Adaptive Neuro-Fuzzy Expert Systems for Malaria Parasite Classification and Severity Level Prediction.

Adeyinka .O. Abiodun [1], Olatubosun Olabode .[2] Sunmola Fatai[2], Emuoyibofarhe Ozichi Justice[1]
[1] Computer Science Department, Bowen University
[2] Information System Department, The Federal University of Technology, Akure

Abstract

Malaria is a serious global health problem and it requires fast and effective diagnosis and classification of the type of infection. The Current malaria diagnosis relies primarily on microscopic examination of Giemsa-stained thick and thin blood films. One major challenge of this method is that, it requires vigorously trained technicians to efficiently detect and classify the malaria parasite species and also predict its severity levels of mild, moderate and severe. This paper present the classification of the Plasmodium falciparum based on severity level by applying the Neuro fuzzy rule based algorithm. The dataset is sourced from Federal Teaching hospital in Nigeria. The Dataset contains 18 features extracted from thin blood smear from 500 patient observations. The dataset is divided into the training set and validation set. The analysis of the results compared with other machine learning methods shows significant improvement with accuracy of .967 and reliability.

Keywords: Malaria, Neuro fuzzy, thick and thin blood film, Plasmodium, classification, Multilayered

Introduction

Research has shown that malaria is caused by four species of the genus malarial parasite Plasmodium which includes the Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale and Plasmodium malariae. Plasmodium is actually a small, single-cell blood organism or protozoan which originated from a species of mosquito called Anopheles (Lalloo, 2016). It is the Anopheles mosquito bites that can spread the Plasmodium into the human blood stream. Currently, there are two tests to detect the malaria parasite, the thick smears and the thin blood smears. The thin blood smear is used to determine the species of malaria parasite, while thick blood smear is used to find the density of malaria parasite per microliter of red blood cells (RBC) (Seman et.al, 2008).

Malaria can occur if a mosquito infected with the Plasmodium parasite bites you. The symptoms of malaria typically develop within 10 days to four weeks following the infection. In some people, symptoms may not develop for several months. Some malarial parasites can enter the body but will be dormant for long periods of time. Common symptoms of malaria include: shaking chills that can range from moderate to severe, high fever, profuse sweating, headache, nausea, vomiting, diarrhea, anemia, muscle pain, convulsions, coma, and bloody stools. Malaria can cause a number of life-threatening complications such as swelling of the blood vessels of the brain, or cerebral malaria, an accumulation of fluid in the lungs that causes breathing problems, or pulmonary edema, organ failure of the kidneys, liver, or spleen anemia due to the destruction of red blood cells and low blood sugar

Malaria is a common, life-threatening infection in endemic tropical areas and one that presents a diagnostic challenge to laboratories in most non-endemic countries. Gatti, et.al (2007) opine, a rapid and accurate diagnosis is a prerequisite for effective treatment, especially for the potentially fatal cases of Plasmodium falciparum infection

The lack of access to good quality diagnostic tests for infectious diseases contributes to the enormous burden of ill health in the developing world, where infectious diseases are the major causes of death and account for more than half of all deaths in children (WHO 2019). Each year, more than 2 million people die of malaria, approximately 4 million of acute respiratory infections and almost 3 million of enteric infections.

It is observed that clinical diagnoses are often made based on doctors' intuitions and heuristics experience rather than on the knowledge rich data hidden in the database. Such diagnosis leads to unwanted biases, errors and excessive medical costs which affects the quality of treatment provided to patients. Motivated by the necessity of such a system, in this paper, a method is suggested to efficiently diagnose the severity level of malaria fever, with the goal to decrease medical errors and superfluous practice variation, decreasing diagnostic time and enhancing patient safety and satisfaction.

In Rajaraman et.al, (2019), Malaria is a life-threatening disease caused by Plasmodium parasites that infect the red blood cells (RBCs). Manual identification and counting of parasitized cells in microscopic thick/thin-film blood examination remains the common, but burdensome method for disease diagnosis. Its diagnostic accuracy is adversely impacted by inter/intra-observer variability, particularly in large-scale screening under resource-constrained settings.

ANFIS is a simple data learning technique that uses a fuzzy inference system model to transform a given input into a target output. This prediction involves membership functions, fuzzy logic operators and if-then rules. There are two types of fuzzy system, commonly known as the Mamdani and Sugeno models. There are five main processing stages in ANFIS operation, including input fuzzification, application of fuzzy operators, application method, output aggregation, and defuzzification (Mora et al, 2006).

Literature Review

Malaria is a serious global health problem and it requires fast and effective diagnosis for detecting and classifying the type of infection (Pinkaew, et.al., 2015). It has been presented in many outlets that proper treatment of malaria should be administered in a timely fashion to prevent an outbreak of epidemics. Microscopic examination of thick blood films is one of the current standards for malaria diagnosis. However, inspecting a thick blood film is time-consuming and requires experienced technicians, particularly in developing countries where malaria are very prevalent but microscopy expertise may not be available. Pinkaew, et.al., (2015) developed a computerized system to aid such diagnosis. In his report, an automated classification system operating on digitized images of thick blood film was developed to classify between Plasmodium falciparum and Plasmodium vivax malaria parasite species. In Adetayo et.al., (2019), Olatubosun et.al., (2020) and Oluborode et.al (2020), ANFIS were used extensively to solve some health problems.

Kaewkamnerd et.al., (2018) opined, Current malaria diagnosis relies primarily on microscopic examination of Giemsa-stained thick and thin blood films. This method requires vigorously

trained technicians to efficiently detect and classify the malaria parasite species such as Plasmodium falciparum (Pf) and Plasmodium vivax (Pv) for an appropriate drug administration. However, accurate classification of parasite species is difficult to achieve because of inherent technical limitations and human inconsistency. To improve performance of malaria parasite classification, many researchers have proposed automated malaria detection devices using digital image analysis. These image processing tools, however, focus on detection of parasites on thin blood films, which may not detect the existence of parasites due to the parasite scarcity on the thin blood film. The problem is aggravated with low parasitemia condition. Automated detection and classification of parasites on thick blood films, which contain more numbers of parasite per detection area, could address the previous limitation. Kaewkamnerd et.al., (2018) in their work presented an automatic device for both detection and classification of malaria parasite species on thick blood film. The system is based on digital image analysis and featured with motorized stage units, designed to easily be mounted on most conventional light microscopes used in the endemic areas.

