbol{N}$ is the population size.

The new generation for the next iteration Big Bang phase is normally distributed around x_c . The new candidates around the center of mass are calculated by adding or subtracting a normal random number whose value decreases as the iterations elapse. This can be formalized as[31]:

$$x^{new} = x_c + \frac{r\alpha(x_{max} - x_{min})}{k} \tag{8}$$

where r is a normal random number, α is a parameter limiting the size of the search space, x_{max} and x_{min} are the upper and lower limits respectively, and k is the iteration step.

The procedure of the BBBC optimization is given as follow:

Step 1 (Big Bang Phase)

The generation of N initial candidates randomly in the search space.

Step 2

The cost function values of all the candidate solutions are computed.

Step 3 (Big Crunch Phase)

The center of mass is calculated. Either the best fit individual or the center of mass is chosen as the point of Big Bang Phase.

Step 4

New candidates are calculated around the new point calculated in Step 3 by adding or subtracting a random number whose value decreases as the iterations elapse.

Step 5

Return to Step 2 until stopping criteria has been met.

Since normally distributed numbers can be exceeding ± 1, it is necessary to limit the population to the prescribed search space boundaries. This narrowing down restricts the candidate solutions into the search space boundaries[31].

Optimizing the Rule Base of the T2FLS with BBBC

The fuzzy rules are defined for the n dimensional pattern problem as follows[14][38]:

$$R_l: IF\ x_1\ is\ A_1^l, \dots, x_n\ is\ A_n^l\ THEN\ y_1\ is\ B_1^l \quad (9)$$

Where $l = (1,2,3,\dots,M), M$ is the total number of rules and i is the index of the rules and $x = (x_1, x_2, \dots, x_n)^T$ is an n-dimensional pattern vector.

To optimize the rule base of the T2FLS, the parameters of the rule base are encoded into a form of a population. The T2FLS rule base can be represented as shown in Figure 10.

m_1^1	...	m_p^1	b_1^1	...	b_q^1	...	m_1^R	...	m_p^R	b_1^R	...	b_q^R

Figure 10: The population representation for the parameters of the rule base

Based on Figure 10, m_x^r are the antecedents and b_k^r is the consequents of each rule respectively, where $x = 1, \dots, p,\ p$ is the number of antecedents; $k = 1, \dots, q, q$ is the number of consequents; $r = 1, \dots, R$ and R is the number of the rules to be tuned. However, the values describing the rule base are discrete integers while the original BBBC supports continuous values. Thus, instead of Equation (8), the following Equation (10) is used in the BBBC paradigm to round off the continuous values to the nearest discrete integer values modelling the indexes of the fuzzy set of the antecedents or consequents.

$$D^{new} = D_c + round\left[\frac{\gamma\rho(D_{max} - D_{min})}{k}\right] \quad (10)$$

Where D_c is the fittest individual, r is a random number, ρ is a parameter limiting search space, D_{max} and D_{min} are lower and upper bounds, and k is the iteration step. in this study, the rule base constructed by the Wang-Mendel approach[14][39] is used as the initial generation of candidates. After that, the rule base can be tuned by BBBC using the cost function depicted in Equation (10).

Optimizing the Type-2 membership functions with BBBC

In order to apply BB-BC, the feature parameters of the type-2 membership function have to be encoded into a form of a population. As depicted in Equation (6), in order to construct the type-2 MFs, the parameter α has to be determined to obtain σ_{s2}^l while σ_{s1}^l is provided by FCM. To be more accurate, the uncertainty factors α_k^j for each fuzzy set of the MFs are computed, where $k = 1, \dots, p,\ p$ is the number of antecedents; $j = 1, \dots, q, q$ is the number of inputs/features. For illustration purposes, as in the MFs of the proposed system, five type-2 fuzzy sets including VERY LOW, LOW, MEDIUM, HIGH and VERY HIGH are utilized for modeling each of the five inputs , therefore, the total number of the parameters for the input type-2 MFs is 5X5=25. In a similar way, parameters for the output MFs are also encoded which has the same type of five MFs like input variable. Therefore, the structure of the population is built as displayed in Figure 11.

α_1^1	...	α_5^1	α_1^2	...	α_5^2	α_1^5	...	α_5^5	α_1^{opt}	...	α_5^{opt}

Figure 11: The population representation for the parameters of the type-2 fuzzy sets

The optimization problem is a minimization task, with the parameters encoded as above, the cost function in Equation (11) is minimized. The model's performance is evaluated by comparing the model output with actual values of the testing data set using RMSE (Root Mean Square Error) in error minimization from the correct value[40]:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2} \quad (11)$$

Where N is the number of actual historical data, y_t is the actual value and $\hat{y}_t$ is the corresponding predicted value for $t = 1$ to N. Within the context of this study finding the best model means obtaining the predicted values that would give the least RMSE result in MEP.

