SUST Journal of Engineering and Computer Sciences (JECS), Vol. 21, No. 1, 2020

A Type-2 Fuzzy Logic Based System for Malaria Epidemic Prediction in Ethiopia

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

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

Received:23/11/2019
Accepted:19/01/2020

ABSTRACT- Malaria is the most prevalent mosquito-borne disease throughout tropical and subtropical regions of the world with severe medical, economic, and social impact. Malaria is 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 approaches, nevertheless, they had no learning capabilities. In this paper, we present a Type-2 Fuzzy Logic Based System for Malaria epidemic prediction in Ethiopia which was trained using real data collected throughout Ethiopia from 2013 to 2017. Fuzzy Logic Based Systems provide a transparent model which employs IF-Then rules for the prediction that could be easily analyzed and interpreted by decision-makers. This is quite important to fight the sources of Malaria and take the needed preventive measures where the generated rules from our system were able to explain the situations and intensity of input factors which contributed to Malaria epidemic incidence up to three months ahead. The presented Type-2 Fuzzy Logic System (T2FLS) learns its rules and fuzzy set parameters from data and was able to outperform its counterparts T1FLS in 2% and ANFIS in 0.33% in the accuracy of prediction of Malaria epidemic in Ethiopia. In addition, the proposed system did shed light on the main causes behind such outbreaks in Ethiopia because of its high level of interpretability.

Keywords*: Type-2 fuzzy logic system, Fuzzy C-means, malaria prediction, machine learning.*

المستخلص - الملاريا هي أكثر الأمراض التي تنتقل عن طريق البعوض في جميع أنحاء المناطق المدارية وشبه المدارية في العالم ولها تأثير طبي واقتصادي واجتماعي شديد. تعد الملاريا مشكلة خطيرة في مجال الصحة العامة في إثيوبيا منذ عام 1959، تم تخفيض معدلات المرضى والوفيات ابتداءً من عام 2001. وقد أجريت دراسات مختلفة للتنبؤ بمرض الملاريا باستخدام مناهج رياضية وإحصائية. في هذه الورقة، نقدم نظام مبني على المنطق من النوع 2 غامض للتنبؤ بمرض الملاريا في إثيوبيا والذي تم تدريبه باستخدام البيانات الحقيقية التي تم جمعها في جميع أنحاء إثيوبيا 2013-2017. توفر النظم المبنية على المنطق المبهم نموذجًا شفافًا يستخدم قواعد IF-Then لقواعد التنبؤ الذي يمكن تحليله بسهولة وتفسيره من قبل صناع القرار. هذا مهم للغاية لمحاربة مصادر الملاريا واتخاذ التدابير الوقائية اللازمة حيث تمكنت القواعد المنبثقة من نظامنا من شرح حالات وكثافة عوامل المدخلات التي ساهمت في حدوث مرض الملاريا حتى ثلاثة أشهر قادمة. يتعلم نظام المنطق الضبابي من النوع الثاني (T2FLS) قواعده ومعاملات المجموعة غامض fuzzy set من البيانات، وكان قادرا على التفوق على نظرائه في دقة التنبؤ بمرض الملاريا في إثيوبيا حيث تفوق على T1FLS ب 2 ٪ و ANFIS ب 0.33 ٪. بالإضافة إلى ذلك ، ألقى النظام المقترح الضوء على الأسباب الرئيسية وراء تفشي المرض في إثيوبيا بسبب ارتفاع مستوى التفسير المبني على النظام.

INTRODUCTION

Malaria is the most prevalent mosquito-borne disease throughout tropical and subtropical regions of the world with huge medical, economic, and social impact. It is caused by protozoan parasites of the genus Plasmodium [1]. In 2017, an estimated 219 million cases of Malaria occurred worldwide and most malaria cases were in the WHO African Region (200 million equivalents to 92%), followed by the WHO South-East Asia Region with 5% of the cases and the WHO Eastern Mediterranean Region with 2%.

There were an estimated 435, 000 deaths from Malaria globally and the WHO African Region accounted for 93% of all malaria deaths in 2017 [2]. According to WHO 2018 report [2], an estimated US$ 3.1 billion was invested in Malaria control and elimination efforts globally by governments of Malaria endemic countries and international partners in 2017. Malaria continues to be a serious public health problem in Sub-Saharan Africa despite efforts that have been made to prevent and control the disease for many decades. Malaria is a leading public health problem in Ethiopia since 1959, even if, its morbidity and mortality have been reduced from 2001 [3, 4].

It is estimated that about 75% of the total area of the country and 68% of the population is living in malarious area [4]. Ethiopia has varying topographical and favorable climatic feature to disease-transmitting vector and parasite development, including elevation, rainfall, and temperature [5]. Figure 1, shows the trend of Malaria cases from 2010 to 2017 and the incidence is still high in Ethiopia.

Different scholars tried to show the importance of climate variables and morbidity data to predict malaria transmission and early detection [5]. However, a relationship between climate, seasonal parasite transmission, and disease outcomes are complex and have been poorly defined for many years [6]. Temperature, rainfall, and humidity are key determinants for malaria incidence and significant change in climate variability has coincided with increased magnitude and frequency of malaria epidemics [7].

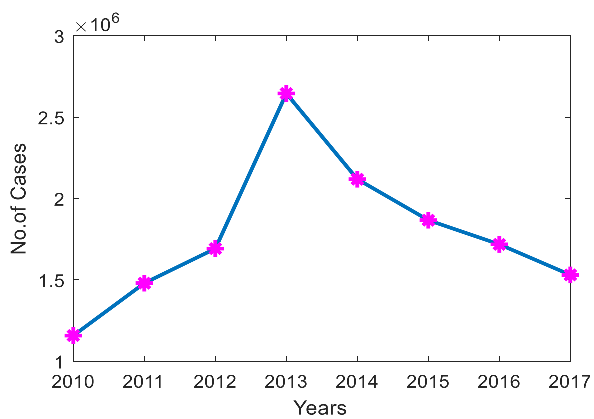

Figure 1: Number of malaria cases per year

The interaction between lag climatic factors and their biological influence on mosquito and parasite life cycle is a vital factor in the association between weather and malaria. The lag pattern for meteorological factors should be considered in the development of malaria incidence prediction and early warning system [8, 9].

The main objective of this paper is to develop intelligent fuzzy logic-based system i.e. type-2 and type-1 Fuzzy Logic Systems (FLSs) to get optimal prediction of Malaria epidemic ahead of time. Further, the results of T1FLS and T2FLS compared with ANFIS.

The rules of T1FLS and T2FLS were generated automatically from training data by specifying parameters of the membership functions of antecedents and consequents of linguistic level using Fuzzy C-Means Clustering techniques.

