Journal of Computer Sciences and Applications, 2018, Vol. 6, No. 2, 82-90
Available online at http://pubs.sciepub.com/jcsa/6/2/5

DOI:10.12691/jcsa-6-2-5

Improved Early Detection of Gestational Diabetes via Intelligent Classification Models: A Case of the Niger Delta Region in Nigeria

A.A. Ojugo*, D. Otakore

Department of Mathematics/Computer, Federal University of Petroleum Resources Effurun, Nigeria
*Corresponding author: ojugo.arnold@fupre.edu.ng

Received July 08, 2018; Revised August 17, 2018; Accepted September 05, 2018

Abstract Diabetes is the most common endocrine-metabolic disorder that features a body characterized by hyperglycaemia – giving rise to risk of microvascular (retinopathy, neupathy and nephropathy) and microvascular (vascular disease, stroke and ischemic heart diseases). Nigeria has become aware of inherent threats of the Type-II diabetes and the consequent metamorphism into gestational diabetes in mothers with or without previous cases of Type-II. We presents a comparative study of classification models using both the supervised and unsupervised evolutionary models. We aim at improved early detection of the disorder via data-mining tools. Adopted dataset is from College of Health and Teaching Hospitals with selected Universities in Niger Delta. Results show that age, body mass index, family ties to second degree, environmental conditions of inhabitance among others are critical factors that increases its likelihood. Gestational diabetes in mothers were confirmed if: (a) history of babies weighing > 4.5kg at birth, (b) insulin resistance with polycystic ovary syndrome, and (c) abnormal tolerance to insulin.

Keywords: *Diabetes, microvascular, macrovascular, fuzzy classifier, DiabCare, Gestational, hyperglycaemia*

Cite This Article: A.A. Ojugo, and D. Otakore, "Improved Early Detection of Gestational Diabetes via Intelligent Classification Models: A Case of the Niger Delta Region in Nigeria." *Journal of Computer Sciences and Applications*, vol. 6, no. 2 (2018): 82-90. doi: 10.12691/jcsa-6-2-5.

1. Introduction

Diabetes mellitus has now become a general chronic disease that affects about 6% of the global population – so that its avoidance and early detection for effective treatment has become imperative and undoubtedly a critical task for health and economic issue in 21st century. Diabetes is a metabolic disease that is characterized by the presence of hyperglycemia or high blood glucose. This result from the body's inability to secrets enough insulin that the body requires for glucose processing as a byproduct of the carbohydrate that we eat, or that the body is resistant to the effects of insulin. Thus, the reason why it is popularly named the *silent killer*. Glucose, as a main source of energy for cells that makes up the muscles and other tissues, is produced from the ***food*** we eat and in our ***liver***. Sugar (or glucose) is absorbed in the bloodstream and enters into a cell by the help of insulin. Liver stores glucose as glycogen so that if glucose becomes low, the liver reconverts stored glycogen into glucose to normalize glucose level [1]. Diabetes is a diagnosis from glycemia, associated with microvascular disease [2].

Recent estimates indicate about 212million people projected worldwide with diabetes in the year 2013, which will increase to about to 366 million by 2030. This increase in prevalence is expected to be more in the Middle Eastern crescent, Sub-Saharan Africa and India. In Africa, the estimated prevalence of diabetes is 1% in rural areas, about 7% in urban sub-Sahara Africa, and between 8-13% in developed areas like South Africa and India. In Nigeria, it varies from 0.65% in rural Mangu to 11% in urban Lagos. And dataset from the World Health Organization (WHO) suggests that Nigeria has the greatest number of people living with diabetes in Africa.

Diabetes is associated with range of complications such as risk of blindness, blood pressure, heart and kidney diseases, and nerve damage to mention a few [3,4]. Its early detection is extremely difficult by experienced physicians, and thus – led to a continued quest for methods to effectively and precisely classify the disease [5]. Ojugo et al [6] Various models have been used for its early detection and identification to include: (a) ***supervised*** classification in which its input variables for the diagnosis are ***known***), and (b) ***unsupervised*** classification in which the variables used for diagnosis and classification are ***unknown***). In both instance, a ***critical*** feat in selecting the appropriate classification model to use is, its accuracy and precision ability in classifying the task at hand.

1.1. Types of Diabetes

Diabetes is generally classified into [7]:

a. Type-I is a chronic state where pancreas secrets little or no insulin. Thus, sugar builds up in the bloodstream to cause real life threats. Type-1 has no cure as its causes are unknown, and it is insulin-dependent. Symptoms include: blurred vision, extreme hunger, increased thirst, fatigue, mood changes, irritability, incessant urination, unintended weight loss, vaginal yeast infection (in females) etc. Some of its known risk factors include: genetics, family history, age, exposure to bacteria and Epstein-Barr virus, early exposure to cow milk, low vitamin D, introduction to cereal/gluten in baby diet, intake of nitrate-contaminated water, mothers with preeclampsia at pregnancy and babies born with jaundice [3,8].

b. Type-II (noninsulin-dependent) is chronic state that affect how a body metabolizes sugar. It develops slowly in a body that either resists the effects of insulin produced by a body, or the body does not produce enough insulin to maintain normal glucose level. It is common in adults and in obese children. While, there is no cure for type-II, it is managed via proper eating habits, exercising, maintaining a healthy weight and sometimes, diabetes medications or insulin therapy. Its symptoms are increased thirst/hunger, weight loss, frequent urination, fatigue, blurred vision, acanthosis nigricians (areas of darkened skin) amongst others [3,9,10]. Chinenye and Young [11] Type-II has asymptomatic preclinical phase, which is not benign and thus, underscores the need for primary prevention and population screening in order to achieve early diagnosis and treatment.

1.2. Gestational Diabetes Diagnosis: The Nigerian Scenario

Gestational Diabetes Mellitus (GDM) is defined as disorder of glucose tolerance occurring first in pregnancy in mothers – whereas, some experts have viewed and believe GDM to be of same entity with Type-II – wherein the former constitutes the early signs and manifestation of the latter. GDM is endemic around the world and its prevalence differs from one region to another. Its risk factors include: family history of DM in first-degree relatives, child bearing with congenital anomaly, baby weighs more than 4000g or more, dying of unknown causes at birth, obesity, age greater than 35years amongst other. Various techniques are available to diagnose GDM as to what to test, when to perform such tests and what method is best. Most authors continually favor the early weeks of third trimester (between 26-to28 weeks) of pregnancy as best time to screen for GDM. Its investigations can be divided into ***screening*** and ***definitive*** tests [12]. The risk factors can be seen in the Table 1.