Prediction Phase

In our fuzzy system, the antecedents are ELV, RF, RH, TEMP and LAGCASE and the system output is the CASE are modelled by five fuzzy sets: VERY LOW, LOW, MEDIUM, HIGH and VERY HIGH. The type-1 fuzzy sets are obtained via FCM and the rules are the same as the T2FLS.

When an input pattern is introduced to the generated model, two cases will happen. The first case is when the input $x^{(t)}$ matches any of the M rules in the generated model, we will obtain the output fuzzy set, $B^{(l)}$ [12] [38]. In the second case, If $x^{(t)}$ does not match any of the existing M rules, we will calculate the similarity (or distances) between each of the fuzzy rules generated by $x^{(t)}$ and each of the M rules stored in the rule base and select the output of the similar rule, $B^{(l)}$ [12][29].

RESULTS AND DISCUSSIONS

In this study, we aimed to predict Malaria epidemic which is the number of Malaria cases in specific place ahead of one up to three months with less false alarm. With a consideration of this aim, we have developed BBBC optimized Type-2 Fuzzy Logic-based Malaria Prediction System and its result compared with black-box model which is Artificial Neural Network (ANN) and Coordination of Adaptive Neuro Fuzzy Inference System (ANFIS) as benchmark. The idea is to monitor sensible outputs and analyze any significant deviation. In order to conduct a fair comparison, all the methods share the same input features and experiment setup.

ANN is the common predictive models but the internal logic is unexplainable and incomprehensible. ANN is black box technique in the sense that while it approximate any function, studying its structure would not give you any insights on the structure of the function being approximated[41]. ANN model training was performed using MATLAB R2019a built-in Toolbox.

With ANN model development, a feed forward three-layer back propagation network was chosen. A Levenberg–Marquardt algorithm for back propagation with a gradient descent and momentum weight and bias learning function, was used to train the network. The error minimization of ANN model with testing data is 0.0942 in RMSE and the accuracy is 90.58% shown in Table 1.

Some sample generated rules by Type-2 Fuzzy Logic System are shown below:

R1: If ELV_VLOW and RF_VLOW and RH_VLOW and TEMP_HIGH and LAGCASE_VLOW Then CASE_VLOW

R2: If ELV_VLOW and RF_VLOW and RH_LOW and TEMP_MEDIUM and LAGCASE_HIGH Then CASE_MEDIUM

R3: If ELV_MEDIUM and RF_HIGH and RH_HIGH and TEMP_VLOW and LAGCASE_HIGH Then CASE_HIGH

Our developed model testing results displayed Table 1 and Figure 12 in term of accuracy, and errors minimization also shown in Figure 13. Based on these results, BBBC via T2FLS outperform T1FLS, non-optimized T2FLS and ANFIS with different percent. On the other hand, optimized T2FLS slightly less than in accuracy to compare black-box model ANN for Malaria prediction.

In Table 1 and on Figure 12, the results of model testing accuracy of T1FLS, T2FLS, ANFIS, ANN, and BBBC optimized T2FLS represented. Hence, we can see that the results 87.85% for BBBC optimized T2FLS outperforms the accuracy of 85.10% for T1FLS (i.e. uplift of 2.75%), the accuracy of 86.66% with ANFIS[21] (i.e. uplift of 1.19 %) and the accuracy of 86.99% non-optimized T2FLS (i.e. Uplift by 0.86%). This result is due to the optimization of membership functions with the appropriate range using the BBBC algorithm. However, the ANN model has an accuracy of 90.58% which is better than BBBC optimized T2FLS by 2.73%.

Even if BBBC via T2FLS has less accuracy to compare with ANN, It has a capability to represent a solution in the form of rules like Equation (9) which is close to natural language for easy understanding and interpreting by layman users when we perform Malaria epidemic prediction. In addition, the BBBC optimized T2FLS can have an optimal MEP accuracy. It gives a better interpretable model as a result of using a smaller number of rules in comparison to the T1FLS and non-optimized T2FLS.