Malaria remains a major burden on global health, with roughly 200 million cases worldwide and more than 400,000 deaths per year (Mahdieh et,al,. 2018). Besides biomedical research and political efforts, modern information technology is playing a key role in many attempts at fighting the disease. One of the barriers towards a successful mortality reduction has been inadequate malaria diagnosis in particular. To improve diagnosis, image analysis software and machine learning methods have been used to quantify parasitemia in microscopic blood slides. The article Mahdieh et,al,, (2018) provided an overview of these techniques and discusses the current developments in image analysis and machine learning for microscopic malaria diagnosis. In their article, they organize the different approaches published in the literature according to the techniques used for imaging, image pre-processing, parasite and cell segmentation, feature computation, and automatic cell classification.

This study presents Diabetes Diagnosis with Maximum Covariance Weighted Resilience Back Propagation Procedure was carried out by Olatubosun et.at., (2019). The Maximum covariance method was divided into three phases. A large number of candidate's hidden units is considered by initializing their various weights with random values. Then the desired number of hidden units is selected amongst the candidates by using the maximum covariance. The weights feeding the output units are calculated with linear regression method. After the maximum covariance initialization, the network is trained with the resilient back propagation which is an adaptive training algorithm. The activation function in the hidden units is hyperbolic tangent function. Ten baseline variables includes, age, sex, body mass index, average blood pressure and six blood serum measurements, were obtained for each of n = 442 diabetes patients, as well as the response of interest, a quantitative measure of disease progression one year after baseline was used. The learning machine was trained, validated and tested. The result shows the algorithm is efficient in the diagnosis of who is a diabetic patient.

Adaptive Neuro-Fuzzy Inference Model

The systems that integrate the principles of artificial neural networks and fuzzy logic are called neuro-fuzzy systems. They use a learning ability of artificial neural networks based on training data in order to adapt the forms of membership fuzzy functions and inference fuzzy rules. In this way, in one system, the advantages of logical inference and learning are combined. One of the most commonly used neuro-fuzzy systems is ANFIS (Adaptive Neuro-Fuzzy Inference System). ANFIS is a multilayer neural network that is based on data (input-output vector) for training, provides a certain value of an output variable for certain inputs. An important feature is that ANFIS can effectively model nonlinear connections of inputs and outputs (Zacharie et.al., 2017). ANFIS training is based on the application of an algorithm of error propagation

backward, either alone or in combination with the method of least squared error. ANFIS uses the Takagi-Sugeno method of inference, and a typical fuzzy rule.

Eleftherios (2012), Jang (1993), Jang and Sun (1995) introduced the adaptive network-based fuzzy inference system (ANFIS). This system makes use of a hybrid learning rule to optimize the fuzzy system parameters of a first order Sugeno system. An example of a two input with two rules first order Sugeno system can be graphically represented by Figure 1.

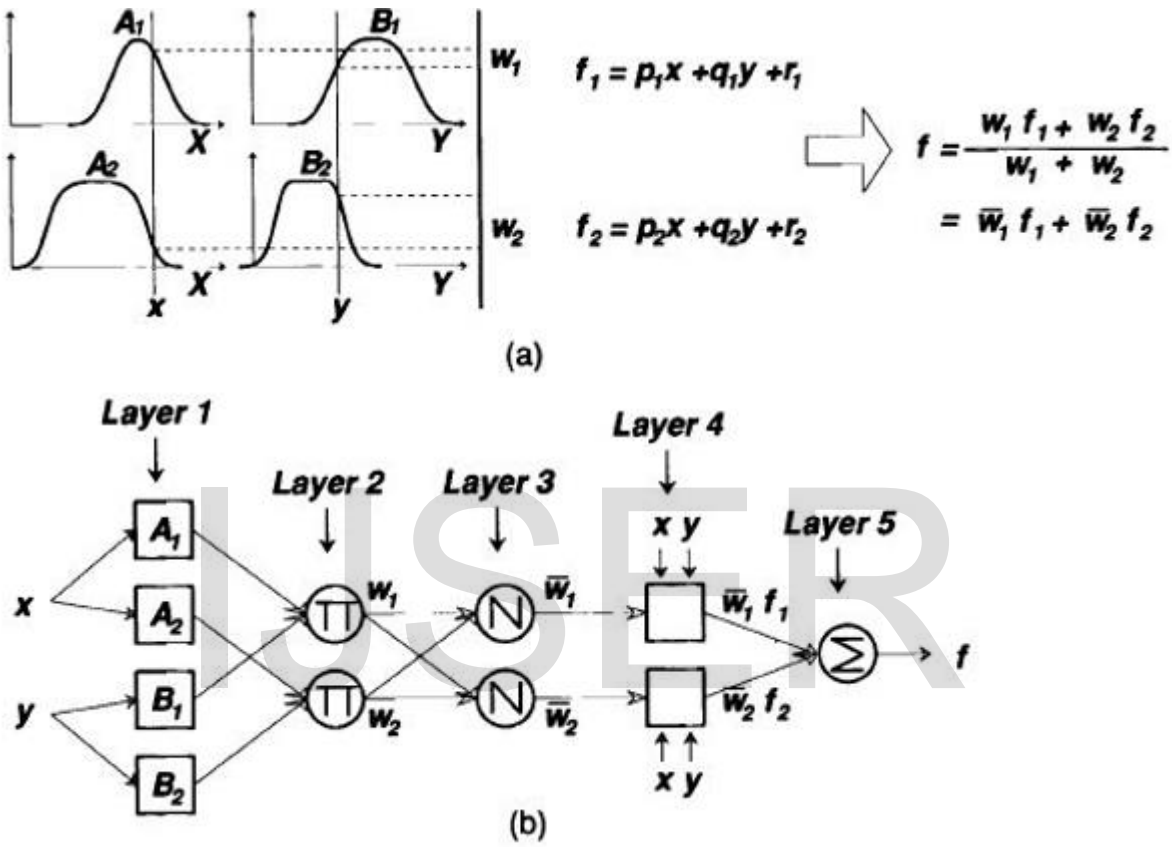


Figure 1. Architecture of an ANFIS model (adapted from Jang 1993).