Type-II diabetes has asymptomatic preclinical phase that is not benign. It underscores the need for primary prevention and population screening in order to achieve early diagnosis and treatment. Reported undiagnosed diabetes have a prevalence as much as 18.9% – whereas Nyenwe et al (2003) reported a 2.8% rate of disease in Port Harcourt and Ajufo et al, [4] reported a rising rate of 5.64% in Agbor, Warri (Delta State) and Yenegoa (Bayelsa State). Arije et al [13] concurs with a satisfactory systolic and diastolic blood pressure control was obtained in only 38.5% and 42.2% of some Nigerian patients attending a tertiary health facility, respectively.

Diabcare Nigeria in 2008 took a sample study conducted across 7-tertiary health centers in Nigeria with the objective of assessing clinical and laboratory profile, evaluating the quality of care of Nigerian diabetics with a view to planning improved diabetes care. Clinical parameters studied: diabetes types, anthropometry, blood pressure, chronic complications therein and treatment types. Laboratory data assessed: fasting plasma glucose (FPG), 2 Hour post-prandial (2-HrPP), glycated haemoglobin (HbA1c), urinalysis, serum lipids, electrolytes, urea and creatinine. Total of 531 patients, 209(39.4%) males and 322(60.6%) females enrolled. Results shows mean age of patients is 57.1±12.3years with mean duration 8.8±6.6years. A majority (95.4%) had Type-II diabetes compared to Type-I (4.6%) via a $p<0.001$. Mean FPG, 2-HrPP glucose and HbA1c were noted at 8.1±3.9mmol/L, 10.6±4.6mmol/L and 8.3±2.2% respectively. Only 170 (32.4%) male and 100 (20.4%) female patients reached ADA and IDF glycaemic targets respectively. 72.8% patients did not practice self-monitoring of blood glucose and hypertension is found in 322 (i.e. 60.9%) patients, with a mean systolic BP of 142.0±23.7mmHg and mean diastolic BP of 80.7±12.7mmHg [11].

Its complications include: peripheral neuropathy 59.2%, cataracts 25.2%, cerebrovascular disease 4.7%, retinopathy 35.5%, nephropathy 3.2% and diabetic foot ulcer 16.0%. It is obvious that the status of Diabetes Care in terms of glycaemic control, control of cardiovascular risk factors, management practices and presence of late complications of diabetes were below the optimum expected; And most screening conducted in pursuance of early detection that are based on risk factors have been found to be insensitive as well as resulted in an increased false positives rates of methods adopted for these test as a little above 40% of these cases are missed. Also, no screening method is consistently reliable. Thus, the rationale for this study to early detect GDM in mother as maternal mortality has been seen to be on increase [3,7].

The ***rationale*** is to advance the early diagnosis and detection of Type-II and GDM in mothers via intelligent classification (supervised and unsupervised) model. Study will propagate observed data as input – as models seek to uncover the stochastic feats of interest to yield an output guaranteed of high quality and void of ambiguities. These models, further tuned can become robust and perform quantitative processing to ensure qualitative knowledge and experience, as its new language [14,15].

2. Materials and Methods

2.1. Dataset Used

Some statistical information of attributes is given in Table 1. The data set consists of 768 samples, about two third of which have negative diabetes diagnosis and one third with a positive diagnosis. The data set is randomly split into equal size of training and test sets of 384 samples each.

Table 1. Risk factor for GDM and Clinical Parameters for Encoding Dataset Schema Used

Attribute Name	*Clinical Associates*
Family Relatives	Type-2 DM in 1st-Degree Relatives
Number of Pregnancy	1 or more
Plasma Glucose Tolerance	History of abnormal glucose tolerance with $\mu = 120.9$ and $\delta = 32$
Diastolic Blood Pressure	$\mu = 69.1$ and $\delta = 19.4$
Body Mass Index (BMI)	$\mu = 32.0 kg/m^2$ and $\delta = 7.9 kg/m^2$
Triceps skin fold thickness	$\mu = 21$, $\delta = 17$ and BMI > 32kg
Diabetes Pedigree Function	Type-2 or GDM in previous pregnancy
Age	35years and above
2-hour serum insulin	$\mu = 79$ and $\delta = 115.2$
Ethnicity	African-American, Hispanic, Asian-American, Pacific Islander etc
Insulin Resistance	Polycystic ovary syndrome
Large Babies	History of babies >4.5kg at birth
Malformation	Birth of a malformed child
Perinatal Events	Unexplained perinatal loss
Diabetes Pedigree Function	$\mu = 0.5$ and $\delta = 0.3$
Maternal Birth/Large Babies	< 2.5kg or >4.5kg

2.2. Statement of Problem

The problem statements are as follows:

1. Its early detection is critical and imperative as unchecked scenarios will lead to increased maternal mortality. Non-robust tests and diagnosis are fast becoming redundant as it often yields inconclusive results due to unknown inputs.
2. Increased incorrect classification of conditions not even related to diabetes, but with symptoms that mimics a class type. And results in an increased rate of false-positive (*unclassified*) and true-negative (classify symptoms as a type when it is not), has become a concern in evolutionary modeling. Proposed model(s) seek to effectively group data into genuine class (GDM) via evolutionary models that use predictive data-mining rules and reinforcement learning (Section III).
3. Many datasets often consist of ambiguities, imprecision, noise and impartial truth that must be resolved via robust search. Also, speed constraint that often gets such solution trapped at local minima (resolved in Section III).
4. Hybrid models have been successfully used in diabetes study [3] with tradeoffs and conflicts that are not easily resolved. These include conflict imposed on model by the various underlying statistical dependencies that exist between the adopted heuristic methods in the hybrid, and conflict imposed on the hybrid by the dataset used. Proposed model resolves this (Section III) via its data encoding and profile creation that seeks to assign scores to rules that effectively classifies each dataset into a type or class of diabetes.
5. Parameter selection is a daunting task when searching a solution space for a complete and optimized solution that will aid effective and efficient classification in a certain domain. Careful selection is required so that the system does not result in model over-fitting of data as well as overtraining cum over-parameterization (resolved in Section III) as the model seeks to discover underlying probability of the data feat(s) of interest.

