**Communication in Physical Sciences, 2022, 8(4): 556-572**

# 2Optimization of Determinant Diagnostic Symptoms for Febrile Diseases using Genetic Algorithm.

**Edith Ugochi Omede and Stella Chiemeke**
**Received: 18 August 2022/Accepted 25 September 2022/Published online: 29 September 2022**

***Abstract:*** *Many diseases especially febrile diseases present numerous mimicking and confusing symptoms that pose great challenges to their proper distinctive syndromic diagnosis. This ambiguity causes inaccurate diagnoses which result on the misappropriation of treatment. Many victims of this situation have been left in worse health conditions or even death.This paper considers two febrile diseases that are believed to be in the blood of every Nigerian, these are Malaria and Typhoid Fever. This challenge of their distinctive syndromic diagnostic symptoms was tackled by optimizing the numerous symptoms using a genetic algorithm based on their manifestation degree (the frequency of occurrence of a symptom in different cases). The genetic algorithm was simulated using matlabR2013a. An optimization degree of 64.06% was obtained. Though the conventional method is the best for disease diagnosis, it is not always available, especially in rural areas where many depend on low-skilled medical practitioners for their health care. The use of these optimized determinant symptoms in the syndromic diagnosis of Malaria and Typhoid fever will reduce the risk of misdiagnosis of these two diseases.*

**Keywords:** *Symptom Optimization, Determinant Symptoms, Genetic algorithm (GA), Malaria diagnosis, Typhoid diagnosis.*

**Edith.U. Omede***
**Department of Computer Science**
**[1]Delta State University, Abraka, Delta State, Nigeria**
**Email: edithomede@delsu.edu.ngg, edigra4Jesus@gmail.com**
**Orcid id: 0000-0002-9627-0552**

**Stella. C. Chiemeke**
**Department of Computer Science**
**[2]University of Benin, Benin City, Edo State,. Nigeria**

## 1.0 Introduction

Malaria and Typhoid fever are among the life-threatening febrile diseases in developing countries. They are curable but are complicated when neglected and not properly diagnosed and treated (Chiemeke and Omede, 2014). According to the WHO report (2018), an estimated 1.4 to 2.6 million deaths per year in sub-Saharan Africa are caused by malaria. These diseases, though caused by different micro-organism Plasmodium and Salmonella typhi respectively, are often present with mimicking or overlapping symptoms, especially in the early stages of typhoid fever (Simon-Oke & Akinbote, 2020; Odikamnoro *et al.*, 2018; Ohanu *et al*. 2003). This situation often presents a diagnostic problem and, in some cases, could lead to diagnostic confusion, especially in syndromic diagnosis (Uneke, 2008; Adehor and Burell, 2008).

This mimicking nature of the two diseases poses a great problem for their proper distinctive diagnosis, so consideration of how to determine their discernable signs and symptoms becomes very important for their proper distinctive diagnosis. In this research, the discernable symptoms termed determinant symptoms are obtained by optimization using a genetic algorithm.

The Genetic Algorithm (GA), as one of the global optimization algorithms, stemmed from the principle of natural selection; the survival of the fittest and extinction of the inferior. It can be used to handle multi-objective optimization problems. GA was proposed and developed by Professor John

***Corresponding Author*: E. U. Omede, *Email:* *edithomede@delsu.edu.ngg***

Holland from the University of Michigan in the late 1950s and early 1960s. GA shows its ability to achieve better performance than other algorithms in solving highly complex spatial problems because of its high versatility and strong robustness (Haibo, 2022, Muhammad *et al*.,2018).

The GA optimization is based on successive generations that combine (crossover operator) the best solutions (selection operator and fitness) to create new solutions. Random changes (mutation operators) are also considered to avoid local optimization. GA is classified into the group of evolutionary algorithms (EA). The optimization algorithms included in EA differ in genetic representation and other implementation details, as well as the nature of the applied problem (Guerrero *et al*., 2022).

A study of previous works on Malaria and Typhoid fever revealed that numerous types of research have been stirred up by the prevalent nature of these two diseases in the Sub-Saharan region of Africa on both conventional diagnoses (Odikamnoro *et al*., 2018; Davis 2022; Nas *et al*., 2018, Ohanu *et al*.,2003;Uneke, 2008; Simon-Oke and Akinbote 2020) and automated diagnosis of Malaria and Typhoid fever (Adehor and Burrel, 2008a-b,; Adetunmbi *et al*., 2012, Agakar and Ghatol, 2010; Chiemeke and Omede, 2014; Djam *et al*., 2011, Oguntimilehim *et al*., 2013, Olabiyisi *et al*. 2011; Samuel and Omisore, 2013). Conventional methods which involve going through laboratory test provides more distinct diagnosis but failed to be sufficient for numerous people being infected especially those in areas where there is little or no health care. Several years of research on Malaria and Typhoid fever-assisted diagnostic tools have birthed expert systems for syndromic diagnosis, many of these systems did not consider the effect of the mimicking symptomatology of malaria and typhoid fever with other febrile diseases which pose difficulty in their proper distinctive syndromic diagnosis. It was also observed that many researchers have successfully carried out optimization processes using genetic algorithms (Yang *et al*. 2010, Muhammad *et al*., 2018; Haibo, 2022; Guerrero *et al.* 2022). Hence this research employs symptoms optimization using a genetic algorithm for the selection of discernable symptoms (regarded as 'determinant symptoms' in this work) for distinctive diagnosis of each of the diseases.

## 2.0 Materials and Methods

### *2.1 Material*

The numerous symptoms of Malaria and Typhoid fever were gotten from a series of consultations with
medical experts (especially those in Delta state university teaching hospital Oghara and Delta state health service centre, Abraka) and from the study of standard literature in the tropical medicine field and records of confirmed patients.

#### 2.1.1 *Dataset*

The dataset is comprised of different symptoms that are manifested by malaria and typhoid fever at different infestation periods which are regarded in this study as early (less than or equal to 2 weeks) and late (greater than 2 weeks). The symptoms and their categories are shown in Table 1.

Table 1 depicts different symptom categories for both Malaria and Typhoid fever with their manifestations in infected patients at different periods (P) from the point of infestation. M and T represent malaria and typhoid respectively, "a"=less or equal to 2 weeks and "b"= greater than 2 weeks. The period factor is very important because most of the symptoms are time-dependent. "*" indicates the presence of a symptom.

The manifestation degree of symptoms at different periods is indicated as: A ≡ "Almost all"; V ≡ "Very common; C ≡"Common"; U ≡ "Uncommon"; R ≡ "Rare";Vr ≡ "Very

rare";Xr ≡ "Extremely rare"; "-" ≡ Symptom not associated with the disease.