For example, the ANFIS architecture for two input variables x1 and x2 is shown in Figure 1b and the fuzzy rules for this example can be presented as a rule

$$Rule\ R_i : IF(x1\ is\ B_i) and\ (x2\ is\ A_i)\ THEN\ f_i \qquad 1$$

$$f_i = p_i x1 + q_i x2 + r_i$$

where p_i, q_i and r_i are consequent parameters that can be estimated by using the ordinary least squares method. The B_i and A_i are linguistic values defined by fuzzy sets.

The ANFIS architecture is consisted of two trainable parameter sets, the antecedent membership function parameters and the polynomial consequent parameters *p*, *q*, *r*. The ANFIS training paradigm uses a gradient descent algorithm to optimize the antecedent parameters and a least squares algorithm to solve for the consequent parameters. Because ANFIS uses two very different algorithms to reduce the error, the training rule is called a hybrid. The consequent parameters are updated first using a least squares algorithm and the antecedent parameters are then updated by backpropagating the errors that still exist. The ANFIS architecture consists of

five layers with the output of the nodes in each respective layer represented by O_i^1, where *i* is the *ith* node of layer l.

The ANFIS structure consists of five layers, as shown in Figure 1b. The nodes of the first layer define fuzzy sets, i.e., membership fuzzy functions corresponding to input variables. This layer is often called a Fuzzification layer, because it determines the membership degree of the value of a variable to a particular fuzzy set (Qasem, et.al., 2017). A membership function of a fuzzy set, as proposed by Zadeh (1965), is a generalization of the indicator function in classical sets with a range covering the interval [0,1]. Fuzzy sets can be constructed using various simple linear or nonlinear functions. The nodes of this layer are adaptive, which means that their parameters are adjusted during a training period. The first-layer nodes that represent the membership functions of the input variable X can be defined as $\mu_{A_j}(x)$ where j $(j = 1, \ldots, 2)$ denotes the number of membership functions.

In Figure 1b, the fuzzification layer (Layer 1) associates a membership function with each node that defines the degree of compatibility with category represented by a fuzzy set (Godjevac 1999). For instance, for a generalized bell-shaped membership function, the output of the ith node of layer 1 (O_i^1) is given by:

$$O_i^1 = \mu_{A_i}(x) = 1/(1 + |(x - c_i)/a_i|^{2b_i}) \qquad 2$$

where μ_{A_i} is the membership degree of x (varying between 0 for a value excluded from the membership to 1 for a value that fully belongs to the category defined by the fuzzy set); and a_i, b_i and c_i are premise parameters. The parameter a_i corresponds to the membership function centre, while b_i and c_i determine the slopes at the inflection points of the function used to define the fuzzy set.

The (Layer 2) second-layer nodes are fixed and perform an operation of multiplying the input signals (operation AND). The *ith* neuron has the output of the form:

$$w_i = \mu_{B_i}(x1) \times \mu_{A_i}(x2) \qquad 3$$

The output of the second layer is equal to the minimum value of the two inputs (Pamučar, et.al 2016). Hence the output of Layer 1 nodes is obtained from:

$$O_i^2 = w_i = \mu_{B_i}(x1) \times \mu_{A_i}(x2) \qquad 4$$

where μ_{A_i} and μ_{B_i} are membership degrees for input variables $x1$ and $x2$, respectively.

The third layer normalizes the values obtained at the output of the nodes of the second hidden layer. In the case shown in Figure 1b, with two nodes in the second layer, the normalized value at the output of the *ith* node of the third hidden layer has the following mathematical form:

$$O_i^3 = \bar{w}_i = w_i/(w_1 + w_2) \qquad 5$$

In Layer 4, a linear combination of input variables is applied. Each node of the fourth layer is an adaptive node with the function it completes. The output of the Layer 4 nodes (O_i^4) is obtained from:

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i(p_i x1 + q_i x2 + r_i) \qquad 6$$

Where p_i, q_i and r_i are conclusion parameters that can be estimated using the least squares method for input variables $x1$ and $x2$.

Finally, the predicted output of Layer 5 (O_i^5) is obtained by a weighted mean of Layer 4 outputs:

$$O_i^5 = \sum_i \bar{w}_i(p_i x1 + q_i x2 + r_i) \quad 7$$

$$f = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad 8$$

The set of fuzzy inference rules that apply to the structure given in Figure 2 consists of two rules:
IF x is $A1$ AND y *is* B1, THEN $f1 = p1x + q1y + r1$
IF x is $A2$ AND y is B2, THEN $f2 = p2x + q2y + r2$

The flow of implementation of ANFIS is presented in Figure 2. To train fuzzy systems, the ANFIS hybrid algorithm uses first the least square estimator to adjust the consequent parameters according to premise parameters. Secondly, it uses the gradient descent algorithm to adjust the premise parameters according to the consequent parameters by adapting the connection weights to minimize errors.

Species of malaria parasite

We can recognize malaria parasites and their stages in thin films based on their appearances. However, the effect the parasite has on red blood cells is also important because it will help you to identify the malaria species. There are four species of malaria that affect humans:

• Plasmodium falciparum: the commonest species in the hotter parts of the world and responsible for much sickness and even death.

• P. vivax: the commonest species in the cooler parts of the tropics, the largest of the malaria parasites found in humans, and the cause of much illness.