3. Intelligent Proposed Model

3.1. Linear Discriminant Analysis (LDA)

LDA is a simple and effective supervised classification method with wide range of applications. Its basic theory is to classify compounds (rules) dividing n-dimensional descriptor space into two regions separated by a hyper-plane that is defined by linear discriminant function. Discriminant analysis generally transforms classification tasks into functions that partitions data into classes; Thus, reducing the problem to an identification of a function. The focus of discriminant analysis is to determine this ***functional*** form (assumed to be linear) and estimate its coefficients. It was introduced in 1936 by Ronald Aylmer Fisher and his LDA function works by finding the mean of a set of attributes for each class, and using the mean of these means as boundary. The function achieves this by projecting attribute points onto the vector that maximally separates their class means and minimizes their within-class variance as expressed in Eq. 1 as follows:

$$LDA = X' S^{-1}(X_2 - X_1) - \frac{1}{2}(X_2 + X_1)' S^{-1}(X_2 - X_1) > c \quad (1)$$

where X is vector of the observed values, X_i (i = 1, 2…) is the mean of values for each group, S is sample covariance matrix of all variables, and c is cost function. If the misclassification cost of each group is considered equal, then c = 0. A member is classified into one group if the result of the equation is greater than c (or = 0), and into the other if it less than c (or = 0). A result that equals c (set to 0) indicates such a sample cannot be classified into ***either*** class, based on the features used by the analysis. LDA function distinguishes between two classes – if a data set has more than two classes, the process must be broken down into multiple two-class problems. The LDA function is found for each class versus all samples that were not of that class (one-versus-all). Final class membership for each sample is determined by LDA function that produced the highest value and is optimal when variables are normally distributed with equal covariance matrices. In this case, the LDA function is in same direction as Bayes optimal classifier [16], and it performs well on moderate sample sizes in comparison to more complex method [17]. Its mathematical function is simple and requires nothing more complicated than matrix arithmetic. The assumption of linearity in the class boundary, however, limits the scope of application for linear discriminant analysis. Real-world data frequently cannot be separated by linear boundary. When boundaries are nonlinear, the performance of the linear discriminant may be inferior to other classification methods. Thus, to curb this – we adopt a decimal encoding of the data to give us a semblance of linear, continuous boundaries.

3.2. Quadratic Discriminant Analysis (QDA)

QDA is another distance-based classifier by Smith [18], which is very similar to and more of an extension of LDA. Both discriminant functions assume that values of each attribute in each class are normally distributed, however,

the discriminant score between each sample and each class is calculated using the sample variance –covariance matrix of each class separately rather than the overall pooled matrix and so is a method that takes into account the different variance of each class. While, LDA assumes that the covariance matrices of the groups are equal; QDA makes no assumption. When the covariance matrices are not equal, the boundary between the classes will be a hyper-conic and in theory, the use of quadratic discriminant analysis will result in better discrimination and classification rates. However, due to the increased number of additional parameters to be estimated, it is possible that the classification by QDA is worse than that of linear discriminant analysis [19]. The QDA is found by evaluating the Eq. 2:

$$QDA = X'\left(S_1^{-1} - S_2^{-1}\right)X + 2\left(Y_2'S_1^{-1} - Y_1'S_2^{-1}\right)X - [Y_2'S_1^{-1}Y_2 - Y_1'S_2^{-1}Y_1 + Ln(\frac{|S_2|}{|S_1|})] > c \quad (2)$$

The same conditions apply to the nature of c as well as the classification, in the case that the result is equal to c or zero. As with LDA, the QDA distinguishes between two classes. For multiple class data sets, this was handled the same as for linear discriminant analysis. Size of differences in variances determines how much better QDA performs better than LDA. For large variance differences, QDA excels when compared to LDA. Additionally, of the two, only QDA can be used when population means are equal. QDA is more broadly applicable than the LDA; But, less resilient in non-optimal conditions. The quadratic discriminant can behave worse than the linear discriminant for small sample sizes. Additionally, data that is not normally distributed results in a poorer performance by the quadratic discriminant, when compared to the linear discriminant. Marks and Dunn [20] found the performance of the quadratic discriminant function to be more sensitive to the dimensions of the data than the linear discriminant, improving as the number of attributes increases to a certain optimal number, then rapidly declining. Linear and nonlinear discriminant functions are the most widely used classification methods. This broad acceptance is due to their ease of use and the wide availability of tools. Both, however, assume the form of the class boundary is known and fits a specific shape. This shape is assumed to be smooth and described by a known function. These assumptions may fail in many cases. In order to perform classification for a wider range of real-world data, a method must be able to describe boundaries of unknown, and possibly discontinuous, shapes.

3.3. K-Nearest Neighbourhood (KNN)

KNN is a well-known supervised learning model for pattern recognition. It was introduced by Fix and Hodges in 1951, and is still one of the most popular nonparametric models for classification problems [21]. KNN assumes that observations, which are close together, are likely to have the same classification. The probability that a point x belongs to a class is estimated by proportion of training points in a specified neighbourhood of x that belong to that class. This point(s) is then either classified by majority vote or by a similarity degree sum of the specified number (*k*) of nearest points. In majority voting, number of points in neighbourhood belonging to each class is counted, and the class to which the highest proportion belongs to is most likely classification of *x*. The similarity degree sum calculates a similarity score for each class based on the *K*-nearest points and classifies *x* into the class with the highest similarity score. Its lower sensitivity to outliers allows the majority voting to be commonly used other than the similarity degree sum [22]. We use majority voting for the data sets to determine which points belongs to neighbourhood so that distances from *x* to all points in the training set must be calculated. Any distance function that specifies which of two points is closer to the sample point could be used [21]). The most common distance metric used in *K*-nearest neighbour is the Euclidean distance [23]. The Euclidean distance between each test point f_t and training set point f_s, each with *n* attributes, is calculated via Eq. 3:

$$d = \left[\left(f_{t1} - f_{s1}\right)^2 + \left(f_{t2} - f_{s2}\right)^2 \ldots + \left(f_{tn} - f_{sn}\right)^2\right]^{\frac{1}{2}} \quad (3)$$

In general the following steps are performed for the *K*-nearest neighbour model [24] (a) chosen of *k* value, (b) distance calculation, (c) distance sort in ascending order, (d) finding *k* class values, and (e) finding the dominant class.

A challenge in using *K*nn is to determine optimal size of *k*, which acts as smoothing parameter. A small *k* is not sufficient to accurately estimate the population proportions around the test point. A larger *k* will result in less variance in probability estimates (but for risk of introducing more bias). *K* should be large enough to minimize probability of a non-Bayes decision, and small enough that all points included, gives an accurate estimate of the true class. Enas and Choi [25] found optimal value of k to depend on sample size and covariance structures in each population, and on proportions for each population in the total sample. In some cases where the differences in covariance matrices and difference between sample proportions are both small, or both large, then optimal *k* is $N^{3/8}$ (*N* is number of samples in training set). If there is a large difference between covariance matrices, and a small difference between sample proportions (or vice-versa), optimal *k* is determined by $N^{2/8}$.