**Table 1: Categorized Symptoms of Malaria and Typhoid Fever with the degree of periodic manifestation**

| S/No | Category | Symptom | S_Label | M | T | a | b |
|---|---|---|---|---|---|---|---|
| **1** | Systemic | High fever | $S_1$ | * | | Vc | |
| **2** | Features ($X_1$) | Stepwise Fever | $S_2$ | | * | Vc | Vc |
| **3** | | Chills | $S_3$ | * | * | U/A | - |
| **4** | | Diaphoresis(Excessive sweating) | $S_4$ | * | * | Vc | - |
| **5** | | Rigors (exaggerated chill) | $S_5$ | * | * | A/U | Uc |
| **6** | | Anorexia (Loss of appetite) | $S_6$ | * | * | C/A | - |
| **7** | | Lethargy (Fatigue) | $S_7$ | * | | A | - |
| **8** | | Insomnia | $S_8$ | | * | Vc | - |
| | | Weight loss | $S_9$ | * | * | C/C | |
| **9** | Neurologic | Malaise(ill feeling) | $S_{10}$ | * | * | C/A | C |
| **10** | Features($X_2$) | Frontal Headache | $S_{11}$ | | * | Vc | Vc |
| **11** | | Headache | $S_{12}$ | * | | C | C |
| **12** | | Psychosis (mental disability) | $S_{13}$ | | * | Vr | C |
| **13** | | Confusion /delirium | $S_{14}$ | * | * | R/Vc | - |
| **14** | Pulmonary | Bronchitic Cough | $S_{15}$ | | * | C | |
| **15** | Feature($X_3$) | Cough | $S_{16}$ | * | | A | A |
| **16** | | Rales (sound from the unhealthy lung) | $S_{17}$ | | * | C | - |
| **17** | | Mild cough | $S_{18}$ | | * | C | - |
| **18** | | Pneumonia | $S_{19}$ | | * | R | C |
| **19** | Ear, Nose, Throat Feature ($X_4$) | Coated Tongue | $S_{20}$ | | * | Vc | - |
| | | Epistaxis(Nose bleed) | $S_{21}$ | | | | |
| **20** | Dermatologic Feature ($X_5$) | Rose spot | $S_{22}$ | | * | R | - |
| **21** | | Dicrotic pulse | $S_{23}$ | | * | R | C |
| **22** | Cardiovascular Feature ($X_6$) | Thrombophlebitis (blood clot) | $S_{24}$ | | * | | Vr |
| **23** | | Nausea and Vomiting | $S_{25}$ | * | | C | C |
| **24** | Gastro intestine | Diarrhea | $S_{26}$ | * | * | C/- | C/C |
| **25** | Feature($X_7$) | Jaundice | $S_{27}$ | * | * | C/C | C/- |
| **26** | | Constipation | $S_{28}$ | | * | Vc | C |
| **27** | | Bloating | $S_{29}$ | | * | Vc | - |
| **28** | | Intestine hemorrhage | $S_{30}$ | | * | Vr | Vc |
| **29** | | Splenomegaly | $S_{31}$ | * | | - | R |
| **30** | | Hepatosplenomegaly | $S_{32}$ | | * | C | |
| | | Intestinal Perforation | $S_{33}$ | | | | |
| **31** | Musculoskeletal Feature($X_8$ ) | Myalgia (Muscle pain) | $S_{34}$ | * | | Vc | |
| **32** | | Arthralgia (Joint pain) | $S_{35}$ | * | | Vc | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **33** | Urogenital Feature ($X_9$) | Urinary retention | $S_{36}$ | | * | C | |
| **34** | | Renal pain | $S_{37}$ | * | * | -/R | Vr/- |

Table 2 shows the assigned weighting scores to different manifestation degrees ranging from 0 to 5. 0 is for a symptom that does not manifest in the case of a particular disease, so 0 weighting is ignored. 1 is the lowest weighting while 5 is the highest weighting. This weighting is used in the fitness evaluation of individual systems during optimization.

**Table 2: The Weighting Score for different Manifestation degrees**

| Manifestation degree (%) | Category | Label | Assigned Weighting scores |
|---|---|---|---|
| **'-' (negative)** | Extremely rare | xr | 0 |
| **<5%** | Rare | R | 1 |
| **5% to 35%** | Uncommon | U | 2 |
| **35% to 65%** | Common | C | 3 |
| **65% to 90%** | Very Common | V | 4 |
| **85% to 99%** | Almost all | A | 5 |

### *2.2 Methods*

#### *2.2.1 Optimization of the Symptoms Using Filtered Genetic Algorithm.*

**Problem Formulation**

The problem of determining the determinant symptoms for Malaria and Typhoid fever can be formulated into a mathematical model as follows:

Let s be a universal set of symptoms and m Manifestation degree of each symptom in a disease, d.

Indices:

Let i = 1 to 2{disease index, in this case}; j = 1 to n {symptom index}; k = 1 to 5 {manifestation degree index} and t = 1 to 2 {period index}

Decision variable

$$s_{ji} = \begin{cases} 1, & symptom\ j\ is\ manifested\ by\ disease\ i \\ 0, & else \end{cases}$$

Parameters

$n = number\ of\ symptoms$

$$p_{ijt} = \begin{cases} 1, & period\ t\ at\ which\ symptom\ j\ manifests\ in\ disease\ i \\ 0, & else \end{cases}$$

$$s_{jk} = \begin{cases} 1, & symptom\ j\ has\ manifestation\ degree\ k \\ 0, & else \end{cases}$$

$\epsilon_{ij} = error\ \ cost\ of\ disease\ (d_i) manifesting\ symptom\ (s_j)$

$a_{ji} = symptom\ j\ manifested\ by\ \ d_i\ within\ 2\ weeks\ of\ infestation.$

$b_{ji} = symptom\ j\ manifested\ by\ \ d_i\ after\ 2\ weeks\ of\ infestation.$

$w_{ijk} = manifestation\ degree\ \ of\ \ s_j\ \ by\ \ d_i\ within\ a\ given\ period\ of\ infestation.$

Note that the manifestation degree is the weighting score for the symptoms, the detail of which is shown in Table 2.

The objective Function is to select the diagnostic determinant symptoms by minimizing the quadratic error (loss) function which is a variation from the mean weighted score and is given thus:

$$f(w_E) = \frac{1}{n}\sum_{j=1}^{n}\left|(w_T - w_j)^2\right| \rightarrow min \qquad (1)$$

*where $w_T$ is a constant, the expected weighting score for a determinant symptom, $w_j$ is the weighting score of symptom $j$, $n$ is the number of symptoms for a disease $i$.*

Expanding; $(w_T - w_j)^2 = (w_T - w_j) * (w_T - w_j)$ (2)

$= w_T * w_T - w_T w_j - w_j w_T - (w_j * -w_j)$ (3)

$= {w_T}^2 - 2w_T w_j + {w_j}^2$ (4)

The quadratic error function becomes:

$$f(w_E) = \frac{1}{n}\sum_{j=1}^{n}\left|(w_T^2 - 2w_T w_j + w_j^2)\right| \quad (5)$$

If n = 1

$$f(w_E) = w_T^2 - 2w_T w_j + w_j^2 \quad (6)$$

Subject to:

Every symptom must be manifested in at least one disease

$$\sum_{j=f(s_j)} w_{ji} \geq 1 \; \forall i \quad (7)$$

1. Each symptom j that determines a disease i, must have a high weighted score of manifestation degree w

$$w_{jk} = \begin{cases} 1, & 3 < k \leq 5 \\ 0, & 0 < k \leq 3 \end{cases} \quad (8)$$

2. The highest manifestation degree is 100% and has the maximum weighting score which is the target score.

$$w_T = 5 \quad (9)$$

Constraint set (1) ensures that every symptom is manifested in at least one of the diseases while constraint (2) ensures that determinant symptoms' *Weighted score* (the actual manifestation of a symptom in a given sample) must be greater than 3 (i.e. above 65% ). The objective is to minimize the number of symptoms that can distinctly determine the presence of a particular disease (Malaria or Typhoid fever as is considered in this study)

### *2.2.2 Filtered Genetic Algorithm*

The Genetic Algorithm (GA) using the Filtered approach was employed in the optimization of the objective function to evolve the best symptoms for efficient distinctive diagnosis of Malaria and Typhoid fever by providing optimal inputs for the neuro-fuzzy classifier and 'determinant' symptoms for auto–rule generator. Filter, Wrapper and Embedded (Hybrid) models are three major feature selection approaches that have been intensively used for data dimension reduction in bioinformatics(Yang et.al. 2010). A brief definition of these approaches is given in this study as the detailed discussion is not within the scope of the study.

The filter approach requires the statistical analysis of the feature set only for solving the feature selection task without utilizing any learning model or classifier. It works fast using a simple measurement, though not satisfactorily.

The wrapper approach involves the predetermined learning model, and selects

features on measuring the learning performance of the particular learning model. The hybrid approach attempts to take advantage of the filter and wrapper approaches. It is often found that the hybrid technique is capable of locating a good lution, while a single technique often traps an immature solution.