The paper is organized as follows. In section II reviewed related works to this paper. Section III provides a brief overview on T1FLS, T2FLS, and Fuzzy C-Means Clustering. Section IV introduces the presented type-2 fuzzy logic based system for Malaria prediction. Section V presents the experiments and results. Finally, the conclusions and future work are summarized in Section VI.

Related work

In areas with unstable Malaria transmission, various studies conducted in the past to predict malaria epidemic and transmission by applying statistical and mathematical techniques using malaria cases and meteorological data. For example, one of the common statistical approaches for Malaria incidence prediction is Auto-Regressive Integrated Moving Average (ARIMA) which is applied to analyze meteorological factors link with malaria appearance [10] and also Seasonal Autoregressive Integrated Moving Average (SARIMA) model developed to predict and forecast monthly malaria incidences based on environmental variables and monthly malaria incidence data[11,12]. Seasonal adjustment methods were also applied to forecast Malaria incidence from historical morbidity data alone with unstable transmission [13]. Polynomial Distributed Lag Models (PDL) were applied to determine the effect of weather factors and their lag distribution on malaria incidence in Ethiopia for the years 1990 to 2000 [8].

However, the traditional statistical and mathematical approach like ARIMA, seasonal adjustment and others techniques needs the data to be stationary overtime and they assumed linearity of data even though the meteorological variables were less likely linearly correlated. In addition, they did not learn the situation from historical data to capture trends properly and they are not accurate enough for forecasting abnormal incidence of Malaria for use in an epidemic early warning system. In addition to statistical regression and mathematical techniques, recent years have seen the application of Artificial Intelligence techniques to develop predictive models for malaria prediction.

For example, Support Vector Regression (SVR) and an Adaptive Neuro Fuzzy Inference System (ANFIS) were proposed to predict the Malaria epidemic up to three months ahead using monthly climate and malaria cases data [14]. R. Rismala *et al* [15] employed Evolving Neural Network (ENN) for Malaria incidence prediction based on environmental factors and the findings proved that there was a sufficient correlation between weather and malaria incidence. Santosh and Ramesh [16] determined malaria abundances using clinical and environmental variables by applying Artificial Neural Network (ANN). Another learning technique is a coupling of Firefly Algorithm (FFA) and Support Vector Machines (SVM) to forecast the malaria incidences using monthly averages of rainfall, temperature, relative humidity and malarial incidences where the SVM-FFA model provided more accurate forecasts compared to Artificial Neural Networks (ANN) and Auto-Regressive Moving Average [17]

Even though such machine learning approaches give good prediction accuracies, they lack being able to handle uncertainties and noises in malaria incidence and climatic data. In addition, such advanced machine learning techniques are black-box and they have limitations to represent in more interpretable and reasonable form the results for decision making so that the models can be less understandable and transparent for normal use and trust. There exist various white-box transparent models, one of these models is Fuzzy Logic Systems (FLS).

FLS provides white-box models which could be easily analyzed and understood by the layman user [18, 19]. Fuzzy Logic concepts have been applied with success in many real-world applications. For example, FLS was applied to predict the house sale price [20] and T2FLS applied in the prediction of the correlation of pressure-volume–temperature (PVT) properties of crude oil [21] and modeling data uncertainty on electric load forecasting [22]. In addition, T2FLS successfully applied in financial summarizations [23]. The association between malaria and meteorological factors is complex due to the lagged and non-linear pattern and this drove the need to build transparent and robust frameworks like Fuzzy Logic System for predicting and assessing risks of Malaria epidemic.

Overview of Type-1 and Type-2 Fuzzy Logic System and Fuzzy C-Means Clustering

Type-1 Fuzzy Logic System

Fuzzy sets were introduced by Zadeh [24] as a means of representing and manipulating data that was not precise, but rather fuzzy. Fuzzy sets and systems have been evolved through time and accepted as a methodology for building systems that can deliver satisfactory performance in the face of uncertainty and imprecision inputs [25]. Conventional Fuzzy Logic systems (i.e. Type-1 Fuzzy Logic Systems) provide a means for calculating intermediate values between the range between 0 and 1 [26]. 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 the membership degree of an element to the set and are defined by the following Equation [27]:

$$A = \{(x, \mu_A(x)) \mid x \in X\} \tag{1}$$

where $\mu_A : x \to [0,1]$. In this definition, $\mu_A(x)$ represents the membership degree of the element $x \in X$ to the set A. Type-1 and Type-2 fuzzy set can be represented using different membership functions such as triangle, trapezoid, and Gaussian. For example, Figure 2 shows an example of type-1 fuzzy set and type-2 fuzzy set represented by Gaussian membership function.

The Footprint of Uncertainty (FOU) for an interval type-2 fuzzy set is created with uncertain standard deviation of the Gaussian Membership Function. The uncertain standard deviation helps to capture the non-stationary behavior of the targets and inputs.

The support of FOU is formed by Equations (2) and (3), where $\overline{\mu}(x)$ and $\underline{\mu}(x)$ are upper and lower membership functions, respectively, $x \in X$ and X is the Universe of discourse, σ_1 and σ_2 are the standard deviations of two type-1 membership functions used to construct the general type-2 fuzzy logic system support, $\sigma_1 < \sigma_2$ and m is the mean used by both type-1 membership functions[27].

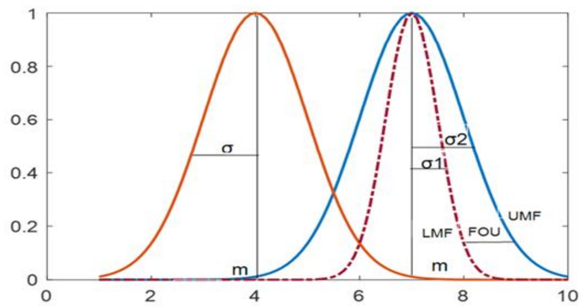

Figure 2: Type-1 and Type-2 Gaussian membership functions

$$\underline{\mu}(x) = \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma_1}\right)^2\right] \quad , \qquad (2)$$

$$\bar{\mu}(x) = \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma_2}\right)^2\right] \qquad (3)$$

The structure of type-1 fuzzy logic system displayed in Figure 3 [28] with four main components.

An input is applied to a type-1 FLS, the inference engine computes the type-1 output set corresponding to each rule. The defuzzifer then computes a crisp output from these rule output sets.