• P. malariae: a less common species but one that occurs throughout much of the world.

• P. ovale: a relatively rare species but reported from time to time in many countries, especially in Africa; sometimes confused with P. vivax.

The simplest guide to distinguishing between the four species of malaria is the effect the parasite has on infected red blood cells. Features to concentrate on include the size of the red blood cell (whether or not it is enlarged) and whether or not staining reveals Schüffner's dots or Maurer's dots (also known as Maurer's clefts) within the cell. The diagnostic features outlined in Figure 3 will help to decide which species of malaria parasite you have found.

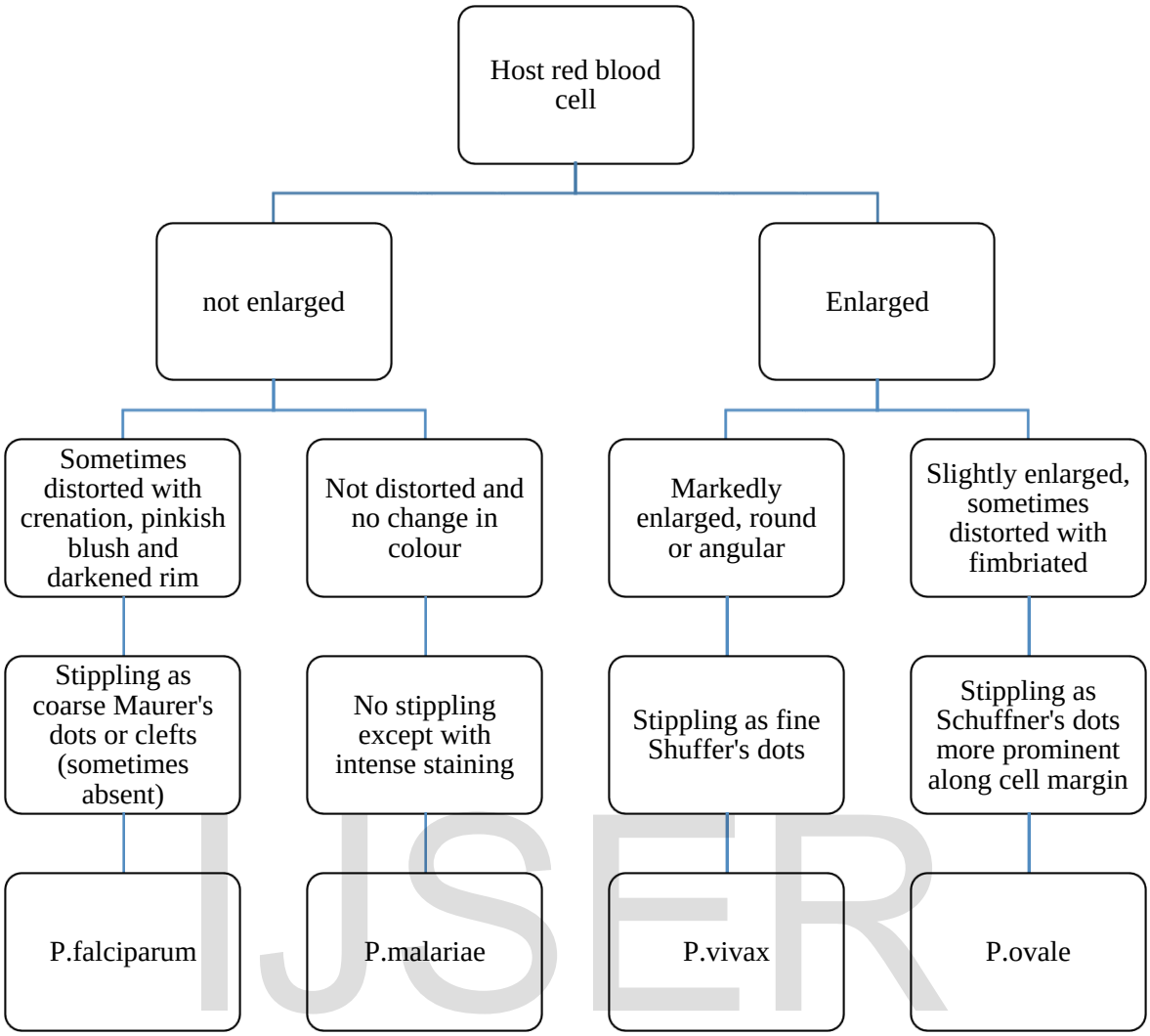


Figure 3. Species differentiation of malaria parasites by host-cell change in Giemsa-stained thin blood films

Fuzzy Logic Description of Input-Output Relations

The inputs $x_1, \ldots, x_n$ and outputs $y_1, \ldots, y_m$ are assumed to be available, the logical linguistic statement, namely,

IF (input x_1) AND ... AND (input x_n)

THEN (output y_1) AND ... AND (output y_m) actually described the unknown system, on the basis of the available data. A finite fuzzy logic implication statement can always be described by a set of general fuzzy IF-THEN rules containing only the fuzzy logic AND operation, in the following multi-input single-output form:

(1) "IF (x_1 is X_{11}) AND ... AND ($x_n$is X_{1n}) THEN (y is Y_1)."
(2) "IF (x_1 is X_{21}) AND ... AND ($x_n$is X_{2n}) THEN (y is Y_2)."
|
(*N*) "IF (x_1 is X_{N1}) AND ... AND ($x_n$is X_{Nn}) THEN (y is Y_N)."

Here, we should recall that the phrase "*x* is *X*" is an abbreviation of the complete statement "*x* belongs to the fuzzy subset *X* with a corresponding membership value $\mu_X(x)$." We should now

restrict our discussion on closed intervals for the fuzzy subsets $X_{11}, \ldots, X_{Nn}$ and $Y_1, \ldots, Y_N$, so that interval arithmetic can be applied.