Typically, filter-based algorithms do not optimize the classification accuracy of the classifier directly but attempt to select features with certain kinds of evaluation criteria. With the filter approach, the gene selection process and the classification process are separated, the advantages are that the algorithms are often fast and the selected genes are better generalized to unseen data classification. Since, at this point, this work is concerned with the only selection of the best symptoms for the diagnosis, the Filtered Genetic algorithm is employed for easy and fast selection.

### *2.2.3 Filtered Genetic Algorithm Parameters*

**Population**

The population contains a set of chromosomes; each chromosome is one complete possible solution to the problem to be solved with a genetic algorithm. The value of the function $f(w_E) for\ an\ s_i$as is shown in Table 3 is chromosome, the set of the chromosomes for each period of manifestation for Malaria and Typhoid fever is a population. All the symptoms are considered for the four cases, thus the population is set to 4N where N in this study is the number of symptoms (37).

Table 3 depicts the binary coding of the chromosomes. Encoding potential solutions (chromosomes) to the problem using a method that a computer can process is very essential in using the genetic algorithm to solve the problem. The most common approach is encoding solutions as binary strings of 1's and 0's where the digit at each position represents the value of some part of the solution and labelled as genes. For the symptoms selection in this study, the genes are represented by binary encoding where unsigned 5-bits are used to represent integers 0 to 25 representing the error functions of the individual symptoms (that is the deviation of their weighting scores ($w_j$) from the target weighting score($w_T$)). The valid genes fall within 00000-11001 respectively since $f(w_E)$is within 0 to 25.

*(Note: $M_a$, $T_a$ represent Early Malaria and Typhoid fever, and $M_b$, $T_b$ represents Late Malaria and Typhoid fever respectively)*

To convert the binary string back to its real value, the following procedure is followed:

Let Dj be equivalent to gene j, then gene j = {b1,b2,…bk} where bk is a binary substring which is either 1 or 0. The Decoded value:

$$DV = \sum_{l=1}^{k} (b_l x 2^{k-1}) \qquad (10)$$

Thus a binary string 00011 can be converted to its real value as:

$$0\ x\ 2^5 + 0\ x\ 2^4 + 0\ x\ 2^3\ x\ 2^2 + 1\ x\ 2^1 + 1\ x\ 2^0$$ which is equivalent to 3.

*Evaluation*

Each chromosome is evaluated using the fitness function

$$F = \frac{1}{1+f(w_E)} \qquad (11)$$

where $f(w_E)$ is the weighting score function The suitability of each chromosome to be selected for production is tested by subjecting it to fitness functions. The fitness value reflects the quality of each chromosome and is the bases for the selection of chromosomes that will be parents for new offspring.

***Selection***

Chromosomes are selected from the population to be parents to offspring. The problem is how to select these chromosomes. According to Darwin's evolution theory, the best ones should survive and create new offspring. The existing methods for selection of the best chromosomes, include among others roulette wheel selection, Boltzmann selection, tournament selection, rank selection, and steady-state selection (Alabsi and Naoum, 2012).

### 2.2.4 *Gene Encoding*

**Table 3: Symptoms encoded based on their corresponding error functions (The bit strings of the error functions are the chromosomes).**

| S_L | $w_j$ | $f(w_E)$ | Chrom | $w_j$ for $M_b$ | $f(w_E)$ | Chrom | $w_j$ for $T_a$ | $f(w_E$ | Chrom | $w_j$ - Tb | $f(w_E)$ | Chrom |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $S_1$ | 4 | 1 | 00001 | 0 | 25 | 11001 | 0 | 25 | 11001 | 0 | 25 | 11001 |
| $S_2$ | 0 | 25 | 11001 | 0 | 25 | 11001 | 4 | 1 | 00001 | 4 | 1 | 00001 |
| $S_3$ | 2 | 9 | 01001 | 0 | 25 | 11001 | 5 | 0 | 00000 | 0 | 25 | 11001 |
| $S_4$ | 4 | 1 | 00001 | 0 | 25 | 11001 | 4 | 1 | 00001 | 0 | 25 | 11001 |
| $S_5$ | 5 | 0 | 00000 | 2 | 9 | 01001 | 2 | 9 | 01001 | 3 | 4 | 00100 |
| $S_6$ | 3 | 4 | 00100 | 0 | 25 | 11001 | 5 | 0 | 00000 | 0 | 25 | 11001 |
| $S_7$ | 5 | 0 | 00000 | 0 | 25 | 11001 | 5 | 0 | 00000 | 0 | 25 | 11001 |
| $S_8$ | 4 | 1 | 00001 | 0 | 25 | 11001 | 4 | 1 | 00001 | 0 | 25 | 11001 |
| $S_9$ | 3 | 4 | 00100 | 0 | 25 | 11001 | 3 | 4 | 00100 | 0 | 25 | 11001 |
| $S_{10}$ | 3 | 4 | 00100 | 3 | 4 | 00100 | 5 | 0 | 00000 | 3 | 4 | 00100 |
| $S_{11}$ | 4 | 1 | 00001 | 4 | 1 | 00001 | 4 | 1 | 00001 | 4 | 1 | 00001 |
| $S_{12}$ | 3 | 4 | 00100 | 3 | 4 | 00100 | 3 | 4 | 00100 | 3 | 4 | 00100 |
| $S_{13}$ | 1 | 16 | 10000 | 3 | 4 | 00100 | 1 | 16 | 10000 | 3 | 4 | 00100 |
| $S_{14}$ | 1 | 16 | 10000 | 0 | 25 | 11001 | 4 | 1 | 00001 | 0 | 25 | 11001 |
| $S_{15}$ | 0 | 25 | 11001 | 0 | 25 | 11001 | 3 | 4 | 00100 | 0 | 25 | 11001 |
| $S_{16}$ | 5 | 0 | 00000 | 5 | 0 | 00000 | 0 | 25 | 11001 | 0 | 25 | 11001 |
| $S_{17}$ | 0 | 25 | 11001 | 0 | 25 | 11001 | 3 | 4 | 00100 | 0 | 25 | 11001 |
| $S_{18}$ | 0 | 25 | 11001 | 0 | 25 | 11001 | 3 | 4 | 00100 | 0 | 25 | 11001 |
| $S_{19}$ | 0 | 25 | 11001 | 0 | 25 | 11001 | 1 | 16 | 10000 | 3 | 4 | 00100 |
| $S_{20}$ | 0 | 25 | 11001 | 0 | 25 | 11001 | 4 | 1 | 00001 | 0 | 25 | 11001 |
| $S_{21}$ | 0 | 25 | 11001 | 0 | 25 | 11001 | 0 | 25 | 11001 | 0 | 25 | 11001 |
| $S_{22}$ | 0 | 25 | 11001 | 0 | 25 | 11001 | 1 | 16 | 10000 | 0 | 25 | 11001 |
| $S_{23}$ | 0 | 25 | 11001 | 0 | 25 | 11001 | 1 | 16 | 10000 | 3 | 4 | 00100 |
| $S_{24}$ | 0 | 25 | 11001 | 0 | 25 | 11001 | 0 | 25 | 11001 | 1 | 16 | 10000 |
| $S_{25}$ | 3 | 4 | 00100 | 3 | 4 | 00100 | 0 | 25 | 11001 | 0 | 25 | 11001 |
| $S_{26}$ | 3 | 4 | 00100 | 3 | 4 | 00100 | 0 | 25 | 11001 | 3 | 4 | 00100 |
| $S_{27}$ | 3 | 4 | 00100 | 3 | 4 | 00100 | 3 | 4 | 00100 | 0 | 25 | 11001 |
| $S_{28}$ | 0 | 25 | 11001 | 0 | 25 | 11001 | 4 | 1 | 00001 | 3 | 4 | 00100 |
| $S_{29}$ | 0 | 25 | 11001 | 0 | 25 | 11001 | 4 | 1 | 00001 | 0 | 25 | 11001 |
| $S_{30}$ | 0 | 25 | 11001 | 0 | 25 | 11001 | 1 | 16 | 10000 | 4 | 1 | 00001 |
| $S_{31}$ | 0 | 25 | 11001 | 1 | 16 | 10000 | 0 | 25 | 11001 | 0 | 25 | 11001 |
| $S_{32}$ | 0 | 25 | 11001 | 0 | 25 | 11001 | 3 | 4 | 00100 | 0 | 25 | 11001 |
| $S_{33}$ | 0 | 25 | 11001 | 0 | 25 | 11001 | 0 | 25 | 11001 | 0 | 25 | 11001 |
| $S_{34}$ | 4 | 1 | 00001 | 0 | 25 | 11001 | 0 | 25 | 11001 | 0 | 25 | 11001 |
| $S_{35}$ | 4 | 1 | 00001 | 0 | 25 | 11001 | 0 | 25 | 11001 | 0 | 25 | 11001 |
| $S_{36}$ | 0 | 25 | 11001 | 0 | 25 | 11001 | 3 | 4 | 00100 | 0 | 25 | 11001 |
| $S_{37}$ | 0 | 25 | 11001 | 1 | 16 | 10000 | 1 | 16 | 10000 | 0 | 25 | 11001 |