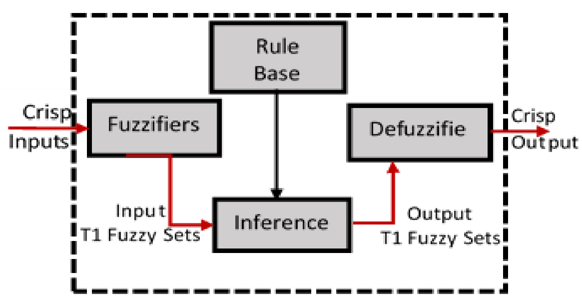

Figure 3: Type-1 FLS Structure

However, the real-world applications are often faced with multiple sources and high levels of uncertainty but type-1 fuzzy sets can only handle a limited level of uncertainty [27].

T1FLS are unable to directly handle rule uncertainties because they use type-1 fuzzy sets that are certain. On the other hand, type-2 FLS are useful in circumstances where it is difficult to determine an exact numeric membership function, and there are measurement uncertainties.

Type-2 Fuzzy Logic System

The concept of type-2 fuzzy logic system was introduced by Zadeh [28] as an extension of type-1 fuzzy logic system by considering the limitation of type-1 fuzzy sets. 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 the membership grade is a crisp number in [0, 1]. Type-2 Fuzzy Logic Systems are useful in circumstances when the training data is affected by noise and uncertainty available in the membership grades themselves; hence, they are useful for incorporating uncertainties [29].

A fuzzy set A on a universe of discourse X is characterized by a membership function $\mu_A : x \to [0,1]$, and is expressed as follow in Equation (4) [27]:

$$A = \{(x, \mu_A(x)) \mid \mu_A(x) \in [0,1] \forall x \in X\}. \qquad (4)$$

The primary membership grade of a type-2 FLS is a Type-1 Fuzzy Set in [0, 1] and the secondary membership is a crisp number in [0, 1] [10]. The secondary membership function and the range of uncertainty are decided by the third dimension of type-2 fuzzy sets and footprint-of-uncertainty (FOU), respectively.

The structure of a type-2 FLS is similar to the structure of a type-1 FLS as shown Figure 4, except type reducer in a type-2 FLS.

In a type-2 FLS, crisp inputs are first fuzzified, usually into input type-2 fuzzy sets. These activate the inference engine and the rule base to produce output type-2 fuzzy sets. They are then processed by the type-reducer, which combines the output sets and performs a centroid calculation, leading to type-1 fuzzy sets known as type-reduced set(s). The defuzzifier can then defuzzify the type-reduced type-1 fuzzy outputs to produce crisp outputs. The type-reduced set of a type-2 FLS shows the possible variation in the crisp output of the FLS due to uncertain natures of the antecedents and/or consequents [28].

However, General type-2 FLSs are computationally intensive due to type-reduction complexity. Things simplify a lot when secondary membership functions (MFs) are interval sets [27, 30]. Interval type-2 fuzzy logic systems (IT2 FLS) employed as a special case of a general T2FS where all the secondary membership grades are equal to one.

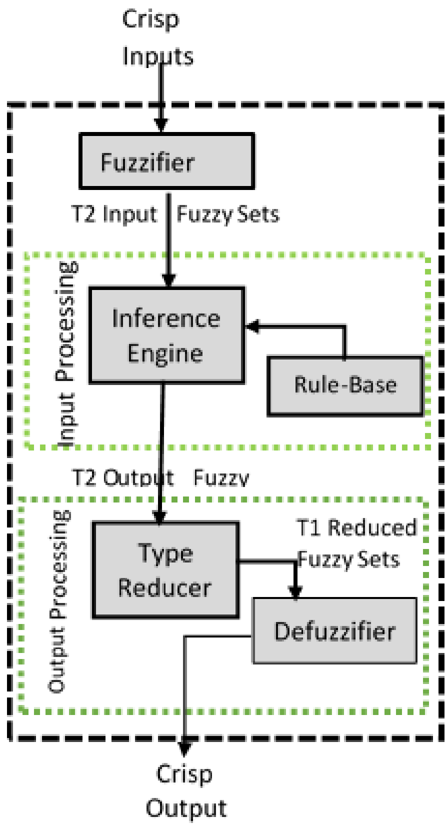

Figure 4: Type-2 FLS structure

The membership functions of IT2FLS are not crisp value rather type-1 interval and it can able to handle uncertainties as shown in Figure 5 and characterized as follow [27]:

$$\tilde{A} = \int_{x \in X} \left[\int_{u \in J_x} 1/u \right] \Big/ x\,,\; J_x \subseteq [0,1] \qquad (5)$$

where the secondary grades of $\tilde{A}$ all are equal 1, x is the primary variable and its measurement domain denoted by X ; u is the secondary domain variable $x \in X$, J_x is the primary membership of x . The footprint of uncertainty (FOU) in IT2FLS is the union of all the embedded T1FS which is marked by the grey area in Figure 5. This grey area is bounded by an upper membership function (UMF) and a lower membership function (LMF) [31]. The FOU of an IT2FSs described using Equation (6) in which the UMF and LMF of the IT2FS are two T1 MFs. The UMF of the IT2FS $\tilde{A}$ is denoted by $\bar{\mu}_{\tilde{A}}(x)$ while it's LMF denoted by $\underline{\mu}_{\tilde{A}}(x)$.

$$FOU(\tilde{A}) = \bigcup_{x \in X} [\underline{\mu}_{\tilde{A}}(x), \bar{\mu}_{\tilde{A}}(x)] \qquad (6)$$

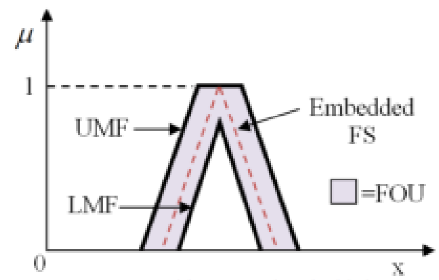

Figure 5: Type-2 Fuzzy Sets with FOU and Embedded Fuzzy Set (FS)

The operations on fuzzy interval type-2 set is almost the same as fuzzy type-1 set; but on the IT2FLS, the operation is performed on two intervals that are UMF and LMF at once.

In this research, our proposed model used the Gaussian primary membership function (MF) of fixed mean and uncertain spread (standard deviation) [27]:

$$\mu_{\tilde{A}}(x) = \exp\left[-\frac{1}{2}\left(\frac{x - v}{\sigma}\right)^2 \right],\; \sigma \in [\sigma_1, \sigma_2] \qquad (7)$$

Where σ_1 and σ_2 are its associated standard deviation for LMF and UMF, v is the fixed mean and x is the input values.