For simplicity of discussion, If there are more than one fuzzy IF-THEN rule ($N > 1$):

$$R^i: \quad IF\ (x_1\ is\ X_{i1})\ AND \ldots AND\ (x_n\ is\ X_{in})$$
$$THEN\ y_i = a_{i0} + a_{i1}x_1 + \ldots + a_{in}x_n,\ i = 1, \ldots, N,$$

then, with the given set of inputs

$$x_1 = x_1^0 \epsilon X_1, \ldots, x_n = x_n^0 \epsilon X_n$$

namely, the same inputs applied to all the different rules, we will have

$$\begin{aligned} y_1^0 &= a_{10} + a_{11}x_1^0 + \ldots + a_{1n}x_n^0 \\ y_2^0 &= a_{20} + a_{21}x_1^0 + \ldots + a_{2n}x_n^0 \\ &\vdots \\ y_N^0 &= a_{N0} + a_{N1}x_1^0 + \ldots + a_{Nn}x_n^0 \end{aligned}$$

The corresponding membership values for these outputs are given, again by the general rule, as

$$\mu_Y(y_I^0) = \sup_{y_1^0 = a_{i0} + a_{i1}x_1^0 \ldots + a_{in}x_n^0} \{\mu_{X_1}(x_1^0) \wedge \ldots \wedge \mu_{X_n}(x_n^0)\}$$

where i = 1, ..., N. In a typical modeling approach, the final single output, y, is usually obtained via the following weighted average formula, using all y_1^0 with the weights $\mu_Y(y_i^0)$, usually called the center-of-gravity formula:

$$y = \frac{\sum_{i=1}^N \mu_Y(y_i^0) \cdot y_i^0}{\sum_{i=1}^N \mu_Y(y_i^0)}$$

where " $\cdot$ " is the ordinary algebraic multiplication (since all quantities here are real numbers). This operation is a convex combination of all outputs.
For the most general situation, we assume that all the coefficients

$\{a_{i0}, a_{i1}, \ldots, a_{in} | i = 1, 2, \ldots, N\}$ are uncertain and belong to certain intervals:

$$a_{i0} \epsilon A_0, \ldots, a_{in} \epsilon A_n \qquad i = 1, \ldots, \quad N$$

where, for example,

$$\begin{aligned} A_0 &= [min\{a_{10}, a_{20}, \ldots, a_{N0}\}, max\{a_{10}, a_{20}, \ldots, a_{N0}\}] \\ A_1 &= [min\{a_{11}, a_{21}, \ldots, a_{N1}\}, max\{a_{11}, a_{21}, \ldots, a_{N1}\}] \\ &\vdots \\ A_0 &= [min\{a_{1n}, a_{2n}, \ldots, a_{Nn}\}, max\{a_{1n}, a_{2n}, \ldots, a_{Nn}\}] \end{aligned}$$

Thus, with the given inputs

$$x_1 \in X_1, x_2 \in X_2, \ldots, x_n \in X_n$$

the output becomes

$$Y = A_0 + A_1 \cdot X_1 + \ldots + A_n \cdot X_n$$

which yields the fuzzy subset (interval) for y, with the membership functions given by the general rule as

$$\mu_{Y,i}(y_i) = \sup_{y_i = a_{i0} + \ldots + a_{in}x_n} sup\{\mu_{X_1}(x_1^0) \wedge \ldots \wedge \mu_{X_n}(x_n^0)\}$$

where i = 1, ..., N. Detail can be found in (Chen and Trung, 2000)

Data collection

The data is sourced from Federal Medical Center, Ido-Ekiti, Nigeria. It contains 18 features extracted from thin blood smear from 500 patient observations. The features in the dataset include sex, age, height, weight, blood pressure, body Temperature, total number of RBC per field, number of RBC infected field, number of normal RBC per field, number of ring form (early trophozite) per field, number of gametophytes per field, number of malaria parasite per RBC per field, number of chromatin dot per field, number of chromatin dot on cytoplasm of RBC per field, number of chromatin dot on membrane of RBC per field, number of chromatin dot on cytoplasm of malaria parasite, number of deformed infected RBC per field and Number of Malaria Parasite per RBC Field. The number of features in the dataset were reduced to 12 for lack of significant contribution to the proper classification. The features are detailed as follows.

Attributes of the Malaria Dataset

x_1 - Age

x_2 - Height

x_3 - Weight

x_4 - Blood Pressure

x_5 - Body Temperature

x_6 - Total Number of RBC per field

x_7 - Number of RBC Infected per field

x_8 - Number of Normal RBC per field

x_9 - Number of Ring Form (Early Trophozites)

x_{10} - Number of Late Trophozites per field

x_{11} - Number of Gametocytes per field

x_{12} - Number of Chromatin Dot per Field

x_{13} - Number of Chromatin Dot on Cytoplasm of RBC per field

x_{14} - Number of Chromatin Dot on Membrane of RBC

x_{15} - Number of Chromatin Dot on Cytoplasm of Malaria Parasite

x_{16} - Number of Deformed Infected RBC per field

x_{17} - Number of Malaria Parasite per field

For the first layer and relation (5) we use the Triangular shape membership functions. We have described the computation procedure for the consequent parameters by using least squares algorithm with Moore-Penrose pseudoinverse matrix. The next step is to describe the training procedure for the antecedent parameters with the error backpropagation algorithm and the simple steepest descent method [Tsoukalas and Uhrig, (1996); Denai et al., (2004); Khan et al., (2010)].

Definition: a membership function for a fuzzy set A on the universe of discourse X is defined as μA:X → [0,1], where each element of X is mapped to a value between 0 and 1. This value, called membership value or degree of membership, quantifies the grade of membership of the element in X to the fuzzy set A.