This model presents several merits [26] in that: (a) its mathematical simplicity does not prevent it from achieving classification results as good as (or better than) other more complex pattern recognition techniques, (b) it is free of statistical assumptions, (c) its effectiveness does not depend on the space distribution of classes, and (d) when the boundaries between classes are not hyper-linear or hyper-conic, K-nearest neighbour performs better than LDA.

Some demerits of *knn* include that it does not work well if large differences are present in samples in each class. K-nearest neighbour provides poor data about the structure of its classes, and relative importance of variables in classification. Also, it does not allow graphical representation of the results, and in case of large number of samples, computation become excessively slow. In addition, Knn requires more memory and processing requirements than other methods. All prototypes in training set must be stored in memory and used to calculate Euclidean distance from every test

sample. The computational complexity grows exponentially as the number of prototypes increases [27].

3.4. Support Vector Machines (SVMs)

SVMs are a new pattern recognition model and tool founded on Vapnik's [28] statistical learning theory. SVMs were designed for binary classification and uses supervised learning to find an optimal separating hyper-plane between two groups of data. With such a plane, SVM predicts the classification of an unlabeled example by seeking which side of the separating plane the datasets lies. SVM acts as a linear classifier in a high dimensional feat space originated by a projection of original input space – resulting in a classifier that is in its general non-linear in the input space and it achieves good generalization performances by maximizing the margin between the two classes. Consider a set of training examples as follows:

$$\{\{x,\underline{y}_1\}\}\ x_i \in R^n, y_1 e(+1,-1); i=1,2,\ldots m$$

where the x_i are real n - dimensional pattern vectors and the y_i are dichotomous labels. SVM maps pattern vectors $x_i \in R^n$ into a possibly higher dimensional feature space $z = \emptyset(x)$) and construct an optimal hyper-plane $w.z + b = 0$ in feature space to separate examples from the two classes. For SVMs with L_1 soft-margin formulation, this is done by solving the primal optimization problem as follows:

$$Min\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\varepsilon_1 \ s.t\ y_i\left(w.z_i + b\right) \geq 1-\varepsilon_1 \qquad (4)$$
$$where\ \varepsilon_1 \geq 0,1,2,\ldots,m$$

C is regularization parameter used to decide a trade-off between training error and margin, and ε_1 (1,2,…m) are slack variables. Its dual form is also of the form:

$$\underset{\propto}{\text{Max}}\sum_{i=1}^{m}\propto_i -\frac{1}{2}\sum_{j=1}^{m}\propto_i \propto_j\ y_i y_j k\left(x_i x_j\right)$$

$$s.t.\sum_{i=1}^{m}\propto_i\ y_i = 0\ such\ that\ 0 \leq \propto_i \leq C,\ i=1,2,\ldots m$$

Where k(xi,xj) is the kernel function that implicitly defines the mapping resulting in the expression that:

$$f\left(x\right) = sgn\left\{\sum_{i=1}^{m}\propto_i\ y_i k\left(x_i, x\right) + b\right\}.$$

All kernel functions have to fulfil Mercer theorem. The most commonly used kernel functions are polynomial kernel and radial basis function kernel, respectively [29]. SVM differ from LDA and QDA in two ways. First, the feature space of a classification problem is not assumed to be linearly separable. Rather, a nonlinear mapping function (also called a kernel function) is used to represent the data in higher dimensions where the boundary between classes is assumed to be linear [30]. Second, the boundary is represented by support vector machines instead of a single boundary. Support vectors run through the sample patterns which are the most difficult to classify, thus the sample patterns that are closest to the actual boundary. Over-fitting is prevented by specifying a maximum margin that separates the hyper plane from the classes. Samples, which violate this margin, are penalized. The size of the penalty is a parameter often referred to as C [31,32].

3.5. Hybrid SVM-NN (Benchmark) Supervised Models

SVM uses associated learning to analyze data and recognize patterns in classification. It takes data input in n-dimensional space, and maps them into classes separated as hyper-planes. It trains the data and assigns them into the classes via a non-probabilistic binary linear classifier [6,7,33,34]. Each data represents a point in the space, and mapped easily into each separate class due to the wide gap between the two classes. To compute the margin, model constructs 2-parallel hyper-planes so that new data are predicted to belong to a class depending on the side of the gap they fall into. It efficiently performs non-linear classification via 'kernel trick' by implicitly mapping inputs into a high-dimensional feature space [3]. The larger the margin, the better its generalization error. Classes may overlap since each data is treated as a separate binary classification problem/task [35].

Knn assumes that data points which are close together, are likely to have the same classification. The probability that a point x belongs to a class is estimated by the proportion of training points in a specified neighbourhood of x that belong to that class. The point is either classified by a majority vote (where number of points in neighbourhood belonging to each class is counted, and the class to which the highest proportion of points belongs to is most likely the classification of *x* [22].

Okesola et al [36] Classification parameter for diabetes can be quite chaotic and non-uniform. Accuracy is improved if classification by model is based on local decision rules. Thus, it uses SVM to provide a global decision rule independent of sample that must be classified. Precise classification is limited to fact that diabetic symptoms are of various genres. Decision rules are localized and applied via collaborative filtering as opposed to the present application of global rules that sees and classifies diabetes based on pre-coded data on genre and type. Such an interchangeability of data is also likely to have local nature (Delany et al, 2004), and this is applicable to genuine diabetes symptoms. However, Okesola et al [36] model has the following errors (implicitly stated as):

a. How does model encode datasets used in the hybrid, and for such dynamic data that is rippled with ambiguities, noise and impartial truth
b. Parameter selection at training and testing was not clearly stated and number of runs that result in their convergence as we seeks to discover underlying probability of the data feat(s) of interest.
c. How did they resolve conflicts imposed by the underlying statistical dependencies in the adopted heuristics as well as that imposed in encoding of the dataset used?
d. What speed constraints were experienced?
e. Because supervised models yield inconclusive solutions of *unclassified* (false-positive) data and

wrongly *classified* (true-negative) data. What improvements are experienced by the model and at what rate (with its predictive data-mining rules and reinforcement learning)?

3.6. Bayesian Profile Hidden Markov Model (PHMM)

Ojugo et al [3] describes the Hidden Markov model as used in examination scheduling. Adapted to GDM diabetes classification problem, probability from one transition state to another is as in Figure 1. The PHMM is a double embedded chain that models complex stochastic processes. Markov process is a chain of state probabilities associated to each transition between states. In *n*-order Markov, its transition probabilities depend on ***current*** and ***n-1 previous*** states. A HMM process determines the state generated for each state observation in a series (output sequence).