The Fuzzy C-Means Clustering

The membership function of T1FLS extracted automatically from data using Fuzzy C-Means Clustering (FCM) method. FCM has a capability to cluster data into specified number of overlapping groups/clusters and each group considered as membership functions/fuzzy sets. The FCM clustering algorithm was proposed by Dunn [32] and generalized by Bezdek [33]. In the classical clustering, each data point x in the given data set $X = \{x_1, x_2, \ldots, x_n\}$, $X \subseteq R^p$ is assigned to exactly in one cluster [34]. However, Fuzzy clustering methods allow the objects to belong to several clusters simultaneously, with different degrees of membership. FCM gives flexibility to data points can belong to more than one clusters. FCM algorithm aims to iteratively adapt the partitioning of the dataset so as to minimize a weighted sum of squared errors between data points and cluster centers [33]. The objective of the FCM algorithm is to minimize the squared error objective function $J(U,V)$ [33]:

$$J(U,V)=\sum_{i=1}^{n}\sum_{j=1}^{c}\left(\mu_{ij}\right)^{m}\left\|x_i-v_j\right\|^2 \qquad (8)$$

Where $V=(v_1,v_2,\ldots,v_c)$ vectors of centers, c is number of clusters and, n is the number of data points, $U=(\mu_{ij})_{nxc}$ is a fuzzy partition matrix, m is fuzzy partition matrix exponent for controlling the degree of fuzziness overlap, with $m\in(1,\infty)$. Fuzzy overlap refers to how fuzzy the boundaries between clusters the number of data points that have significant membership in more than one clusters. The x_i is the i^{th} data point and v_j is the center of the jth cluster. u_{ij} is the degree of membership of x_i in the jth cluster. For a given data point, x_i the sum of the membership values for all clusters is one.

$$u_{ji}\in[0,1],\ j=1,\ldots,c;\ i=1,\ldots,n \quad \text{and} \quad \sum_{j=1}^{c}\mu_{ji}=1$$

for $\forall i$.

The FCM work as follows [33]:

1. Randomly initialize the cluster membership values, u_{ij}.
2. Calculate the cluster centers
$$v_j=\frac{\sum_{i=1}^{n}(u_{ij})^m x_i}{\sum_{i=1}^{n}(u_{ij})^m} \qquad (9)$$
3. Update u_{ij} according to the following
$$\mu_{ij}=\left(\sum_{k=1}^{c}\left(\frac{d_{ji}}{d_{ki}}\right)^{\frac{2}{m-1}}\right)^{-1} \qquad (10)$$
where $d_{ji}=\left\|x_i-v_j\right\|$.
4. Calculate the objective function, $J(U,V)$
5. Repeat steps 2–4 until $J(U,V)$ improves by less than a specified minimum threshold or until after a specified maximum number of iterations.

FCM assigns each data values to a cluster which has smallest error value to its center.

The Proposed Type-2 Fuzzy Logic Based System for Malaria Epidemic Prediction

The real dataset we used for experiment are monthly confirmed Malaria case (Morbidity Data) and climate data. The Malaria cases data and climate data were collected from 2013 to 2017, in five years, throughout the country with areas of unstable malaria transmission as specified by Ministry of Health in Ethiopia and Ethiopia National Meteorology Agency respectively. The climatic inputs data are Monthly Average of Temperature (TEMP), rainfall (RF), Relative Humidity (RH) and Elevation (ELV) collected from ground Meteorology Stations. The other input is monthly lag outpatient malaria case (LAGCASE). The output is a number of outpatient malaria cases (CASE). Since these variables have different units, it would be difficult to establish correlation between them. Thus, min-max normalization method was applied to limit the value of these five predictors and output between 0 and 1. The formula for the normalization is as follow [35]:

$$Normalized\ (z_{new})=\frac{z_i-z_{\min}}{z_{\max}-z_{\max}} \qquad (11)$$

Where z_i and z_{new} is the old and transformed value of each input values, $z_{\min}$ is the minimum value and $z_{\max}$ is the maximum value from the data sample z, z represents each inputs.

The proposed type-2 fuzzy logic based system for Malaria Epidemic Prediction works into two phases as shown in Figure 6, a modeling phase and a prediction phase.

Modeling Phase

In this phase, Fuzzy Sets specified and the rule bases of Type-1 and Type-2 Fuzzy Logic System constructed.

Stage 1) Extracting Fuzzy Sets:-For each s continuous Input-Output variable generates all type-1 fuzzy sets, A_s^q for antecedent and B^q for consequent, where q is represents linguistic labels such as “Very Low”, ”Low”, ”Medium”, ”High”, ”Very High”, $q=1,2,..,V$, where V is the number of labels specified using FCM clustering techniques in MATLAB Toolbox.

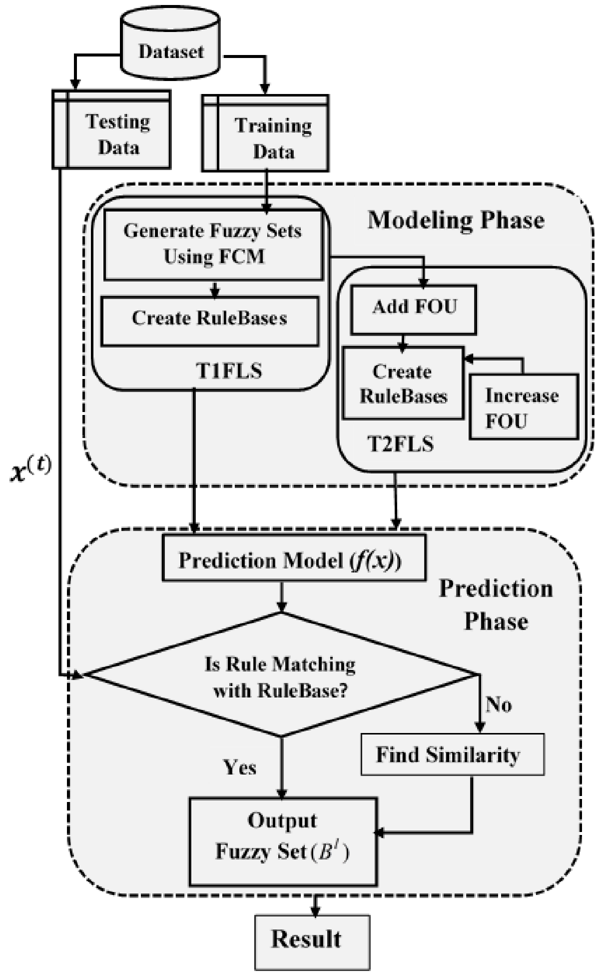

Figure 6: Proposed method for Malaria Epidemic prediction

Stage 2) Fuzzy Rule Extraction: A one pass method that was suggested by Wang-Mendel [27] employed to generate all rules from the accumulated data and finally, a malaria epidemic prediction model designed based on these combined fuzzy rules bases. The fuzzy rules are defined for the n-dimensional pattern problem as follows [27, 36]:

$$R_l: \text{IF } x_1 \text{ is } A_1^l \text{ and} \ldots \text{and } x_n \text{ is } A_n^l \text{ THEN } y_1 \text{ is } B_1^l \tag{12}$$