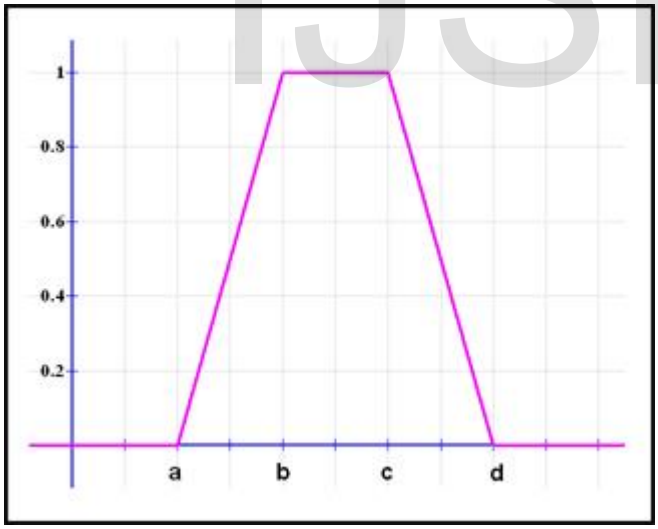
Membership functions allow us to graphically represent a fuzzy set. The x axis represents the universe of discourse, whereas the y axis represents the degrees of membership in the [0,1] interval.

Simple functions are used to build membership functions. Because we are defining fuzzy concepts, using more complex functions does not add more precision.

Below is a list of the membership functions we will use in the practical section of this tutorial.

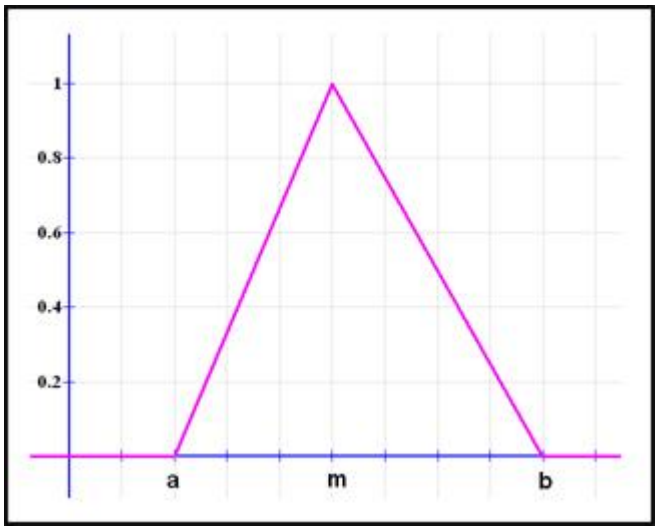
Trapezoidal function: defined by a lower limit a, an upper limit d, a lower support limit b, and an upper support limit c, where a < b < c < d

$$\mu_A(x) = \begin{cases} 0, & (x < a) or\ (x > d) \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \dfrac{d-x}{d-c}, & c \le x \le d \end{cases}$$



Triangular function: defined by a lower limit a, an upper limit b, and a value m, where a < m < b

$$\mu_A(x) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{m-a}, & a \le x \le m \\ \dfrac{b-x}{b-m}, & m \le x \le b \\ 0, & x \ge b \end{cases}$$



The triangular function is defined as equ 21.

$$\mu_{ij}(x_j: a_{ij}, b_{ij}) = \begin{Bmatrix} 1 - \frac{|x_j - a_{ij}|}{\frac{b_{ij}}{2}}, & if\ |x_j - a_{ij}| \le \frac{b_{ij}}{2} \\ 0 & otherwise \end{Bmatrix} \quad 21$$

where $a_{ij}$is the peak or center parameter and $b_{ij}$is the spread or support parameter.

Fuzzy Structure of the ANFIS for the Model

To create the fuzzy structure, the triangular function is defined in equation 21 is adopted as follows using Table 1.

Table 1: Linguistic Values and ranges for Fuzzy variables

Symbols	Very low	Low	Average	High	Very High
x_1	-	15-30	30-45	45-60	60-75
x_2		0-1	1-2	2-3	
x_3	10-30	30-50	50-70	70-90	90-110
x_4		70/40-90/60	100/70-139/80	140/90-100/100	
x_5		36-37	37-38	38-39	
x_6	128-228	228-328	328-428	428-528	528-628
x_7	12-24	24-36	36-48	48-60	
x_8	80-178	178-276	276-374	374-472	472-570
x_9	4-19	19-34	34-49	49-64	64-79
x_{10}	0-6	6-12	12-18	18-24	24-30
x_{11}	0-1	1-2	2-3	3-4	4-5
x_{12}	18-33	33-48	48-63	63-78	
x_{13}	0-2	2-4	4-6	6-8	8-10
x_{14}	0-4	4-8	8-12	12-16	16-20
x_{15}	0-4	4-8	8-12	12-16	16-20
x_{16}		0-1	1-2	2-3	

The dataset was randomly divided into two main categories namely: Training and Testing data set, which comprise of 80% and 20% of the total Malaria dataset respectively. Experiments with the dataset have concentrated on simply attempting to distinguish the severity level of malaria parasite (values s1, s2, and s3). Therefore, three main outputs are identified, where the value s1, s2 and s3 mean the presence of malaria attack is moderate, high and severe in the patient.

Fuzzy System Design

First step of fuzzy system designing is determination of input and output variables. There are 15 input variables and 1 output variable in the malarial dataset. However, the dimension was reduced to 12 input variables and 1 output variable with Principal Component Analysis. After that, we must design membership functions (MF) of all variables. These membership functions determine the membership of objects to fuzzy sets.

At first, we will describe some of the input variables with their membership functions. In second step, we introduce the output variable with its membership functions. In next section, we will show the rules of system.