TABLE 1: ACCURACY RESULT OF DIFFERENT TECHNIQUES

T1FLS	T2FLS	ANFIS	ANN	BBBC Via T2FLS
85.10%	86.99%	86.66%	90.58%	87.85%

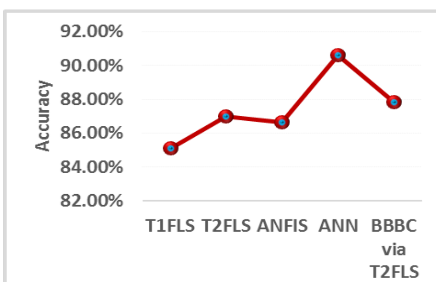

Figure 12: T1FLS, T2FLS, ANFIS and ANN performance with Testing

The error minimization of T1FLS, T2FLS and BBBC optimized T2FLS with testing sample data displayed in Figure 13. The error value of T2FLS with BBBC close to 0 than others (T1FLS and T2FLS) which means it has better accuracy and will have less false alarm in Malaria epidemic prediction.

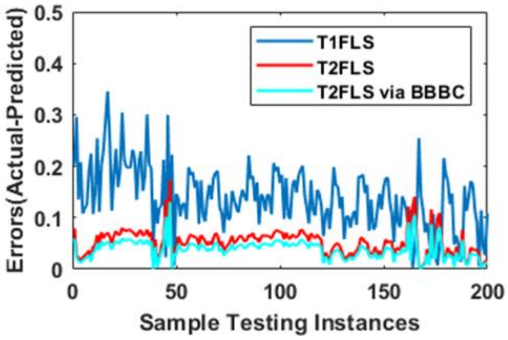

Figure 13: Error minimization by T1FLS, T2FLS and T2FLS via BBBC

The main importance of this Fuzzy Logic-based System is that it can be explained and interpreted by the main stakeholders and decision-makers. We can use the fuzzy IF-THEN rules to solve the MEP problem and understand the Malaria occurrence pattern from generated rules. For example, there are generated rules below which can be justified and used by health officers and experts for MEP in Ethiopia based on climatic factors and lag Malaria cases. Generated rules with High and Very High outputs are more interesting which are close to the Malaria epidemic. For example:

R3: If ELV_HIGH and RF_LOW and RH_LOW and TEMP_HIGH and LAGCASE_HIGH Then CASE_HIGH.

From the above rule, we can understand that the Malaria case occurrence will be high at a high level of elevation when the temperature becomes high (i.e. hot) even if a high level of elevation in Ethiopia was considered a Malaria low or free zone by experts. This might be climatic factors change due to global warming. Another example can be found in the following rule:

R4: If ELV_LOW and RF_HIGH and RH_MEDIUM and TEMP_MEDIUM and LAGCASE_MEDIUM Then CASE_HIGH.

We can infer from the above rule that the number of Malaria infected persons will be high (i.e. CASE_HIGH) after one and half months since the current number of Malaria cases is medium (i.e. LAGCASE_MEDIUM) and the amount of relative humidity (i.e. RH_MEDIUM) and temperature (i.e. TEMP_MEDIUM) are Medium. This is might be due to high amount of rainfall (i.e. RF_HIGH) creates water pools which lead to hatching a large number of Anopheles mosquitos (i.e. Malaria parasite vectors); and a medium level of temperature and relative humidity are a favorable condition for Plasmodium parasite growth inside the mosquito which is the cause of Malaria disease.

The proposed model can be used in the real situation for the prediction of Malaria epidemic ahead up to three months in Ethiopia by health workers and decision-makers to get control Malaria morbidity and deaths and try to take preventive measures based on the facts showing by the extracted rules.

CONCLUSIONS

Malaria is a life-threatening disease which has a huge medical, economic, and social impact. Diversity of modeling technique have been used to develop predictive models of malaria, no work has made use of Fuzzy Logic-based System. In this paper, we presented a BBBC optimized type-2 fuzzy logic-based Malaria epidemic prediction system which has a capability to predict up to three months ahead. Based on the testing result showed in Table 1 and Figure 12, BBBC optimized T2FLS outperforms its counterpart T1FLS, non-optimized T2FLS and ANFIS. This result is due to the optimization of membership functions with the appropriate range using the BBBC algorithm. On the other way, the BBBC optimized T2FLS prediction accuracy slightly less than to compare with ANN prediction accuracy. However, the proposed T2FLS via BBBC optimization gave an opportunity to explain the cause of why the epidemic happened by investigated the amount and intensity of input factors using IF-THEN rules. This system has the capability of handling the encountered uncertainties and generating transparent result from a pre-specified number of linguistic rules, which enables the user to easily understand and analyze the generated Malaria epidemic prediction results. The proposed BBBC optimized T2FLS is able to predict Malaria occurrence with sufficient leading time on specific places by using the current climatic factor values and Malaria cases.