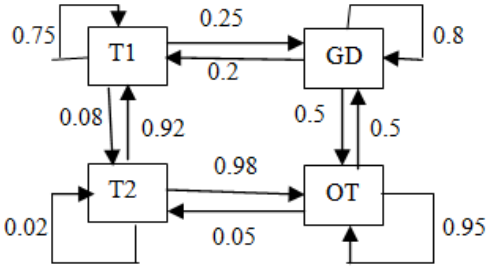

Figure 1. Actual State Transition with P(*x*)

For GDM diabetes analysis, a rule not accepted by the trained HMM, yields high probability of either a false-positive or true-negative result [3]. Traditional HMM scores data via clustering based on profile values. Probabilities of initial set of rules are sampled – then classified into GDM or non-GDM class. HMM maintains a log in memory to help reduce high true-negatives (rules of symptoms with semblance of diabetic feats) and high-false positives (unclassified rules for diabetes). Thus, our HMM is initially trained to assimilate normal behaviour of the various types or diabetes class/types. It then creates a profile of the rules, classifying them into type-1, type-2, gestational and other profile ranges were possible [3].

The Profile HMM as a variant of HMM, proffers solution to the fundamental problems of the HMM by: (a) makes explicit use of positional (alignment) data contained in observations or sequences, and (b) allows null transitions, where necessary so that the model can match sequences that includes insertion and deletions [33]. Used in GDM early detection, O is each rules contained therein to define the various symptoms of GDM diabetes type, T is time it takes each rule to classify data input, N is number of unclassified rules and those with symptom semblance that results in false-alarm rates, M is the number of rules accurately classified, π is the initial state or starting rule, A is state transition probability matrix, a_{ij} is the probability of a transition from a state *i* to another state *j*, B contains the N probability distributions for the codes in the knowledgebase from where profiles have been created (one rule for each state of the process); while HMM $\lambda = (A,B,\pi)$. Though, parameters for HMM details are incomplete as above; But, the general idea is still intact [3].

We can also align multiple codes (data) rules as sequence with significant relations. Its output sequence determines if an unknown code is related to sequence belonging to either of the diabetes (type class) or its variant (or those not) contained in the Bayesian net. We then use the profile HMM to score codes and make decision. Circles are ***delete*** state that detects rules as classified into GDM-diabetes types, rectangle are ***insert*** states that allows us to ***accurately*** classify rules of symptoms that have been previously unclassified inputs into a class type and consequently, update knowledgebase of the classified false-positives and true-negatives; diamonds are ***matched*** states that accurately classifies rules of symptoms into variants of similar symptom or unclassified rules, as in standard HMM [3,33]. Delete and insert are emission states in which an observation is made as PHMM passes through all the states. Emission probabilities, corresponding to B in standard HMM model is computed based on frequency of symbols that can be emitted at a particular state in the model; But, are positional-dependent (in contrast to standard model). Also, the emission probabilities are derived from Bayesian net, which represents our training phase. Finally, ***match*** states allow the model to pass through gaps, existing in the Bayesian net to reach other emission states. These gaps prevent model from over-fitting and overtraining as in Figure 2 [3]. Our forward algorithm computes (recursively) probabilities of all possible case by reusing scores calculated for partial sequences using Eq. 5 to Eq. 7 respectively as thus:

$$F_j^M = Log\frac{eM_j(x_i)}{qx_i} + \log(aM_{j-1}M_j \exp\left(F_{j-1}^M(i-1)\right) + aI_{j-1}M_j \exp\left(F_{j-1}^I(i-1)\right) + \bar{a}D_{j-1}M_j \exp\left(F_{j-1}^D(i-1)\right) \quad (5)$$

$$F_j^I = Log\frac{eI_j(x_i)}{qx_i} + \log(aM_jI_j \exp\left(F_j^M(i-1)\right) + aI_jI_j \exp\left(F_j^I(i-1)\right) + aD_jI_j \exp\left(F_j^D(i-1)\right) \quad (6)$$

$$F_j^D = \log(aM_{j-1}D_j \exp\left(F_{j-1}^M(i)\right) + aI_{j-1}D_j \exp\left(F_{j-1}^I(i)\right) + aD_{j-1}D_j \exp\left(F_{j-1}^D(i)\right). \quad (7)$$

3.7. Fuzzy Genetic Algorithm Trained Neural Network Model

The GANN is initialized with if-then rules. Individual fitness is computed as 30-individual are selected via the *tournament* method to determines new pool and individuals for mating. Crossover and mutation is applied to help ***net*** learn dynamic and non-linear feats in the dataset and feats of interest using a multi-point crossover. As new parents contribute to yield new individuals whose genetic makeup is combination of both parents, mutation is reapplied and are allocated new random values that still conforms to belief space. Number of mutation applied depends on how far CGA is progressed on the network (how fit is the fittest individual in the pool), which equals fitness of the fittest individual divided by 2. New individuals replace old with low fitness so as to create a new pool. Process continues until individual with a fitness value of 0 is found – indicating solution is reached [14].

4. Result Findings and Discussion

4.1. Model Performance

Ojugo et al [14] Performance is evaluated via as thus:

Table 2. Model Convergence Performance Evaluation

Model	*MSE*	*MRE*	*MAE*	*COE*
LDA	0.73	0.79	0.75	0.581
QDA	0.56	0.43	0.49	0.762
KNN	0.67	0.65	0.56	0.481
SVM	0.41	0.51	0.45	0.781
SVM-NN	0.32	0.39	0.32	0.791
PHMM	0.36	0.31	0.23	0.853
FGANN	0.36	0.37	0.46	0.818