Input Variables are:

Age: In this field we divide 4 fuzz sets. The fuzzy sets are infant (<38), young adult (40-45), meddle aged adult (40-58) and old adult (52>). Membership functions of "Infant" and "Old adult" sets are trapezoidal and membership functions of "young adult" and "middle aged adult" sets are triangular. We have defined fuzzy membership expressions for Age input field as follows

$$\mu_{Infant}(x_1) = \begin{cases} 1 & x_1 < 29 \\ \frac{8 - x_1}{9} & 29 \le x_1 < 38 \end{cases}$$

$$\mu_{young\ adult}(x_1) = \begin{cases} \frac{x_1 - 33}{5} & 33 \le x_1 < 38 \\ 1 & x_1 = 38 \\ \frac{45 - x_1}{21} & 24 \le x_1 < 45 \end{cases}$$

$$\mu_{middle\ aged\ adult}(x_1) = \begin{cases} \frac{x_1 - 40}{28} & 40 \le x_1 < 48 \\ 1 & x_1 = 48 \\ \frac{58 - x_1}{10} & 48 \le x_1 < 58 \end{cases}$$

$$\mu_{Old\ adult}(x_4) = \begin{cases} \frac{x_1 - 52}{30} & 52 \le x_1 < 60 \\ 1 & x_1 \ge 60 \end{cases}$$

Blood Pressure: Different values of blood pressure change the result easily. In this field, we use systolic blood pressure. This input variable is divided into 4 fuzzy sets. Fuzzy sets are "Low" (<134), "Medium" (127-153), "High" (142-172) and "Very high" (154>). The

Membership functions of “Low” and “Very high” sets are trapezoidal and membership functions of “medium” and “high” sets are triangular. We have defined fuzzy membership expressions for blood pressure input field as follows.

$$\mu_{low}(x_4) = \begin{cases} 1 & x_4 < 111 \\ \dfrac{134 - x_4}{23} & 111 \le x_4 < 134 \end{cases}$$

$$\mu_{Medium}(x_4) = \begin{cases} \dfrac{x_4 - 127}{12} & 127 \le x_4 < 139 \\ 1 & x_4 = 139 \\ \dfrac{153 - x_4}{14} & 139 \le x_4 < 153 \end{cases}$$

$$\mu_{High}(x_4) = \begin{cases} \dfrac{x_4 - 142}{15} & 142 \le x_4 < 157 \\ 1 & x_4 = 157 \\ \dfrac{172 - x_4}{15} & 157 \le x_4 < 172 \end{cases}$$

$$\mu_{Very\ high}(x_4) = \begin{cases} \dfrac{x_4 - 154}{17} & 154 \le x_4 < 171 \\ 1 & x_4 \ge 171 \end{cases}$$

Red Blood Count Total Number of RBC per field

$$\mu_{very\ low}(x_6) = \begin{cases} 1 & x_6 < 178 \\ \dfrac{228 - x_6}{50} & 178 \le x_6 < 228 \end{cases}$$

$$\mu_{low}(x_6) = \begin{cases} \dfrac{x_4 - 228}{50} & 228 \le x_4 < 278 \\ 1 & x_4 = 278 \\ \dfrac{328 - x_4}{50} & 278 \le x_4 < 328 \end{cases}$$

$$\mu_{mediu}(x_6) = \begin{cases} \dfrac{x_4 - 328}{50} & 328 \le x_4 < 378 \\ 1 & x_4 = 378 \\ \dfrac{428 - x_4}{50} & 378 \le x_4 < 428 \end{cases}$$

$$\mu_{High}(x_6) = \begin{cases} \dfrac{x_4 - 428}{50} & 428 \le x_4 < 478 \\ 1 & x_4 = 478 \\ \dfrac{528 - x_4}{50} & 478 \le x_4 < 528 \end{cases}$$

$$\mu_{Very\ high}(x_6) = \begin{cases} \dfrac{x_4 - 528}{50} & 528 \le x_4 < 578 \\ 1 & x_4 \ge 578 \end{cases}$$

Output Variable:

The "goal" field refers to the severity level of malaria parasite in the patient. It is integer value from 1 (moderate), 2 (Mild), to 3 (Severe). By increasing of integer value, the malaria severity level increases in patient. In this system, we have considered a different output variable, which divides to 4 fuzzy sets (Healthy (He), Moderate (Mo), Mild Sick (Mi) and Severe (Se)). Table 2 shows these fuzzy sets with their ranges. Membership functions of "Healthy" and "Severe" fuzzy set is trapezoidal membership functions while the Moderate and Mild fuzzy set are triangular membership function.

Table 2: Classification Output

Output Field	Range	Fuzzy Set
Results	<1.28	Healthy
	1-2.57	Moderate
	1.88-3.45	Mild
	3.2-4.5	Severe

Results and Discussion

Table 3 Summary of ANFIS technique for Classification Malaria Parasite and the Prediction of Severity Level.

Object model	*Specification/Measures*
Type of fuzzy inference system	sugeno
Input neuron	$x_1, x_2, \ldots, x_{17}$
Output neuron	Y
Input Membership function type	Triangular
Number of input membership functions	4
Logical operations	AND
Output membership function type	linear
Network type	Feed-forward backpropagation
Number of rules	164
Training error goal	0.2
Performance function	MSE
Input neuron	Seventeen
Output neuron	One
Maximum epochs (cycles) set	950

Conclusion

Creation, training and testing of the ANFIS model is performed in the Python software package. The total set of data on the variables that enters the system, is divided into three parts:
The training data, consisting of 70% input-output vectors, providing the Learning with a teacher, where the outputs from the network are known in advance for appropriate inputs.
The checking data, which is primarily aimed at preventing the occurrence of training data overfitting. The ANFIS model monitors the value of the checking error in each training epoch and retains learned Parameters at its minimum value. Checking data consists of 13% input-output vectors. And the testing data which enables us to perform an evaluation of the abilities of the ANFIS model to perform a prediction of the malaria severity level in the system as accurately as possible. The outputs of the ANFIS model are compared with known values, and the goal is to select a model that makes a minimum error. As well as checking data, testing data consists of 17% of the total set of data. The Table 4 presents the best model with .967 accuracy.