Hence, the decision-makers will use as warning system to prevent and control Malaria disease early before it leads to human morbidity and death. In future work, additional input factors will include for modeling and the proposed T2FLS based system of Malaria prediction apply in the real-world intelligent system to develop applications for serving the users and scale-up to other diseases which have similar behaviors.

References

[1] D. G. Ayele, T. T. Zewotir, and H. G. Mwambi, (2012) "Prevalence and risk factors of malaria in Ethiopia " Malaria Journal 11: 1–9

[2] World Health Organization,(2019) "World Malaria Report 2019," Geneva.

[3] H. S. Taffese et al., (2018) "Malaria epidemiology and interventions in Ethiopia from 2001 to 2016 " Infect. Dis. Poverty 7: 1–9.

[4] WHO, (2017) "country profile, Ethiopia 2017"

[5] T. A. Abeku, S. J. De Vlas, G. Borsboom, A. Teklehaimanot, and A. Kebede (2002) "Forecasting malaria incidence from historical morbidity patterns in epidemic-prone areas of Ethiopia : a simple seasonal adjustment method performs best " Trop. Med. Int. Heal 7: 851–857.

[6] A. E. Kiszewski and A. Teklehaimanot (2004) "A review of the clinical and epidemiologic burdens of epidemic malaria " Am. J. Trop. Med. Hyg.71: 128–135.

[7] G. Zhou, N. Minakawa, A. K. Githeko, and G. Yan (2004) "Association between climate variability and malaria epidemics in the East African highlands " PNAS 101:2375–2380

[8] L. A. Zadeh (1965) "Fuzzy Sets," Inf. Control 8:338–353.

[9] E. H. Mamdani (1974) "Application of Fuzzy Algorithms for Control of Simple Dynamic Plant." Proc. Inst. Electr. Eng.121:1585–1588.

[10] D. Bernardo, H. Hagras, and E. Tsang (2013) "A genetic type-2 fuzzy logic-based system for the generation of summarised linguistic predictive models for financial applications " Soft Comput 17:2185–2201.

[11] C. Wagner and H. Hagras (2010) "Toward general type-2 fuzzy logic systems based on zSlices," IEEE Trans. Fuzzy Syst 18: 637–660.

[12] B. Chekol and H. Hagras (2020) "A Type-2 Fuzzy Logic-based System for Malaria Epidemic Prediction in Ethiopia " SUST J. Eng. Comput. Sci.21:42-54.

[13] L.A.Zadeh (1975) "The Concept of a Linguistic Variable and its Application to Approxirmate Reasoning-III* " Inf. Sci. (Ny).9:43–80.

[14] J. Mendel (2017) Uncertain Rule-Based Fuzzy Systems. Springer International Publishing, Switzerland

[15] D. Wu and J. M. Mendel (2014) "Enhanced Karnik-Mendel Algorithms for Interval Type-2 Fuzzy Sets and Enhanced Karnik-Mendel Algorithms for Interval Type-2 Fuzzy Sets and Systems " IEEE Xplore.

[16] M. M. Kifle, T. T. Teklemariam, A. M. Teweldeberhan, E. H. Tesfamariam, A. K. Andegiorgish, and E. Azaria Kidane (2019) "Malaria Risk Stratification and Modeling the Effect of Rainfall on Malaria Incidence in Eritrea" J. Environ. Public Health: 1–11.

[17] A. Midekisa, G. Senay, G. M. Henebry, P. Semuniguse, and M. C. Wimberly (2012) "Remote sensing-based time series models for malaria early warning in the highlands of Ethiopia " Malaria Journal 11:1–10.

[18] H. D. Teklehaimanot, M. Lipsitch, A. Teklehaimanot, and J. Schwartz (2004) "Weather-based prediction of Plasmodium falciparum malaria in epidemic-prone regions of Ethiopia I . Patterns of lagged weather effects reflect biological mechanisms " Malaria. Journal 3:1–11.

[19] H. H. Hussien, F. H. Eissa, and K. E. Awadalla, "Statistical Methods for Predicting Malaria Incidences Using Data from Sudan (2017) " Malar. Res. Treat.:1–9.

[20] E. L. Darkoh, J. A. Larbi, and E. A. Lawer (2017) "A weather-based prediction model of Malaria prevalence in Amenfi West District, Ghana" Malar. Res. Treat.

[21] B. E. Chekol and H. Hagras (2018) "Employing Machine Learning Techniques for the Malaria Epidemic Prediction in Ethiopia " in 2018 10th Computer Science and Electronic Engineering Conference, CEEC 2018 - Proceedings, 2019: 89–94.