Where $l=(1,2,3,\ldots,M)$, M is the total number of rules and i is the index of the rules and $x=(x_1,x_2,\ldots,x_n)^T$ is an n-dimensional pattern vector. The different steps involved for rule extraction:

Step 1- For a fixed input-output pairs $(x^{(t)},y^{(t)})$ in the dataset, $t=1,\ldots,N$, (N is total number of data for modeling phase). For the case of T1FLS, compute the membership values $\mu_{A_s^q}(x_s^{(t)})$ for each antecedent fuzzy sets $q=1,\ldots,V$ (V_i is the total number of fuzzy sets representing the input pattern s where $s=1..n$, n is number of input variables). For the case of T2FLS, type-1 fuzzy sets transformed into type-2 fuzzy sets with adding some values on standard deviation as FOU to accommodate uncertainties. Then compute the upper $\overline{\mu}_{A_s^q}(x_s^{(t)})$ and lower $\underline{\mu}_{A_s^q}(x_s^{(t)})$ membership values for each antecedent and consequent fuzzy sets. Extract all rules combining the matched fuzzy sets A_s^q ($\mu_{A_s^q}(x_s^{(t)}) > 0$ for T1FLS and either $\overline{\mu}_{A_s^q}(x_s^{(t)})>0$ or $\underline{\mu}_{A_s^q}(x_s^{(t)}) > 0$ for T2FLS for all $s=1..n$. Find $q^* \in \{1,\ldots,V\}$ such satisfies the conditions expressed in (11) [36, 37]:

$$\mu_{A_s^{q^*}}^{cg}\left(x_s^{(t)}\right) \geq \mu_{A_s^q}^{cg}(x_s^{(t)}) \tag{13}$$

Where $\mu_{A_s^q}^{cg}(x_s^{(t)})$ is the center of gravity of the membership of A_s^q at $x_s^{(t)}$ as can be seen below [26 3537]:

$$\mu_{A_s^q}^{cg}\left(x_s^{(t)}\right) = \frac{1}{2}\left[\overline{\mu}_{A_s^q}(x_s^{(t)}) + \underline{\mu}_{A_s^q}(x_s^{(t)})\right] \tag{14}$$

The following rule will be referred to as the rule generated by $(x^{(t)};y^{(t)})$ [27, 36, 38]:

IF x_1 is $A_1^{q^{*(t)}}$ … and x_n is $A_n^{q^{*(t)}}$ THEN y is centered at $y^{(t)}$ (15)

For all of the input variables x_s there are V_i type-2 fuzzy sets A_s^q, which enables the greater amount of potential rules equal to V_i^n [38]. However, when considering a dataset, there will be the generation of those rules amongst the V_i^n possibilities that show a dominant region comprising a minimum of one data point. In step 1, one rule is generated for each input-output data pairs, where for each input the fuzzy set that achieves the maximum membership value at the data point is selected as the one in the "IF" part of the rule. The firing strength of the generated rule is calculated as follows [36]:

$$w^t = \prod_{s=1}^{n} \mu_{A_s^q}^{cg}(x_s^{(t)}) \tag{16}$$

The weight of a rule w^t is a measure of the strength of the points $x^{(t)}$ belonging to the fuzzy region covered by the rule.

Step 2- Step 1 repeated for all the t from 1 to N to obtain N data generated rules in the form of Equation (12). Since the number of data points pretty large consisting of many rules generated in step 1 that share the same IF part and different consequent i.e. rule conflicting. In this step, rules with the same IF part are combined into a single rule. The N divided into groups that share the same IF part and suppose there are M such groups. Let group l $(l = 1,\ldots,M)$ have N_l in the following forms [36]:

IF x_1 is A_1^l and… and x_n is A_n^l THEN y is centered at $y^{(t_u^l)}$ (17)

Where $u = 1,\ldots,N_l$ and t_u^l is the index for the data points in group l. Therefore, resolving conflicting rules will involve computing the weighted average $(av^{(l)})$ for the conflicting group. Then the weighted average computed as follows [36, 37]:

$$av^{(l)} = \frac{\sum_{u=1}^{N_l} y^{(t_u^l)} w^{(t_u^l)}}{\sum_{u=1}^{N_l} w^{(t_u^l)}} \tag{18}$$

These N_l rules are combined into a single rule, utilizing the following format [36, 38]:

IF x_1 is A_1^l and… x_n is A_n^l THEN y is B^l (19)

Where there is the selection of the output fuzzy set B^l based on the following: amongst the V_0 output type-2 fuzzy sets $B^l, \ldots, B^{V_0}$ find the B^{h*} such that [36, 38]:

$$\mu_{B^{h*}}^{cg}(av^{(l)}) \geq \mu_{B^h}^{cg}(av^{(l)}) \tag{19}$$

For $h = 1,2, \ldots, V_0$

B^l is selected owning to the fact that B^{h*}, where $\mu_{B^h}^{cg}$ is the center of gravity of the membership of B^h at $av^{(l)}$ as in Equation (14).

Prediction Phase

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. In this case, we will follow the process explained in case 1. If $x^{(t)}$ does not match any of the existing M rules, we will follow the process explained in case 2.

Case 1-The Input Matches with Existing Rule

In case the incoming input $x^{(t)}$ matches any of the existing M rules, the following formula will be used to obtain the output fuzzy set, $B^{(l)}$ [36]

$$\mu_{B_C^{h*}}^{cg}(y_c) \geq \mu_{B_C^h}^{cg}(y_c) \tag{20}$$

Where for $h = 1,\ldots,V_o$ and $c = 1,\ldots,k$. The consequent fuzzy set B_c is chosen as B_c^{h*}.

Case 2-The Input Doesn't Match with Existing Rule

In case the incoming input $x^{(t)}$ does not match any of the existing M, we need to find the closest rule in the rule base that match $x^{(t)}$. In order to do this, we need to calculate the similarity (or distances) between each of the fuzzy rule generated by $x^{(t)}$ and each of the M rules stored in the rule base [39, 40]. The rules generated by $x^{(t)}$ are found by taking each element in $x^{(t)}$ and taking all matching fuzzy sets with either $\overline{\mu}_{A_S^q}(x_s)$ or $\underline{\mu}_{A_S^q}(x_s)$ greater than 0 [39,40].