4.2. Classification Accuracy

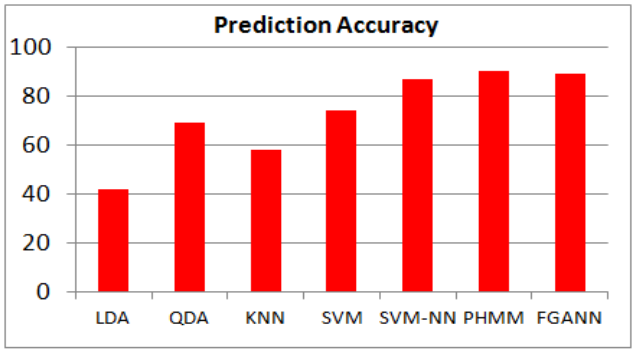

Figure 2. Prediction Accuracy of Algorithms in percentage

4.3. Processing Speed

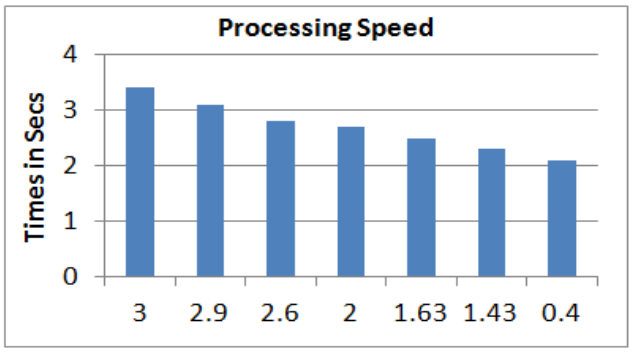

Figure 3. Processing time in Seconds

4.4. Convergence Time

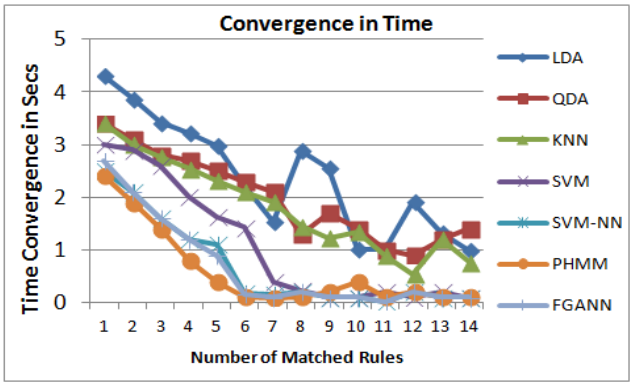

Figure 4. Convergence time of matches

4.5. Result Findings and Discussion

The ***rationale*** for model choice(s) adopted is to compare between: (a) supervised versus unsupervised model, (b) seek a measure to lay superiority claims to between supervised and unsupervised on task at hand, (c) compare clustering (profile) versus hill-climbing heuristic, and (d) measure convergence behavior and other statistic between the various heuristics and model. Also, LDA, QDA, SVM and SVM-NN converged after 405-iterations, 354-iterations, 287-iterations and 301-iterations respectively for supervised model. While, their convergence speed is as in Figure 4 above – it was also observed that PHMM converged after 253-iterations and FGANN converged after 213-iterations. FGANN significantly outperforms PHMM in some tasks; while PHMM was found to outperform FGANN in clustering task classification accuracy. We ***note***, model's speed is traded-off for greater accuracy of classification, more number of rule set generated to update the knowledge database for optimality and greater functionality.

To ensure model's effectiveness and accuracy, we compute the misclassification rate for each model and its corresponding improvement percentages for both training and testing dataset as summarized in Table 2 and Table 3 respectively. Equations for the misclassification rate and its improvement percentage of the unsupervised (B) model against those of the supervised (A) model, is respectively calculated as follows:

$$Misclassification\ Rate(MR) = \frac{No.\ of\ \ Incorrect\ \ Diagnosis}{No.\ of\ \ Sample\ set} \quad (8)$$

Table 3. Misclassification Rate of Each model

Model	*Classification Errors*	
	Training Data	Testing Data
LDA	36.6%	34.9%
QDA	29.9%	27.3%
KNN	43.4%	39.7%
SVM	21.01%	18.9%
SVM-NN	12.39%	17.21%
PHMM	18.7%	15.8%
FGANN	19.3%	18.3%

$$Improvement\ Percentage = \frac{MR(A) - MR(B)}{MR(A)} x100. \quad (9)$$

Table 4. Improvement Percentage

Model	*Improvement %*	
	Training Data	Testing Data
LDA	45.83%	41.16%
QDA	45.01%	44.05%
KNN	41.79%	43.09%
SVM	52.1%	54.67%
SVM-NN	67.02%	68.89%
PHMM	78.78%	76.33%
FGANN	69.30%	69.91%

Table 3 and Table 4 shows ***unsupervised*** model with lowest error in comparison to supervised models. PHMM and FGANN had misclassification of 18.7% and 19.3%

respectively – with improvement of 78.78% and 69.30% respectively. Conversely, LDA/KNN has misclassification error rate of 36.6% and 43.4% respectively; And shows an improvement rate of 45.83% and 41.79% respectively. It was also observed that though *K*nn is quite sensitive to the relative magnitude of different attributes, all attributes are scaled by their z-scores before using *K*-nearest neighbour model in tandem with Antal et al [37].

5. Conclusion and Recommendations

As for GDM, its risk factors are many and must be assessed regularly in all pregnant women. Placental mass and hormonal changes during pregnancy may contribute to the pathogenesis of GDM. Insidious onset of most cases of GDM necessitates a diligent search and screening. Thus, we advise RBG, FBG, and OGTT to be used in GDM diagnosis (as agreed by [38]) and the parameters can then be adapted to unsupervised model. A significant number of GDM cases in pregnancy require insulin treatment. Evidence now abounds that sulphonylureas and metformin are safe in pregnancy. The management and follow-up of GDM is for life. Also, study used ***supervised*** and ***unsupervised*** classification, which consists of 5-phases: (a) train models with available data, (b) determine minimal fuzziness via the obtained weights and same criterion, (c) delete outliers in data, (d) compute membership probability of output, and (e) assign output to appropriate class by largest probability.

Unsupervised models do not assume the shape of partition unlike LDA and KDA. In contrast to *KNN*, PHMM and FGANN do not require storage of training data. Once model is trained, it performs much faster than KNN, because it does not need to iterate through individual training samples. Also, both the PHMM and FGANN does not require experimentation, final selection of kernel function and a penalty parameter as with SVM; But rather, it solely relies on a training process in order to identify final classifier model. Lastly, unsupervised models does not need large amount of data in order to yield accurate results.

References

[1] Ojugo, A., Eboka, A., Okonta, E., Yoro, R and Aghware, F., (2012). *GA rule-based intrusion detection system*, Journal of Computing and Information Systems, 3(8), pp 1182-1194.

[2] Goldenberg, R and Punthakee,, Z (2013). *Definition, classification and diagnosis, prediabetes and metabolic syndrome*, 37(1), S8-S11.

[3] Ojugo, A.A., Allenotor, D and Okobah, I.P., (2016). *Intelligent Classification models for Gestational Diabetes: a comparative approach*, FUPRE-Lecture Series and Technical Report, Vol. 11, pp 34-45.

[4] Ajufo, B.C., Utomi, L.N., Mekanya, C.O and Ojugo, A.A., (2014). Discovering the rising rate of Diabetes in the Niger Delta region of Nigeria: On the way forward, Technical report in Medicine, Federal Medical Centre, FMCTR-1320, Vol. 46.