[22] P. Haddawy et al. (2018) "Spatiotemporal Bayesian networks for malaria prediction" Artif. Intell. Med. 84:127–138.

[23] T. Santosh and D. Ramesh (2019) "Artificial neural network based prediction of malaria abundances using big data: A knowledge capturing approach" Clin. Epidemiol. Glob. Heal.7:121–126.

[24] C. Sudheer et al. (2014) "A Support Vector Machine-Firefly Algorithm based forecasting model to determine malaria transmission," Neurocomputing 129:279–288.

[25] A. J. Starkey, H. Hagras, S. Shakya, and G. Owusu (2018) "A Genetic Algorithm Based System for Simultaneous Optimisation of Workforce Skills and Teams" KI - Künstliche Intelligenz 32:245–260.

[26] C. Wagner and H. Hagras (2007) "A genetic algorithm based architecture for evolving type-2 fuzzy logic controllers for real world autonomous mobile robots" in IEEE International Conference on Fuzzy Systems.

[27] A. Starkey, H. Hagras, S. Shakya, and G. Owusu (2016) "A multi-objective genetic type-2 fuzzy logic-based system for mobile field workforce area optimization," Inf. Sci. (Ny).329: 390–411.

[28] D. Bernardo, H. Hagras, and E. Tsang (2013) "A Genetic type-2 fuzzy logic-based system for financial

applications modelling and prediction " IEEE Int. Conf. Fuzzy Syst(FUZZY-IEEE).

[29] R. Chimatapu, H. Hagras, A. Starkey, and G. Owusu (2018) "A big-bang big-crunch type-2 fuzzy logic system for generating interpretable models in workforce optimization" IEEE Int. Conf. Fuzzy Syst(FUZZY-IEEE).

[30] S. Naim and H. Hagras (2015) "A Big-Bang Big-Crunch Optimized General Type-2 Fuzzy Logic Approach for Multi-Criteria Group Decision Making " J. Artif. Intell. Soft Comput. Res. 3:117–132.

[31] O. K. Erol and I. Eksin (2006) "A new optimization method: Big Bang-Big Crunch," Adv. Eng. Softw.37:106–111.

[32] M. A. Bo Yao, Hani Hagras (2016) "A Big Bang-Big Crunch Type-2 Fuzzy Logic System for Machine Vision-Based Event Detection and Summarization in Real-world Ambient Assisted Living " IEEE.

[33] L. A. O. Castillo and L. Amador-angulo (2016) "A new fuzzy bee colony optimization with dynamic adaptation of parameters using interval type-2 fuzzy logic for tuning fuzzy controller " Soft Comput 22:571–594.

[34] O. Castillo, R. Martínez-marroquín, and P. Melin (2010) "Bio-inspired Optimization of Fuzzy Logic Controllers for Robotic Autonomous Systems with PSO and ACO " Fuzzy Inf. Eng.: 119–143.

[35] E. Ogasawara, L. C. Martinez, D. De Oliveira, G. Zimbrão, G. L. Pappa, and M. Mattoso (2010) "Adaptive Normalization: A novel data normalization approach for non-stationary time series" in The 2010 International Joint Conference on Neural Networks (IJCNN).

[36] B. James, E. Robert, And F. William (1984) "FCM : THE FUZZY c-MEANS CLUSTERING ALGORITHM " Comput. Geosci. 10:191–203.

[37] B. Yao, H. Hagras, D. Alghazzawi, S. Member, and J. Mohammed (2016) "A Big Bang–Big Crunch Type-2 Fuzzy Logic System for Machine-Vision-Based Event Detection and Summarization in Real-World Ambient-Assisted Living" IEEE Trans. Fuzzy Syst.24:1307–1319.

[38] F. Doctor, H. Hagras, and V. Callaghan (2006) "An Incremental Adaptive Life Long Learning Approach for Type-2 Fuzzy Embedded Agents in Ambient Intelligent Environments," IEEE Int. Conf. Fuzzy Syst: 915–922.

[39] L. X. Wang (2003) "The WM Method Completed: A Flexible Fuzzy System Approach to Data Mining " IEEE Trans. Fuzzy Syst.11: 768–782.

[40] R. J. Hyndman and A. B. Koehler (2006) "Another look at measures of forecast accuracy" Int. J. Forecast.22: 679–688.

[41] M. Yao, B. Vuthaluru, O. Tadé, and D. Djukanovic (2005) "Artificial neural network-based prediction of hydrogen content of coal in power station boilers," Fuel 84: 1535–1545.