****Chrom = Chromosomes, $w_j$ = $w_j$ for $M_a$, S-L = S_Label**

Roulette wheel selection (RWS) which selects parents based on their fitness is employed. RWS is a common selection approach that assigns a probability of selection Pj to each j based on its fitness value. The probability of an individual $s_j$ to be a member of the next generation at each iteration is proportional to its fitness value *F* and is calculated thus:

Step 1: Finding of the fitness value of each chromosome in the population using the fitness function as is shown in Equation 12

$$F_j = \frac{1}{1+f(w_E)} \quad (12)$$

Step 2: Calculation of the sum of fitness for all chromosomes in the population using Equation 13

$$f_T = \sum_{j=1}^{n} F_j \quad (13)$$

Where n is the number of chromosomes in the initial population.

Step 3: Calculation of the average fitness of the chromosomes in the population ($f_{AVG}$) using Equation (14)

$$f_{AVG} = \frac{f_T}{n} \quad (14)$$

Step 4: The expected fitness $F$ for each chromosome which is the probability of a chromosome being selected is calculated using Equations 15.

$$eF_j = \frac{F_j}{f_{AVG}} \quad (15)$$

Thus $P_j = eF_j$

An individual (chromosome) $s_j$ is selected for the next generation if its probability $P_j > 0$. The summary of chromosomes' selection using RWS is shown in Table 4.

**Table 4: Summarized Evaluation of initial population**

| Chromosome ID | Weighting score, $w_j$ | Init. P Chromosomes | Fitness($F_j$) $F = \frac{1}{1+f(w_E)}$ | $eF_j = \frac{F_j}{f_{AVG}}$ | Expected count for a chromosome |
|---|---|---|---|---|---|
| **1** | 4 | 00001 | 0.5000 | 2.2262 | 2 |
| **2** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **3** | 2 | 01001 | 0.1000 | 0.4452 | 0 |
| **4** | 4 | 00001 | 0.5000 | 2.2262 | 2 |
| **5** | 5 | 00000 | 1.0000 | 4.4525 | 4 |
| **6** | 3 | 00100 | 0.2000 | 0.8905 | 1 |
| **7** | 5 | 00000 | 1.0000 | 4.4525 | 4 |
| **8** | 4 | 00001 | 0.5000 | 2.2262 | 2 |
| **9** | 3 | 00100 | 0.2000 | 0.8905 | 1 |
| **10** | 3 | 00100 | 0.2000 | 0.8905 | 1 |
| **11** | 4 | 00001 | 0.5000 | 2.2262 | 2 |
| **12** | 3 | 00100 | 0.2000 | 0.8905 | 1 |
| **13** | 1 | 10000 | 0.0588 | 0.2619 | 0 |
| **14** | 1 | 10000 | 0.0588 | 0.2619 | 0 |
| **15** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **16** | 5 | 00000 | 1.0000 | 4.4525 | 4 |
| **17** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **18** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **19** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **20** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **21** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **22** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **23** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **24** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **25** | 3 | 00100 | 0.2000 | 0.8905 | 1 |
| **26** | 3 | 00100 | 0.2000 | 0.8905 | 1 |
| **27** | 3 | 00100 | 0.2000 | 0.8905 | 1 |
| **28** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **29** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **30** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **31** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **32** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **33** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **34** | 4 | 00001 | 0.5000 | 2.2262 | 2 |
| **35** | 4 | 00001 | 0.5000 | 2.2262 | 2 |

| | | | | | |
|---|---|---|---|---|---|
| **36** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| **37** | 0 | 11001 | 0.0385 | 0.1712 | 0 |
| $f_T$ **= 8.3100** | | | | | |
| $f_{AVG}$ **=0.2246** | | | | | |
| $eF_T$ **=37.0000** | | | | | |

In Table 4 which summarizes the initial population evaluation, the Expected count shows the number of a chromosome that can be selected for the production of offspring that will make up the population for the next generation. The integer part of the expected fitness determines the number of chromosomes while the fraction part shows its chance of being repeated. For instance, chromosome 00001 with expected fitness $eF_j$ =2.2262 has a chance of being selected twice while stringing 11001 with expected fitness $eF_j$ =0.1712 has zero (0) chance of being selected. A new generation of solutions is produced by picking from existing chromosomes with a preference for ones that are more significant than others.

Fig. 1 depicts the count for a particular chromosome (symptom) in the next population for reproduction. From the graph, it is observed that those with higher values of expected fitness value have the opportunity of being selected more times than those with low expected fitness value.

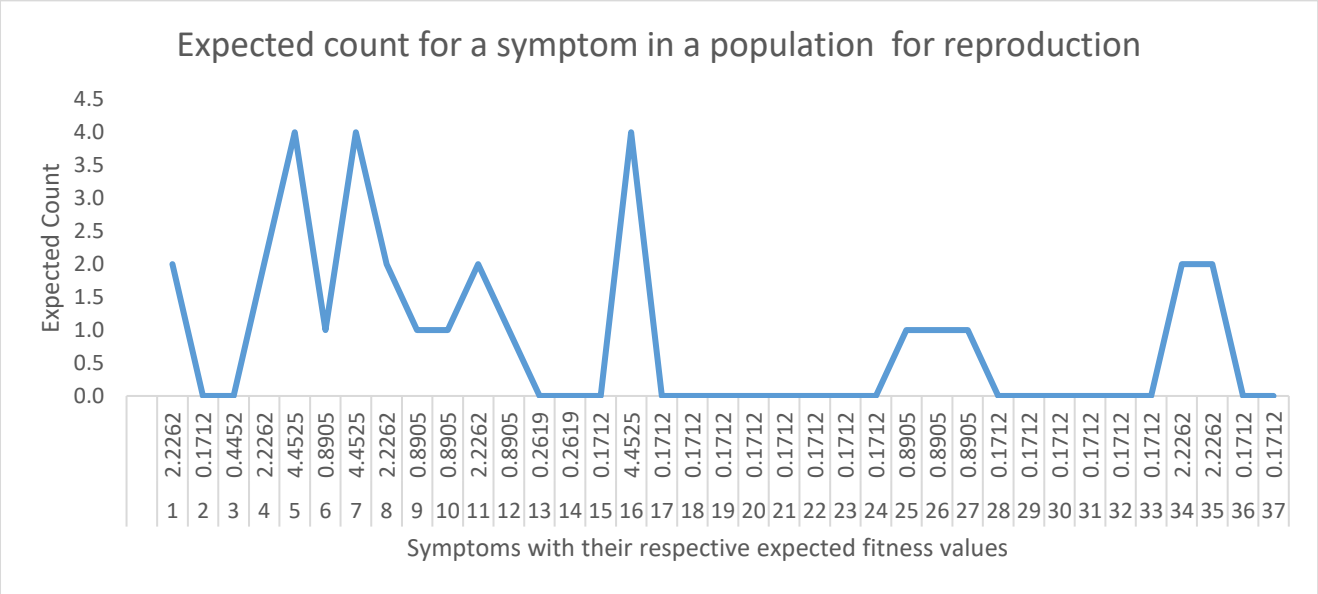


**Fig. 1: Graph of expected count for a symptom in a population for reproduction**

### A. Recombination

Recombination is the process by which selected chromosomes from a source population are recombined to form members of a successor population. This parameter simulates the mixing of genetic material that can occur when organisms reproduce. Two main components of recombination are the genetic operators: *Crossover* and *Mutation*.