[5] Khashei, M., Eftekhari, S and Parvizian, J. (2012). *Diagnosing diabetes type-II using a soft intelligent binary classifier model*, Review of Bioinformatics and Biometrics, 1(1), pp9-23.

[6] Ojugo, A.A., Allenotor, D., Oyemade, D.A., Longe, O.B and Anujeonye, C.N., (2015b): *Comparative stochastic study for credit-card fraud detection models*, African Journal of Computing & ICTs. Vol 8, No. 1, Issue 1. Pp 15-24.

[7] Ojugo, A.A, A.O. Eboka., R.E. Yoro., M.O. Yerokun., F.N. Efozia., (2015a). *Hybrid model for early diabetes diagnosis*, Mathematics and Computers in Science and Industry, Series 50, Pp 176-182.

[8] American Diabetes Association. *Standards of Medical Care in Diabetes – 2009*. Diabetes Care, 32: S13-61.

[9] Canadian Diabetes Association. *Standards of Medical Care in Diabetes – 2014*. Journal of Diabetes Care, 32: S13-61.

[10] Perez, M and Marwala, T., (2011). *Stochastic optimization approaches for solving Sudoku*, IEEE Transaction on Evol. Comp., pp.256-279.

[11] Chinenye, S and Young, E., (2011). State of diabetes care in Nigeria: a review, The Nigerian Health Journal, 11(4), pp101-106.

[12] Adebisi, S.A., Oparinde, D.P., Olarinoye, J.K., Aboyeji, P.A and Ogunro, P.S., (2012).

[13] Arije A, Kuti M, Fasanmade A, Akinlade K, Ashaye A, Obajimi M et al. Control of hypertension in Nigerians with diabetes mellitus; A report of the Ibadan diabetic/kidney disease study group. Int J Diabetes Metabolism. 2007; 15: 82-86.

[14] Ojugo, A.A., Emudianughe, J., Yoro, R.E., Okonta, E.O and Eboka, A., (2013). *Hybrid neural network gravitational search algorithm for rainfall runoff modeling*, Progress in Intelligence Computing and Application, 2(1), doi: 10.4156/pica.vol2.issue1.2, pp22-33.

[15] Heppner, H and Grenander, U., (1990). S*tochastic non-linear model for coordinated bird flocks*, In Krasner, S (Ed.), *The ubiquity of chaos* pp.233-238. Washington: AAAS.

[16] Billings, S. and Lee, K., "Nonlinear Fisher discriminant analysis using a minimum squared error cost function and the orthogonal least squares algorithm", Neural Networks, Vol. 15, pp. 262-270, 2002.

[17] Ghiassi, M. and Burnley, C., "Measuring effectiveness of a dynamic artificial neural network algorithm for classification problems", Expert Systems with Applications, Vol. 37, pp. 3118-3128, 2010.

[18] Smith, C.A., (1947). *Some examples of discrimination*, Annals of Eugenics, Vol. 13, pp. 272-282.

[19] Malhotra, M., Sharma, S. and Nair, S., Decision making using multiple models", European Journal of Operational Research, Vol. 114, pp. 1-14, 1999.

[20] Marks, S. and Dunn, O., "Discriminant functions when covariance matrices are unequal", Journal of the American Statistical Association, Vol. 69, pp. 555-559, 1974.

[21] Fix, E. and Hodges, J., "Discriminatory analysis -Nonparametric discrimination: Consistency properties", Project No. 2-49-004, Report No. 4, Contract No. AF 41(128)-31, USAF School of Aviation, Randolph Field, Texas, 1951.

[22] Chaovalitwongse, W., "On the time series k-nearest neighbor classification of abnormal brain activity", IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, Vol. 37, 2007.

[23] Viaene, S., Derrig, R., Baesens, B., and Dadene, G., (2002). *A comparison of state-of-the art classification techniques for expert automobile insurance claim fraud detection*, The Journal of Risk and Insurance, Vol. 69, pp. 373-421.

[24] Yildiz, T., Yildirim, S., Altilar, D., (2008). *Spam filtering with parallelized KNN algorithm*, Akademik Bilisim

[25] Enas, G. and Choi, S., (1986). *Choice of the smoothing parameter and efficiency of k-nearest neighbor*, Computers and Mathematics with Applications, Vol. 12, pp. 235-244.

[26] Berrueta, L., Alonso-Salces, R., Heberger, K., "Supervised pattern recognition in food analysis", Journal of Chromatography A, 1158, pp. 196-214, 2007.

[27] Muezzinoglu, M. and Zurada, J., "RBF-based neurodynamic nearest neighbor classification in real pattern space", Pattern Recognition, Vol. 39, pp. 747-760, 2006.

[28] Vapnik, V., "Statistical learning theory", Wiley, New York, 1998.

[29] Song, J. and Tang, H., "Support vector machines for classification of homo-oligomeric proteins by incorporating subsequence distributions", Journal of Molecular Structure: THEOCHEM 722, pp. 97-101, 2005.

[30] Duda, R., Hart, P., and Stork, D., "Pattern classification", New York: John Wiley & Sons, Inc.2001.

[31] Brown, M., Grundy, W., Lin, D., Cristianini, N., Sugnet, C., Furey, T., Ares, M., and Haussler, D., "Knowledge - based analysis of microarray gene expression data by using support vector machines"

Proceedings of the National Academy of Sciences of the United States of America, Vol. 97, pp. 262-267, 2000.

[32] Christianini, N.and Shawe-Taylor, J., "An introduction to support vector machines", Cambridge University Press, 2000.

[33] Ojugo, A.A., Ben-Iwhiwhu, E., Kekeje, D.O., Yerokun, M.O and Iyawa, I.J.B., (2014). *Malware propagation on time varying network*, Int. J. Modern Edu. Comp. Sci., 8, pp25-33.

[34] Cukier W., Cody, S and Nesselroth, E. (2006). Genres of spam: Expectations and deceptions, Proc. of the 39th Annual Hawaii International Conference on System Sciences, Vol. 3. www.computer.org/csdl/proceedings/hicss/2006/2507/03/250730051a.pdf.

[35] Longe, O.B., Robert, A.B.C., Chiemeke, S.C and Ojo. F.O., (2008). *Feature Outliers And Their Effects On The Efficiencies Of Text Classifiers In The Domain Of Electronic Mail*, The Journal of Computer Science and Its Applications, 15(2), pp. 45-56.

[36] Okesola, J.O., Ojo., F.O., Adigun, A.A and Longe, O.B., (2015). *Adaptive high probability algorithms for filtering advance fee fraud emails using the concept of data locality*, Computing, Information Systems, Development Informatics and Allied Research Journal, Vol. 6, No. 1, pp 7-12.