In order to calculate the similarity in the antecedent parts between the rule generated by the input $x^{(t)}$ and each rule R_l in the rule base, a function distance, defined as $dis(A_s^{q^*}, A_s^l)$, is used between $A_s^{q^*}(x_s^{(t)})$ and the i[th] antecedent part of the "IF" part of that rule. With this aim, we define a distance that find the difference between the linguistic labels. For example, the distances between "Very Low", "Low" is 1 and the distance between "Very Low" and "Medium" is 2. The similarity or the distance between the rule created by the input $x^{(t)}$ with each rule R_l in the rule base is calculated as [39, 40]:

$$s(x^{(t)}, R_l) = \frac{\sum_{s=1}^{n}\left(1 - \frac{dis(A_s^{(t)}, A_s^l)}{v-1}\right)}{n} \tag{21}$$

Where $s(x^{(t)}, R_l) \in [0,1]$, V is the number of fuzzy sets and $s = 1,\ldots,n$, where n is the number of input variables which is the number of antecedents of the rule, i.e. n-dimensional problem. The higher the similarity distance value is, the more similar the rule generated by the incoming inputs and the existing rule in the rule base. Therefore, for each incoming input $x^{(t)}$ which have not rule matching in the rule base, compute the similarity values $s(x_s, R_l)$ for each rule R_l, $l = 1,\ldots,M$ and find R_{l^*}, $l^* \in \{1,\ldots,M\}$, such that [39]:

$$s(x^{(t)}, R_{l^*}) \geq s(x^{(t)}, R_l) \text{ for } l = 1,\ldots,M \qquad (22)$$

Consequently, the new input classified with the consequent class of the rule R_{l^*}.

Experiments and Results

The dataset which has 11,281 instances divided into 70% to construct the model and the remaining instances to evaluate the model. The model's performance is evaluated by comparing the model output with the events that took place in the past of testing data set. The accuracy of type-1 and type-2 FLS compared using RMSE (Root Mean Square Error) in error minimization from the correct value. RMSE is commonly used to measure the accuracy of errors for numerical predictions have widely been used in evaluating the accuracy of a prediction system, given by Equation (23) [41]:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2} \qquad (23)$$

Where N is the number of actual historical data, y_t is actual value and $\hat{y}_t$ is 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 malaria epidemic prediction. In other words, the accuracy of model can be calculated as accuracy=1-RMSE.

In this paper, each input was represented by five fuzzy sets which are Very Low, Low, Medium, High and Very High by specifying Gaussian membership parameters (i.e. mean and spread) dynamically from data using FCM clustering method. The generated fuzzy sets of antecedents and consequents have overlapping domain interval ranges. For example, Type-1 fuzzy sets membership of temperature and relative humidity represented in Figure 7 and 8 respectively.

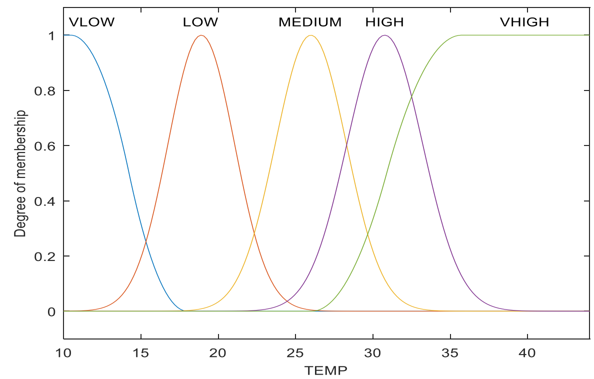

Figure 7: Type-1 membership function for Temperature

These membership functions are distributed over the total range of data values for each parameter and the membership functions at the boundaries are modified such that they are extended indefinitely beyond their respective centers with membership value of 1. A semantic meaning can be associated with each of the resulting fuzzy sets.

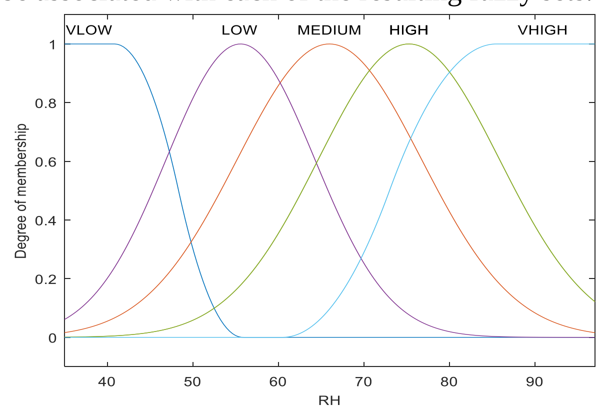

Figure 8: Type-1 membership function for Relative Humidity

The FOUs for the type-2 fuzzy sets were formed by adding some percent of uncertainty to either side of the spread σ_z^j for each of the Gaussian type-1 membership functions as shown Figure 9 and Figure 10 for temperature and relative humidity respectively. The uncertain standard deviation helps to capture the non-stationary behavior of the target outputs and the Gaussian interval type-2 membership functions are used to describe the type-2 fuzzy sets $\tilde{A}_z^j$ (where z=1, 2,…, V and V represents the number of type-2 fuzzy sets for a variable j) with the mathematical definition of Equation(24)[27]:

$$\mu_{A_z^j}(x) = \exp\left\{-\frac{1}{2}\left(\frac{x-m_z^j}{\tilde{\sigma}_z^j}\right)^2\right\}, \tilde{\sigma}_z^j \in \left[\sigma_{z_1}^j, \sigma_{z2}^j\right] \tag{24}$$

Where m_z^j is the value of the center (average) and $\tilde{\sigma}_z^j$ are the values of the spreads for each Gaussian interval type-2 membership function z, for the j^{th} input/output variable.

The WM method applied to generate Fuzzy rules by associating ach training data elements with pre-specified fuzzy sets by using java programming.

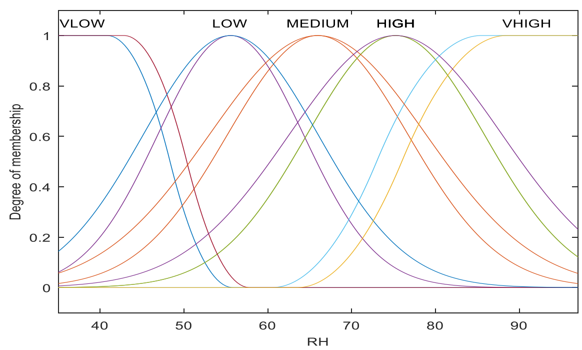

Figure 9: Type-2 Fuzzy Membership for Temperature

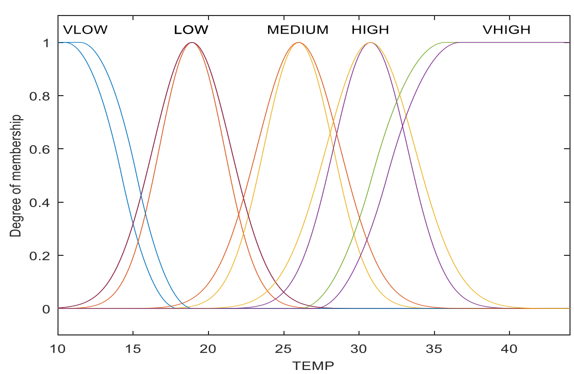

Figure 10: Type-2 Fuzzy Membership for Relative Humidity

Sample rules generated 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

Figure 11 shows the results of accuracy of T1FLS, T2FLS, and ANFIS for model testing. Hence, we can see that the testing results 86.99% for T2FLS (with uncertainty level of 10%) which outperforms the result of 86.66% for ANFIS (i.e. uplift of 0.33%) and the result of 85.1% with T1FLS (i.e. T2FLS uplift of 2% above the T1FLS).