*Crossover*

The crossover operator represents the mixing of genetic material from two selected parent chromosomes to produce one or two child chromosomes by combining the information extracted from the parents. To employ a single point crossover, for a chromosome of length L, a random number c between 1 and L is first generated. The first child chromosome is formed by appending the last L−c elements of the first parent chromosome to the first c elements of the second parent chromosome. The second child chromosome is formed by appending the last L−c elements of the second parent chromosome to the first c elements of the first parent chromosome. Typically, the probability for crossover ranges from 0.6 to 0.95. Crossing these two strings 10000 and 01100 yields:

10000 → 10000 → 10100

01100→ 01100 → 01000

### B. Mutation

Mutation operators act on a child chromosome to flip one or more allele values. In the case of bit-string chromosomes, the normal mutation operator is applied to each position in the chromosome. A random number in the interval [0,1] is generated with uniform probability and compared to a pre-determined "mutation rate". If the random number is greater than the mutation rate, no mutation is applied at that position. If the mutation rate is greater than or equal to the random number, then the allele value *is flipped from 0 to 1 or vice versa. The mutation rate applied in this study is 0.01*

### C. Evolution

After recombination, resultant chromosomes are passed into the successor population. The processes of selection and recombination are then repeated until a complete successor population is produced. At that point, the successor population becomes a new source population (the next generation). The GA is iterated through several generations until convergence to a best-fitness s solution is observed.

The activity diagram depicts all the activities involved in the optimization process which have been discussed in detail in subsection 3.2.(iii)