[37] Antal, P., Fannes, G., Timmerman, D., Moreau, Y., and Moor, B.D., "Bayesian applications of belief networks and multilayer perceptrons for ovarian tumor classification with rejection", Artificial Intelligence in Medicine Vol. 29, pp. 39-60, 2003.

[38] Oputa, R.N and Nzeribe, E.A., (2013). *Gestational diabetes mellitus: a clinical challenge in Africa*, African Journal of Diabetes Medicine, Vol. 21, No. 2, pp 29-31.

[39] Barakat, N.H., Bradley, A.P., Barakat, M.N.H., (2010). *Intelligible support vector machines for diagnosis of diabetes mellitus*, IEEE Transactions on Information Technology in Biomedicine, 14(4), pp1114-1120.

[40] Berardi, V. and Zhang, G. P., "The effect of misclassification costs on neural network classifiers", Decision Sciences, Vol. 30, pp. 659-68, 1999.

[41] Berks, G., Keyserlingk, D., Jantzeen J., Dotoli, M and Axer H. (2000). *Fuzzy clustering: versatile means to explore medical database*, ESIT, Aachen, Germany.

[42] Calisir D. and Dogantekin, E., "An automatic diabetes diagnosis system based on LDA-Wavelet Support Vector Machine Classifier", Expert Systems with Applications, Vol. 38, pp. 8311-8315, 2011.

[43] Caudill M., (1987). *Neural Networks Primer*, AI Expert, pp 46-52.

[44] Chakraborty, R., (2010). Soft computing and fuzzy logic, Lecture notes 1-37, [web]: www.myreaders.info/07_fuzzy_systems.pdf.

[45] Charya, S., Odedra, D., Samanta, S., and Vidyarthi, S., "Computational Intelligence in Early Diabetes Diagnosis: A Review", The Review of Diabetes Studies, Vol. 7, pp.252-262, 2010.

[46] Coello, C., Pulido, G and Lechuga, M., (2004). *Handling multiple objectives with particle swarm optimization*, Proc. of Evolutionary Computing, 8, pp 256-279.

[47] Dawson, C and Wilby, R., (2001). *Comparison of neural networks in river flow forecasting*, J. of Hydrology and Earth Science, SRef-ID: 1607-7938/hess/2001-3-529.

[48] Edo, A.E., Edo, G.O., Ohehen, O.A., Ekhator, N.P and Ordiah, W.C., (2015). *Age and diagnosis of type-2 diabetes in Benin City Nigeria*, African Journal of Diabetes Medicine, Vol. 23, No. 1, Pp18-19.

[49] Fausett L., (1994). *Fundamentals of Neural Networks*, New Jersey: Prentice Hall, pp.240.

[50] Fisher, R. A., "The use of multiple measurements in taxonomic problems", Annals of Eugenics, Vol. 7, pp. 465-475, 1936.

[51] Ganji M. F. and Abadeh, M. S., "A fuzzy classification system based on Ant Colony Optimization for diabetes disease diagnosis", Expert Systems with Applications, Vol. 38, pp. 14650-14659, 2011.

[52] Giles, D and Draeseke, R., (2001). *Economic Model Based on Pattern recognition via fuzzy C-Means clustering algorithm*, Economics Dept Working Paper, EWP0101, University of Victoria, Canada.

[53] Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., and Lander, E.S., "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring", Science, Vol. 286, pp. 531-537, 1999.

[54] Guo, W.W and Xue, H., (2011) *An incorporative statistic and neural approach for crop yield modelling and forecasting*, Neural Computing and Applications, 21(1), pp109-117.

[55] Inan, G and Elif, D., (2005). *Adaptive neuro-fuzzy inference system for classification of EEG using wavelet coefficient*, retrieved from http://www.rorylewis.com/PDF/04.

[56] Jang, J.S., (1993). *Adaptive fuzzy inference systems*, IEEE Transactions on Systems, Man and Cybernetics, vol.23, pp. 665-685.

[57] Kuan, C and White, H., (1994). *Artificial neural network: econometric perspective*", Econometric Reviews, Vol.13, Pp. 1-91 and Pp.139-143.

[58] Ludmila I. K. (2008), *Fuzzy classifiers* Scholarpedia, 3(1), pp2925.

[59] Mandic, D and Chambers, J., (2001). *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*, Wiley & Sons: New York, pp56-90.

[60] Menezes, A.C., Pinheiro, P.R., Pinheiro, M.C.D., Pequeno, T., (2012). Towards the applied hybrid model in decision making: support the early diagnosis of type-2 diabetes, [online] http://www.researchgate.net/publication/262166315.

[61] Minns, A., (1998). *Artificial neural networks as sub-symbolic process descriptors*, published PhD Thesis, Balkema, Rotterdam, Netherlands.

[62] Nascimento L., (1991). Comparative Genomus of two Leptospira interrogans Serovirs, http://jb.asm.org/cgi/content/full/186/7/2164.

[63] Ojugo, A.A., Allenotor, D., Oyemade, D.A., Yoro, R.E and Anujeonye, C.N., (2015c): Immunization Model for Ebola Virus in Rural Sierra-Leone, African Journal of Computing & ICTs. Vol 8, No. 1, Issue 1. Pp 1-10.

[64] Peter, S., (2014). *An analytical study on early diabetes and classification of diabetes mellitus*, Bonfring International Journal of Data Mining, Vol. 4, No. 2, Pp 7-11.

[65] Reynolds, R., (1994). *Introduction to cultural algorithms*, Transaction on Evolutionary Programming (IEEE), pp.131-139.

[66] Ursem, R., Krink, T., Jensen, M.and Michalewicz, Z., (2002). *Analysis and modeling of controls in dynamic systems*. IEEE Transaction on Memetic Systems and Evolutionary Computing, 6(4), pp.378-389.

[67] Vaarala, O., Knip, M., Paronen, J., Hamalainen, A.M., Muona, P., Vaatainen, M., Ilonen, J., Simell, O and Akerblom, H.K., (1999). *Cow's milk formula feeding induces primary immunization to insulin in infants at genetic risk for type-1 diabetes*, Diabetes, 8(7), pp1389-1394.

[68] Antal, P., Fannes, G., Timmerman, D., Moreau, Y., and Moor, B.D., "Bayesian applications of belief networks and multilayer perceptrons for ovarian tumor classification with rejection", Artificial Intelligence in Medicine Vol. 29, pp. 39-60, 2003.

[69] Temurtas, H., Yumusak, N and Temurtas, F., (2009). *A comparative study on diabetes disease diagnosis using neural networks*, Expert Systems with Applications, Vol. 36, pp. 8610-8615.