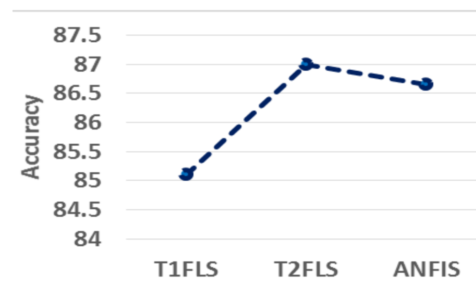

Figure 11: T1FLS, T2FLS and ANFIS Performance with Testing

As shown in Table 1, the accuracy deteriorated as the uncertainty factor of FOU increased from 10% to 40% while the number of rules decreased.

TABLE 1: RESULT OF T2FLS WITH DIFFERENT FOU

	FOU			
	10%	**20%**	**30%**	**40%**
RMSE	0.1301	0.1346	0.1361	0.1380
Accuracy (%)	86.99	86.54	86.39	86.2
No. of rules	651	601	447	290

The error minimization of T1FLS and T2FLS against each testing instances displayed in Figure 12 in which T2FLS error values are closer to 0 values thanT1FLS.

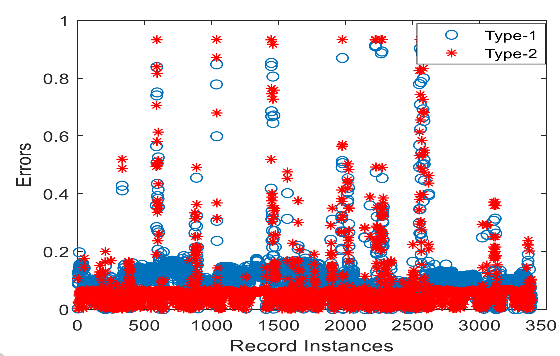

Figure 12: The error value of T1FLS and T2FLS for testing result

The main importance of this Fuzzy Logic Based System model is that the model can be explained and interpreted by the main stake holders and decision-makers like health experts. We can use the fuzzy IF-THEN rules to solve Malaria epidemic prediction 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 Malaria epidemic prediction in Ethiopia based on weather condition, elevation variation, and lag malaria cases. Generated rules with Medium, High and Very High outputs are more interesting which are close to malaria epidemic. For example:

R_3: 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 high level of elevation when the temperature becomes high (i.e. hot) even if high level of elevation in Ethiopia was considered a malaria 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 due to that high amount of rainfall (i.e. *RF_HIGH)* creates water pools which lead to hatching large number of Anopheles mosquitos (i.e. malaria parasite vectors) and medium level of temperature and relative humidity are 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 institutions and decision-makers to get control Malaria morbidity and deaths and try to take preventive measures based on the facts exposed by the extracted rules.

Conclusions and Future Work

In this paper, we presented an intelligent computational fuzzy logic-based malaria epidemic prediction system developed to predict malaria epidemic up to three months ahead of time.

Various scholars tried to predict complex malaria epidemic problem using statistical and mathematical models which assumed linearity of data and they have no capability to handle uncertainties. And also, machine learning techniques like Artificial Neural Network were used to predict malaria occurrence even if it is less interpretable due to its black-box nature.

We have proposed a fuzzy logic based system that is capable of handling the encountered uncertainties and capable of generating optimized models from pre-specified number of linguistic rules, which enables the user to understand the generated malaria occurrence model, thus generating a transparent and easy to read and analyze model.

In this paper, we have investigated T1FLS, T2FLS and ANFIS in predicting malaria epidemic using real climatic and lag malaria cases data which is collected by National Meteorology Agency and Ministry of Health in Ethiopia. Based on the testing result showed on Table 1, T2FLS at 10 % of FOU and ANFIS have high performance compared to T1FLS. Of course, T2FLS is more understandable and interpretable than ANFIS.

The proposed Type-2 Fuzzy Logic System is able to predict malaria epidemic ahead up to three months by using the current weather factor values and lag outpatient malaria cases; and the decision-makers will use as warning system while using the facts revealed by the rules to prevent and control malaria disease early before it leads to human morbidity and death.

In the future work, we will use optimization techniques like Genetic Algorithm and Big bang-Big crunch algorithms [42, 43, 44] that would help finding the optimal parameter values of the membership functions and the optimal number of rules for the automatic implementation of the proposed method.

REFERENCES

[1] Abebe A, Gemeda A, Wondewossen T, Lemu G. (2011). Climatic variables and malaria transmission