### *2.2.4 The Genetic Algorithm activity diagram.*

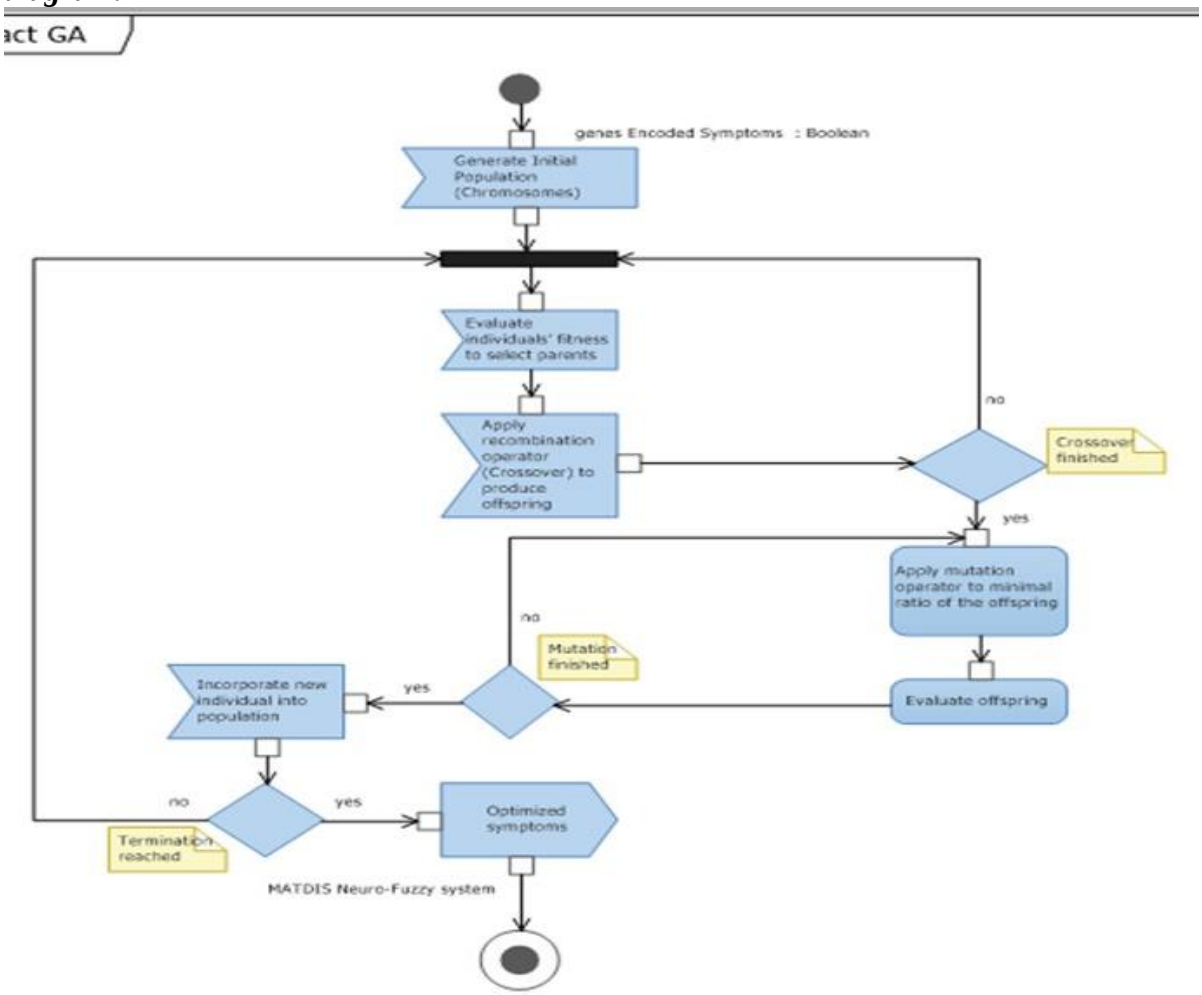


***Fig. 2: Pre-processor (GA) activity diagram.***

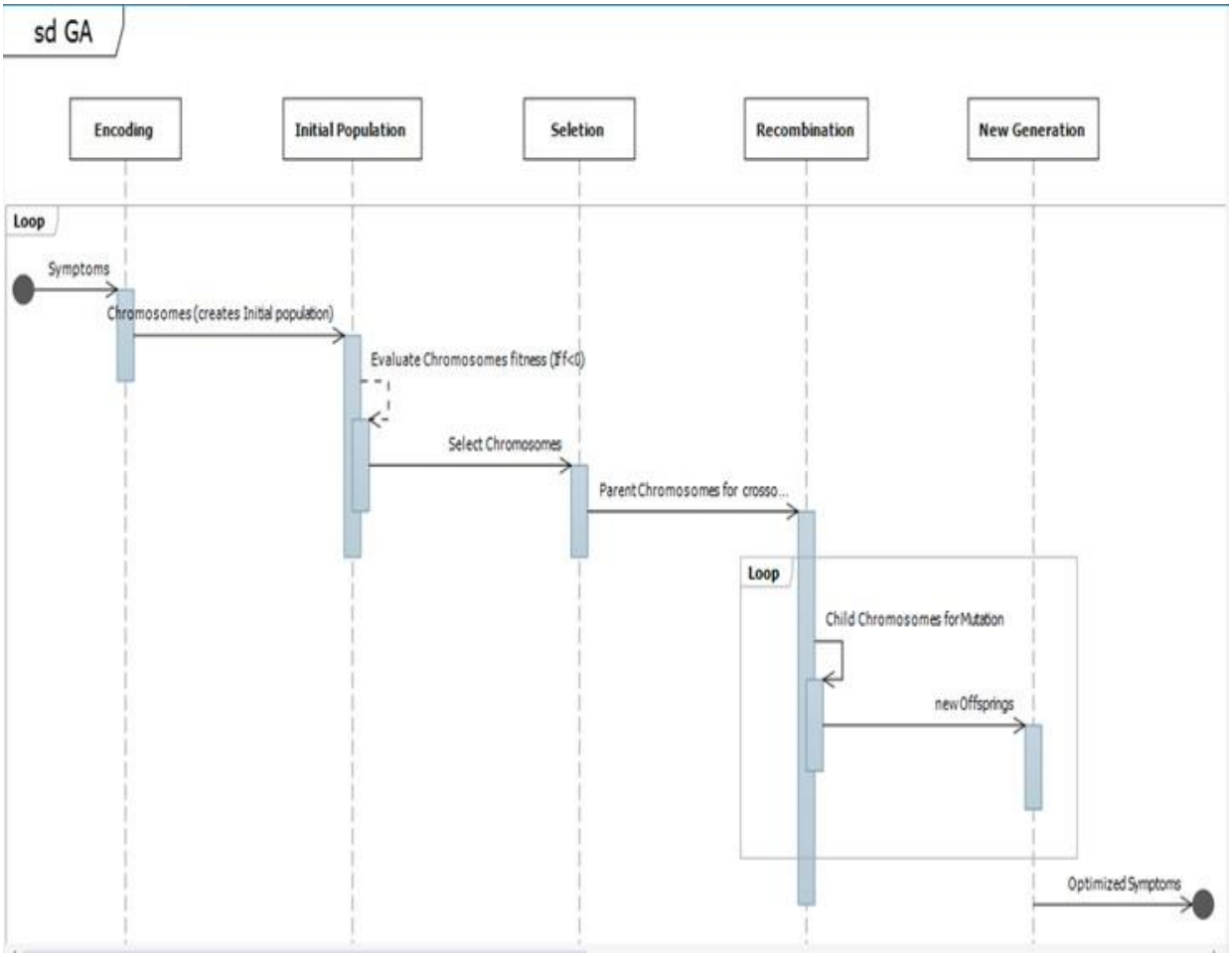


***Fig. 3: GA sequence diagra***

**2.2.5 The Genetic Algorithm Sequence diagram**.

The diagram in Fig. 3 depicts the interactions between the different GA components discussed above to achieve the system's requirement which is the production of diagnostic determinant symptoms for Malaria and Typhoid fever. The sequence diagram shows the distribution of optimization tasks between the components.

### *2.2.6 The Pre-Processing (using GA) Algorithm*

Below is the algorithm (genetic algorithm) for the pre-processing.

*A: Declare arrays of symptoms, (S), period (t) and weighting, (w), disease (d) and variables m for manifestation degree, f, Fval,* $F_T$

*Fval* = $1/(1 + F(w_E))$

$F(w_E) = w_T^2 - 2w_T w_j + w_j^2$ *//objective function to be minimized//*

*Let* $w_T = 5$ *//target weight//*

*Let n be a number of symptoms*

*For( d = 1 to 2: t = 1 to 2: j = 1 to n)*

*Input S(j) //enter symptoms for each disease at a period of infestation//*

*Input m(j) // degree of occurrence of a symptom (in a disease) in percentage//*

*If m(j) < 5% then w(j) = 1 else if m(j) > 5% < = 35% then w(j) = 2: else if m(j) > 35% < = 65% then w(j) = 3: else if m(j) > 65% < = 90% then w(j) = 4: else if m(j) > 90% then w(j) = 5*

*End if*
*Next j: t: d*
*Initialization*
*Let strength and posize denote the binary bits sequence and a number of individuals (chromosomes) in the population, let [a,b] be domain of weighting score for symptom j.*
*The data structure of the population is matrix of po * (strlength) +2)*
*For specified number of generations do*
*B:*$function\ [po] = initialise\ (posize, strlength, Fval)$;
*//Recombination: Crossover and Mutation operators//*
*Crossover*
*Let pc be the probability of crossover*
*Function [child1, child2] = crossover (parent1, parent2, pc),*
*Generate random number r [1,strlength]*
*If (r <pc)*
*cpt = strlength – r //cpt is the crossover point//*
*Child1 = Concatenate (strlength – r) elements of parent1with r elements of parent2;*
*Child2 = Concatenate (strlength – r) element of parent2 with r elements of parent1;*
*Else*
*Child1 = parent1*
*Child2 = parent2*
*end*
*//Mutation operation on the child chromosomes using flip method//*
*Mutation*
*Let pm be mutation probability,* $pm \in [0,1]$*; set pm to 0.01*
*Declare Function [child] = mutation (parent, pm);*
*For i = 1 to n //number of child chromosomes//*
*Generate random number, v*
*If v < = pm*
*mp = rand[1, strlength] //mp mutation point//*
*Child = parent;*
*Child[mp] = abs(parent[mp] -1);*
*Child(:, strlength + 1)= sum(2 ^ (size(child):,1:strlength),2)-1:-1:0)* child(:,1:strlength))*(b-a)/ (2^strlength -1)+a; child(:, strlength +2) = Fval(child(:,strlength + 1))*
*Else: child = parent;*
*End: End*
*Next i*
*Declare Function[newpo] = roulette[oldpo];*
$f_{T\ ;} = sum(oldpo(:, strlength + 2))$ *//calculation of total finess//*
$pr = (oldpo(:, strlength + 2))/f_T$ *// calculation of probability of each chromosome being selected//*
$pr = cumsum(pr); generate\ random\ number, rn;$
$rn = sort(rand(posize, 1); : fitin = 1; newin = 1;$
$while\ newin \leq posize:$
$if\ \big(rn(newin) < pr(fitin)\big): newpo(newin) = oldpo(fitin): newin = newin + 1;$
$else\ fitin = fitin + 1$
*end*

*Generate P(z+1): set z to z+1*
*REPEAT the process from B UNTIL optimum symptoms is exhausted*
*End process.*
*Return Optimized symptoms as the determinant symptoms*

**Table 5: Genetic Algorithm Components specification**

| Component | Value |
|---|---|
| **Search Method** | GA |
| **Population Size** | 37 |
| **Encoding** | Binary coding |
| **Evaluation** | Fitness function |
| **Selection** | RWS |
| **Crossover function** | Single point (0.8) |
| **Mutation rate (Flip bit)** | 0.01 |
| **Stopping Criterion** | Till convergence to best solution is observed |
| **Generation** | 50 |

## 3.0 Results and Discussion

The symptoms were optimized with the genetic algorithm (GA) tool in MatlabR2013a using the components specified in Table 5 as is shown in Fig. 4.