Table 4 Accuracy of ANFIS results

Index	*ANFIS technique*
Mean square error	83.710
Mean absolute deviation	6.985
Coefficient of correlation	0.967

Reference

Adetayo A. Adebayo, Tomilayo Oguntuyi, Olatubosun Olabode, Mosud Y. Olumoye, John C. Nwaiwue (2019) Malaria Parasite Classification based on Severity level: A Comparative Evaluation of Sigmoid and Radial Basis Function. International Journal of Scientific & Engineering Research Volume 10, Issue 7, July-2019, ISSN 2229-5518.

Chakrabortya K, Chattopadhyayb A, Chakrabarti A, Acharyad T, Dasguptae AK., 2015. A Combined Algorithm for Malaria Detection from Thick Smear Blood Slides. *J Health Med Informat* 6: 179. doi:10.4172/2157-7420.1000179.

Chen Guanrong, Trung Tat Pham. (2000). Introduction to Fuzzy Sets, Fuzzy Logic, and Fuzzy Control Systems. CRC Press Boca Raton London New York Washington, D.C. ISBN 0-8493-1658-8 (alk. paper)

Eleftherios Giovanis (2012). Application of Adaptive Neuro-Fuzzy Inference System In Interest Rates Effects on Stock Returns. Indian Journal of Computer Science and Engineering (IJCSE). ISSN: 0976-5166 Vol. 2 No. 1

Eleftherios Giovanis 2012. Application of Adaptive Neuro-Fuzzy Inference System in Interest Rates Effects on Stock Returns. Indian Journal of Computer Science and Engineering (IJCSE). ISSN: 0976-5166 Vol. 2 No. 1

Gatti S, Gramegna M, Bisoffi Z, Raglio A, Gulletta M, Klersy C, et al. A comparison of three diagnostic techniques for malaria: A rapid diagnostic test (NOW Malaria), PCR and microscopy. Ann Trop Med Parasitol 2007;101:195-204.

Jang, J.S.R. & Sun, C.T. (1995). Neuro-Fuzzy Modeling and Control, Proceedings of the IEEE, Vol. 83, pp. 378-406.

Jang, J.S.R. (1993). ANFIS: adaptive-network-based fuzzy inference system, IEEE Transactions on Systems, Man and Cybernetics, Vol. 23, No.3, pp. 665–685.

Kaewkamnerd, S., Uthaipibull, C., Intarapanich, A., Pannarut, M., Chaotheing, S. and Tongsima, S., 2012. An automatic device for detection and classification of malaria parasite species in thick blood film. *Bmc Bioinformatics*, *13*(17), p.S18.

Mahdieh Poostchi, Kamolrat Silamut, Richard Maude, Stefan Jaeger, George Thoma, Image analysis and machine learning for detecting malaria, *Translational Research* (2018), https://doi.org/10.1016/j.trsl.2017.12.004

Mora, J.J., Carrillo, G. and Perez, L. (2006) 'Fault location in power distribution system using ANFIS nets and current patterns', IEEE/PES Transmission and Distribution Conference and Exposition, TDC'06, Latin America, 15–18 August, pp.1–6.

Olatubosun, O., Fasoranbaku, O., and Fagbuwagun, A. 2015. Diabetes Diagnosis with Maximum Covariance Weighted Resilience Back Propagation Procedure. *British Journal of Mathematics & Computer Science* 6(5): 381-393, Article no.BJMCS.2015.085. ISSN: 2231-0851.

Olatunbosun Olabode, Oladunni Daramola, Racheal Akinbo (2020). Mobile Application for Monitoring and Management of Out-Patients. International Journal of Scientific and Research Publications, Volume 10, Issue 3, March 2020. ISSN 2250-3153 DOI: 10.29322/IJSRP.10.03.2020.p9931, http://dx.doi.org/10.29322/IJSRP.10.03.2020.p9931.

Oluborode K.O, Obe O. O, Olabode O, Adeboje O. T (2020). Adaptive Neuro-fuzzy Controller for Double Lane Traffic Intersections. International Journal of Advanced Trends in Computer Science and Engineering. Volume 9, No.4, ISSN 2278-3091. doi.org/10.30534/IJATCSE/2020/330942020.

Pamučar, D.; Ljubojević, S.; Kostadinović, D.; Đorović, B. Cost and risk aggregation in multi-objective route planning for hazardous materials transportation. A neuro-fuzzy and artificial bee colony approach. *Expert Syst. Appl.* **2016**, *65*, 1–15.10.1080/02626667.2017.1367396

Pinkaew, A., Limpiti, T. and Trirat, A., 2015, November. Automated classification of malaria parasite species on thick blood film using support vector machine. In *Biomedical Engineering International Conference (BMEiCON), 2015 8th* (pp. 1-5). IEEE

Qasem, S.N.; Ebtehaj, I.; Riahi Madavar, H. Optimizing ANFIS for sediment transport in open channels using different evolutionary algorithms. *J. Appl. Res. Water Wastewater* **2017**, *4*, 290–298.

Rajaraman S, Jaeger S, Antani SK. 2019. Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images. PeerJ 7:e6977 https://doi.org/10.7717/peerj.6977.

Seman, N.A., Isa, N.A.M., Li, L.C., Mohamed, Z., Ngah, U.K. and Zamli, K.Z., 2008. Classification of malaria parasite species based on thin blood smears using multilayer perceptron network. *International Journal of the Computer, the Internet and Management*, *16*(1), pp.46-52.

Takagi, T. and Sugeno, M., 1985. Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man, and Cybernetics, Adaptive neuro-fuzzy inference system methods

World Health Organization. Malaria Rapid Diagnostic Test Performance - Results of WHO Product Testing of Malaria RDTs: Round 1-3: World Health Organization; 2008. p. 1-2. Back to cited text no. 11.

Zacharie Sirabahenda, André St-Hilaire, Simon C. Courtenay, Ashley Alberto & Michael R. Van Den Heuvel (2017) A modelling approach for estimating suspended sediment concentrations for multiple rivers influenced by agriculture, Hydrological Sciences Journal, 62:13, 2209-2221, DOI:

Zadeh, L. (1965). Fuzzy sets. Inf Cont, Vol. 8, pp. 338–353.