dynamics in Jimma town, South West Ethiopia. Parasites and Vectors 4.
[2] World Health Organisation. (2018).World malaria report. Geneva
[3] Taffese H., Hemming-Schroeder E, Koepfli C, Tesfaye G, Lee M., Kazura J, Yan G., Zhou G. (2018). Malaria epidemiology and interventions in Ethiopia from 2001 to 2016. Infectious Diseases of Poverty 7:1–9.
[4] Ethiopia FMOH. (2007). Entomological Profile of Malaria in Etiopia.
[5] Thomson M, Connor S. (2001). The development of Malaria Early Warning Systems for Africa. TRENDS in Parasitology 17:438–445.
[6] National Malaria Control Team, Ethiopian Public Health Institute, World Health Organisation, Addis Ababa University, INFORM Project. (2014). An Epidemiological Profile of Malaria in Ethiopia Report. Nairobi, Kenya.
[7] Zhou G., Minakawa N., Githeko A.andYan G. (2004). Association between climate variability and malaria epidemics in the East African highlands. Proceedings of the National Academy of Sciences of the United States of America 101:75–80.
[8] Teklehaimanot H., Lipsitch M., Teklehaimanot A. and Schwartz J. (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 11:1–11.
[9] Zhao X., Chen F., Feng Z., Li X. and Zhou X. (2014).The temporal lagged association between meteorological factors and malaria in 30 counties in south-west China: A multilevel distributed lag non-linear analysis. Malaria Journal 13:1–12.
[10] Eskinder .L, and Bern L. (2010). Model variations in predicting incidence of Plasmodium falciparum malaria using 1998-2007 morbidity and meteorological data from south Ethiopia. Malaria journal 9.
[11] Kifle M., Teklemariam T., Teweldeberhan A., Tesfamariam E., Andegiorgish A. and Azaria K. (2019). Malaria Risk Stratification and Modeling the Effect of Rainfall on Malaria Incidence in Eritrea. Journal of Environmental and Public Health 2019:1–11.
[12] Midekisa A, Senay G, Henebry G, Semuniguse P, Wimberly M. (2012). Remote sensing-based time series models for malaria early warning in the highlands of Ethiopia. Malaria Journal 11:1–10.
[13] Abeku T., Van Oortmarssen G., Borsboom G., De Vlas S.and Habbema J. (2003). Spatial and temporal variations of malaria epidemic risk in Ethiopia: Factors involved and implications. Acta Tropica 87:331–340.
[14] Chekol B., Hagras H. (2019). Employing Machine Learning Techniques for the Malaria Epidemic Prediction in Ethiopia. In: Proceedings of 10th Computer Science and Electronic Engineering (CEEC).IEEE: 89–94.
[15] Rismala R., Liang H. and Ardiyanti A. (2013). Prediction of Malaria Incidence in Banggai Regency Using Evolving Neural Network. In: International Conference on Technology, Informatics, Management, and Engineering & Environment (TIME-E 2013) Bandung, Indonesia.
[16] Santosh T. and Ramesh D. (2019). Artificial neural network based prediction of malaria abundances using big data: A knowledge capturing approach. Clinical Epidemiology and Global Health 7:121–126.
[17] Sudheer Ch., Sohani S., Deepak K., Anushree M., Chahar B., Nema A, Panigrahi B. and Dhiman R. (2014). A Support Vector Machine-Firefly Algorithm based forecasting model to determine malaria transmission. Neurocomputing 129:279–288.
[18] Hagras H. and Wagner C. (2009). Introduction to Interval Type-2 Fuzzy Logic Controllers - Towards Better Uncertainty Handling in Real World Applications. The IEEE Systems and Cybernetics eNewsletter 27.
[19] Sakalli A., Kumbasar T., Yesil Y. and Hagras H. (2014). Analysis of the performances of type-1, self-tuning type-1 and interval type-2 fuzzy PID controllers on the Magnetic Levitation system. In: Proceedings of the 2014 IEEE International Conference on Fuzzy Systems, Beijing, China.
[20] Doctor F, Hagras H. and Callaghan V. (2005). A type-2 fuzzy embedded agent to realise ambient intelligence in ubiquitous computing environments. Information Sciences 171:309–334.
[21] Olatunji S., Selamat A. and Raheem A. (2011). Expert Systems with Applications Predicting correlations properties of crude oil systems using type-2 fuzzy logic systems. Expert Systems with Applications 38:10911–10922.
[22] Wang C. and Dong M. (2012). Engineering Applications of Artificial Intelligence Modeling data uncertainty on electric load forecasting based on Type-2 fuzzy logic set theory. Engineering Applications of Artificial Intelligence 25:1567–1576.
[23] Bernardo D., Hagras H. andTsang E. (2013). A genetic type-2 fuzzy logic based system for the generation of summarised linguistic predictive models for financial applications. Soft Computing 17:2185–2201.
[24] ZADEH L. (1965). Fuzzy Sets. Information and Control 8:338–353.

[25] Wagner C. and Hagras H. (2010). Toward general type-2 fuzzy logic systems based on zSlices. IEEE Transactions on Fuzzy Systems18:637–660

[26] Mendel J., Hagras H., Tan W. and Melek W, Ying H. (2017). Introduction to type-2 fuzzy logic control. IEEE Press Canada

[27] Mendel J. (2017). Uncertain Rule-Based Fuzzy Systems Second Edition. Springer, Los Angeles.

[28] Zadeh L. (1975). The Concept of a Linguistic Variable and its Application to Approxirmate Reasoning-III*. Information Sciences 9:43–80.

[29] Karnik N, Mendel J. (1999). Applications of type-2 fuzzy logic systems to forecasting of time-series. Information Sciences 120:89–111.

[30] Hagras H., Wagner C. (2009). Introduction to Interval Type - 2 Fuzzy Logic Controllers - Towards Better Uncertainty Handling in Real World Applications. The IEEE Systems, Man and Cybernetics eNewsletter June 2009:1–9.

[31] Liang Q. and Mendel J. (2000). Interval type-2 fuzzy logic systems: Theory and design. IEEE Transactions on Fuzzy Systems 8:535–550.

[32] Dunn J. (1974). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. Journal of Cybernetics 3:32–57.

[33] James B., Robert E., William F. (1984). FCM : The Fuzzy C-Means Clustering Algorithm. Computers & Geosciences 10:191–203.

[34] Valente de O. and Pedrycz W. (2007). Advances in Fuzzy Clustering and its Applications. John Wiley & Sons, Chichester England.

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

[36] Doctor F., Hagras H. and Callaghan V. (2006). An Incremental Adaptive Life Long Learning Approach for Type-2 Fuzzy Embedded Agents in Ambient Intelligent Environments. IEEE International Conference on Fuzzy Systems 15:915–922.

[37] Wang L. (2003). The WM Method Completed: A Flexible Fuzzy System Approach to Data Mining. IEEE Transactions on Fuzzy Systems 11:768–782

[38] Ferreyra E, Hagras H, Mohamed A, Owusu G, Telecom B, Park A. (2017). A Type-2 Fuzzy Logic System for Engineers Estimation in the Workforce Allocation Domain. IEEE: 0–5.

[39] Garcia-Valverde T., Garcia-Sola A., Gomez-Skarmeta A., Botia J., Hagras H, Dooley J. and Callaghan V. (2012). An adaptive learning fuzzy logic system for indoor localisation using Wi-Fi in Ambient Intelligent Environments. IEEE International Conference on Fuzzy Systems/FUZZ-IEEE.

[40] Chimatapu R., Hagras H., Starkey A.and Owusu G. (2018). A big-bang big-crunch type-2 fuzzy logic system for generating interpretable models in workforce optimization. IEEE International Conference on Fuzzy Systems 2:1–8.

[41] Hyndman R. andKoehler A. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting 22:679–688.

[42] Bo Y., Hagras H., Daniyal A.and Mohammed A. (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.

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

[44] Michela A., Dario B., Hani H. (2016). Multi-Objective Evolutionary Optimization of Type-2 Fuzzy Rule-based Systems for Financial Data Classification. IEEE Transactions on Fuzzy Systems 25:993-1005