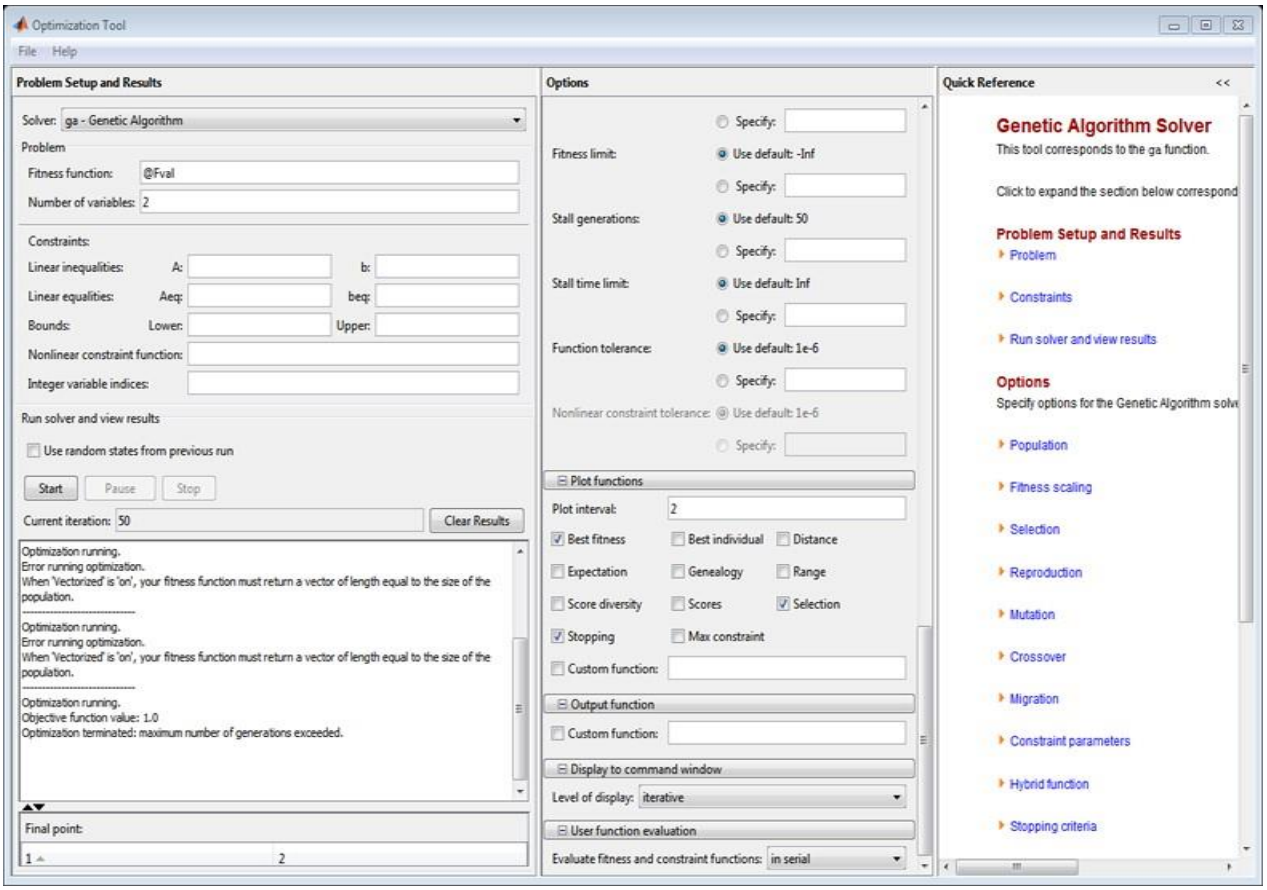


***Fig. 4: Optimization using GA tool.***

Fig. 5 shows the graph of the highest fitness value as 1, the number of children for each population and the stopping criteria which were met at 50$^{th}$ generation.

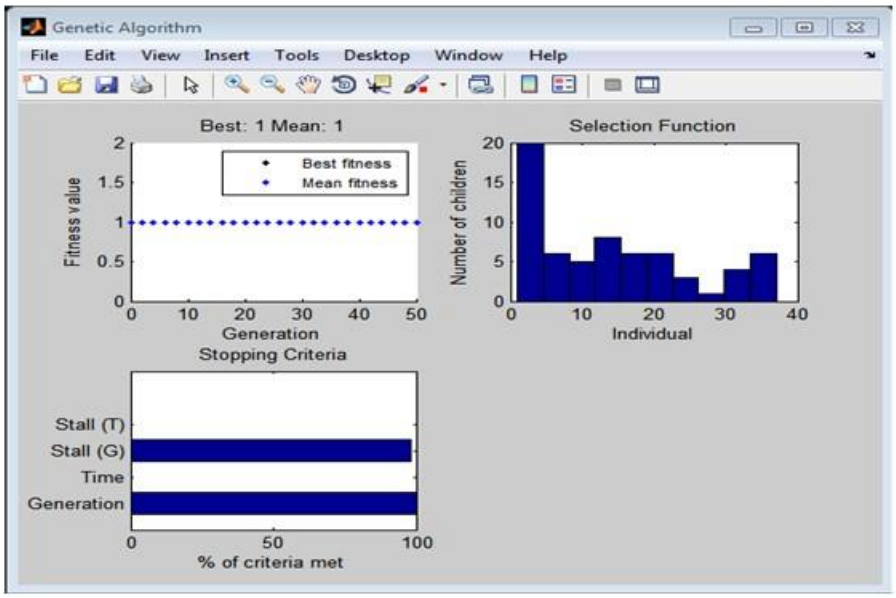


**Fig. 5: The graph of Genetic Algorithm output**

Table 6 shows the predominant symptoms manifested by Malaria and Typhoid fever during early and later (less than and greater than 2 weeks) periods of infestation respectively.

**Table 6: shows the predominant symptoms manifested by malaria and 'typhoid fever**

| **Disease** | **≤ 2 weeks** | | **> 2 weeks** | |
|---|---|---|---|---|
| | **Label** | **Symptom** | **Label** | **Symptom** |
| **Malaria** | $S_1$ | High fever | | Symptom |
| | $S_4$ | Diaphoresis | | |
| | $S_5$ | Rigors | $S_{16}$ | |
| | $S_7$ | Lethargy | | Cough |
| | $S_{16}$ | Cough | | |
| | $S_{27}$ | Jaundice | | |
| | $S_{34}$ | Myalgia | | |
| | $S_{35}$ | Arthralgia | | |
| | $S_2$ | Stepwise fever | | |
| | $S_3$ | Chills | | |
| | $S_4$ | Diaphoresis | $S_2$ | Stepwise fever |
| | $S_6$ | Anorexia | $S_{11}$ | Intestine hemorrhage |
| | | | $S_{30}$ | Intestine hemorrhage |
| **Typhoid Fever** | $S_8$ | Insomnia | | |
| | $S_{10}$ | Malaise | | |
| | $S_{14}$ | Delirium | | |
| | $S_{20}$ | Coated tongue | | |
| | $S_{28}$ | Constipation | | |
| | $S_{29}$ | Bloating | | |

### 3.1. Degree of Optimization

This is the measure of how much the symptoms were optimized. This optimization process aims to reduce the number of symptoms needed for effective diagnosis of Malaria and Typhoid fever by producing the best features known as the determinant symptoms for each disease at a given infestation period.

The degree of optimization was calculated using Relative change.

$$Relative\ change = \frac{New\ value - Reference\ value}{Reference\ value} \quad (1)$$

Let $B_o$ (number of manifested symptoms) be the reference value, and $A_o$ (number of selected symptoms, the determinant symptoms after the optimization process) be the new value.

$$Relative\ change(A_oB_o) = \frac{A_o - B_o}{B_o} \quad (2)$$

Let the degree of optimization $d_{o(ip)}$ for each manifestation period of a disease, $d_i$ be equivalent to the percentage of absolute Relative change at each period for a disease $d_i$.

$$D_{o(ip)} = \left|A_{o(ip)} - B_{o(ip)}\ /B_{o(ip)}\right| * 10 \quad (3)$$

The total degree of optimization, $D_o$ is the percentage absolute change in the total number of all manifested symptoms for both Malaria and Typhoid fever after optimization and is given by

$$D_o = |\ B_o - A_o\ /B_o| * 100 \quad (4)$$

**Table 7: The Optimization degree**

| Case | No of symptoms before optimization $B_o$ | No of ymptoms after optimization $A_o$ | Degree of optimization at p for $d_i$ % $D_{oip} = \left\|\frac{A_{oip} - B_{oip}}{B_{oip}}\right\| \times 100$ | Overall degree of optimization % $D_o = \left\|\frac{A_o - B_o}{B_o}\right\| \times 100$ |
|---|---|---|---|---|
| **$M_a$** | 18 | 9 | 50.00 | |
| **$M_b$** | 11 | 1 | 90.91 | |
| **$T_a$** | 27 | 10 | 62.96 | 64.06 |
| **$T_b$** | 8 | 3 | 62.50 | |
| **$S_T$** | 64 | 23 | 64.06 | |

Table 7 summarizes the calculation of the degree of optimization considering the manifested symptoms at each period of infestation. $M_a$, $M_b$ represent early and late Malaria, $T_a$ and $T_b$ represent early and late Typhoid fever respectively while $S_T$ is the total number of symptoms. The overall degree of optimization is the percentage ratio of change in the number of symptoms after optimization to the total number of symptoms before optimization.

## 4.0 Conclusion

This paper presented symptoms optimization using GA to obtain the best discerning symptoms ('determinant symptoms') used to design a diagnosis system for distinctive diagnosis of Malaria and Typhoid fever. The determinant symptoms can also be used in rural areas where the syndromic diagnosis of malaria and typhoid fever is still prevalent, though is not the best practice but most of these people do not have access to adequate medical care, equipped laboratories and skilled personnel. The determinant symptoms can be a guide to their health care officers and also reduce misappropriation of treatment. Though the conventional method is the best for diseases diagnosis, it is not always available, especially in rural areas where many depend on low-skilled medical practitioners for their health care. The use of these optimized determinant symptoms in the syndromic diagnosis of Malaria and Typhoid fever will reduce the risk of misdiagnosis of these two diseases.

## 5.0 Acknowledgment

We heartily appreciate all who contributed in one way or the other towards the successful completion of this study especially the medical officers at Delta State University Teaching hospital, Oghara and Delta State University Medical Centre, Abraka. May God reward you abundantly.

## 6.0 References

Adehor, A.B., & Burrell, P. (2008). The Integrated management of health care strategies and differential diagnosis by expert system technology: a single-dimensional approach. World Academy of Science, Engineering and Technology, 44, pp. 533-538.

Adehor, A.B. and Burrell, P.R., (2008b), *An Intelligent decision support system for the prompt diagnosis of malaria and typhoid fever in the Malaria Belt of Africa in IFIP*. Conference: Artificial Intelligence in Theory and Practice II, IFIP 20th World Computer Congress, TC 12: IFIP AI Stream, September 7-10, , Milano, Italy, pp. 287–296

Adetunmbi A.O, Oguntimilehin A., and Falaki S.O , (2012). "Web-based medical assistant system for Malaria diagnosis and therapy", GESJ: *Computer Science and Telecommunications*, 1, 33, , pp 42-53.

Alabsi F. and Naoum R. (2012). Comparison of Selection methods and crossover operations using Steady State Genetic Based Intrusion Detection System. *Journal of Emerging Trends in Computing and Information Sciences* 3, 7, , *pp 1053-1057.*

Agarkar A.M. & Ghatol, A.A (2010); FFANN based Cost Effective Major Infant disease Management. *International Journal of Computaer Application, 7, 11,* doi: 10.5120/1289-1755

. Guerrero C., Lera I. & Juiz C. (2022). Genetic-based optimization in fog computing: Current trends and research opportunities. *Swarm and Evolutionary Computation* 72, https://doi.org/10.1016/j.swevo.2022.101094.

Chiemeke, S., & Omede, E.U. (2014). Mal-typho diagnosis intelligent system (matdis): the auto-diagnostic rule generation algorithm. *Computing, Information Systems, Development Informatics & Allied Research Journal, 5, 4, pp. 83-92*

Davis C.P.(2022) Is Malaria Contagious? https://www.medicinenet.com/is_malaria_contagious/article.htm#what_is_malaria. viewed 7/09/2022

Ding, X., Zheng, M. & Zheng, X. (2021). The Application of Genetic Algorithm in Land Use Optimization Research: 10, 526. https://doi.org/10.3390/land10050526

Djam, X. Y., Wajiga, G. M., Kimbi, Y. H. & Blamah N.V. (2011). A Fuzzy expert system for the management of malaria. *International Journal of Pure and Applied Sciences and Technology* 5, 2, pp. 84-108.

Haibo, W. (2022). Research on the Application of Genetic Algorithm in Physical Education Journal of Mathematics 8477945,. https://doi.org/10.1155/2022/8477945

Amjad, M. K., ,Butt S.I., Kousar R., Ahmad, R., Agha, M. H., Faping Z. Anjum,N. & Asgher, U. (2018). Recent Research Trends in Genetic Algorithm Based Flexible Job Shop Scheduling Problems. *Mathematical Problems in Engineering, 9270802, https://doi.org/10.1155/2018/9270802*

Nas, F. S., Ali, M. & Yahaya, A. (2018). Malaria and typhoid fever co-infection among febrile patients in Kumbotso Local Government Area Kano, Nigeria.

*Bayero Journal of Pure and Applied Sciences,* 10, 2, pp. 122-125. http://dx.doi.org/10.4314/bajopas.v10i2.21

Odikamnoro, O. O., Ikeh ,I. M., Okoh, F.N., Ebiriekwe ,S.C. , Nnadozie I.A., Nkwuda, J.O & Asobie, G. C. (2018).Incidence of malaria/typhoid co-infection among adult population in Unwana Community, Afikpo North Local Government Area, Ebonyi State, Southeastern *Nigeria. Afr J Infect*

*Dis*., 12, 1, pp. 33-38. doi: 10.21010/ajid.v12i1.6

Oguntimilehin, A., Adetunmbi A.O. & Abiola O. B. (2013). A Machine Learning Approach to Clinical Diagnosis of Typhoid Fever. *International Journal of Computer and Information Technology*, 2, 4

Ohanu, M. E., Mbah, A.U., Okonkwo, P. O. & Nwagbo, F.S..(2003). Interference by malaria in the diagnosis of typhoid using Widal test alone. *West Africa Journal of Medicine,* , 22, 3, pp, 250–252doi: 10.4314/wajm.v22i3.27961

Olabiyisi, S.O Omidiora, E.O Olaniyan O.M & Derikoma,O. (2011). A Decision support system model for Diagnosing tropical diseases using Fuzzy logic. *African Journal of Computing & ICT* . 4, 2., pp. 1-6.

Samuel, O.W. & Omisore, M. O.(2013). Hybrid Intelligent System for the Diagnosis of Typhoid Fever. *Journal of Computer Engineering &Information Technology*, 2, 2, http://dx.doi.org/10.4172/2324-9307.1000109

Simon-Oke, I., & Akinbote, M. (2020). Prevalence of malaria and typhoid Coinfection in relation to haematological profile of University Students in Akure, Nigeria. *J Infect Dis Epidemiol, 6, 166. doi.org/10.23937/2474-36658/1510166*,

Uneke C. J (2008). Concurrent malaria and typhoid fever in the tropics: the diagnostic challenges and public health implications. .*J Vector Borne Dis* 45, pp. . 133-142

World Health Organization (2018). *World Malaria Report. Federal Republic of Ethiopia Ministry of Health*, National Guide Lines (3rd edition)

Yang, C., Chuang, L. & Yang C. (2010). IG-GA: A hybrid filter/wrapper method for feature selection of microarray data. J. Med. Biol. Eng., 30, 1, doi: 10.1109/ICSMC.2008.4811698

**Consent for publication**

Not Applicable

**Availability of data and materials**

The publisher has the right to make the data public

**Competing interests**

The authors declared no conflict of interest. This work was carried out in collaboration among all authors.

**Funding**

There is no source of external funding

**Authors' contributions:**

Edith Omede conceived the idea, designed the model and the computational framework and evaluated the result.

Stella Chiemeke supervised the work from start to completion, did necessary corrections.