International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Comparative Study of Diabetic Patient Data Using Classification Algorithm in Weaka

Ahlam A. Sharief[1], Alaa F. Sheta[2], Talaat M. Wahbi[3]

[1]Computer Science, Sudan University of Science& Technology Khartoum, Sudan

[2]Computer Science Department, Taif University, Taif, Saudiarabia,

[3]Computer Science, Sudan University of Science& Technology Khartoum, Sudan

Abstract: ***Diabetes mellitus is now a big growing health problem as it is fourth biggest cause of death world wide particularly in the industrial and developing countries.Early detection of diabetes is of vital importance nowadays.There have been many techniques used in Machine learning that were applied over diabetes diagnosis to help physicians. Diabetes mellitus or simply diabetes is a disease caused due to the increase level of blood glucose. Various available traditional methods for diagnosing diabetes are based on physical and chemical tests. These methods can have errors due to different uncertainties. A number of Data mining algorithms were designed to overcome these uncertainties.***

Keywords: diabetes, Classification, Feature Selection, Pima Indians Diabetes Dataset (PIDD), Weaka

1. Introduction

Studies in the field of medical decision support systems have been secured and, because of the high achievement rate of these studies, intense interest for this field is expanding consistently. These frameworks often utilize different artificial intelligence techniques and data mining [1]. In medical decision support systems, there is a more intense interest in the study of diseases that are basic all through the world, anddiabetes is one of them. Past studies in this field are specified below: Authors in [2] expressed that diabetes mellitus is adisorder with cluttered digestion system and unseemlyhyperglycemia because of either a lack of insulin emission or the blend of insulin resistance and deficient insulin discharge to repay. Diabetes is a metabolic infection where a person suffer from increased sugar level, this is because of either the pancreas does not produce sufficient insulin for the body or the cell does not react to the insulin that is produced [3]. Generally there are two types of diabetes, type1diabetes, where the pancreas does not create sufficient insulin and this is essentially because of the annihilation of insulin creatingbeta cells of pancreas via auto immunity. Type2 diabetes is the placethe cellsneglect to take the insulin because ofdifferent variables [4].

As per World Health Organization(WHO) 2016 report

- 422 million individuals worldwide have Diabetes Mellitus.
- More than 80% of Diabetes Mellitus passings happen in low and center wage nations.
- The WHO estimates that diabetes resulted in 1.5 million deaths in 2012, making it the 8th leading cause of death.

However another 2.2 million deaths worldwide were attributable to high blood glucose and the increased risks of associated complications (e.g. heart disease, stroke, kidney failure), which often result in premature death and are often listed as the underlying cause on death certificates rather than diabetes. Healthy regimen, systematic physical activity, keeping up a typical body weight and evading utilization of tobacco can prevent or delay the onset of type2 Diabetes Mellitus[5].

Diabetes can be controlled by utilizing diverse measures like insulin and eating routine. For this it ought to be distinguished as right on time as would be prudent and accordingly give fitting treatment. The majority of the grouping, recognizing and diagnosing medicines is in view of synthetic and physical tests. In light of the derivation acquired from these outcomes, a specific illness can be anticipated. Forecast may have blunders. This is because of diverse vulnerability of different parameters utilized for testing [6]. Such vulnerabilities make the expectations wrong and keep the shots of curing the ailment. The registering office has been advanced with incredible headways. These headways gave by data innovation, serves to arrange the information, foresee the results and conclusion of numerous illnesses all the more precisely. The fundamental focal point of data innovation is that colossal information stockpiling of pasttolerant's records are kept up and observed by healing facilities consistently for different references [7]. This restorative information helps the specialists to inspect distinctive examples inthe information set. The examples found in information sets may be utilized forarrangement, forecast and judgment of the illnesses [6].

1.1 Related work Diabetes

The technique for diagnosing diabetes using back spread neural framework figuring is proposed in this study [8]. The beat, glucose obsession in blood, serium insulin, BMI, number of times a man pregnant and age are the crucial parametersforthediagnosingdiabetes.Theessentialdrawback of this structure is missing estimations of the information. Authors in [9] improved this framework to overcome the disservice of this structure. The information dataset is changed using the missing qualities and it upgrades the structure by overhauling the portrayal precision. Data preprocessing framework similarly proposed in this study which is used to improve the pace of the structure. Author in [10] proposed a procedure utilizing a back-propagation

neural system alongside parallel characterization. The inputs to this system are glucose test, post plasma, age. The rate of execution of this structure is 92.5% when it is contrasted and past system. The expectation of glucose level for future in light of the present level is utilizing fake neural system. Artificial Neural Network is prepared utilizing the parameters to anticipate the glucose level. The Pima Indian diabetes dataset is utilized for exploration work. The structures with 22 distinct classifiers are utilized for the exact arrangement utilizing info as Pima Indian diabetes dataset. The execution assessment of structure taking intoaccountkmeansandknniscontrastedandexistingsystem and exactness scope of structure is 66.6 to 77.7% in [11]. Themodel utilizing cross breed k-means and choice tree with the assistance of bolster vector machine, GDA, coming to the arrangement exactness of 82.05%. The creator accomplished distinctive grouping exactness for system with different mixes of ANN, DT ANN and fell GA CFS ANN in [12], [13] and [14]. The paper is structured as follows. Section 2 the basic concepts and modelling approaches for classification are briefly reviewed. Section 3 gives review about features selection. Section 4 contain the performance criterion. Section 5is divided into tow subsections, the first one gives an insight about the area of study and the data collected, the second describe the resultof summachine learning algorithms. The last section is conclusions and feature work.

2. Classification Approaches

In this section, the basic concepts and modeling approaches for classification are briefly reviewed.

2.1. Multi-Layer Perceptrons (MLPs)

Artificial neural networks (ANNs) are computer systems developed to mimic the operations of the human brain by mathematically modelling its neurophysiological structure. Artificial neural networks have been shown to be effective at approximating complex nonlinear functions [15].

Forclassificationtasks,thesefunctionsrepresenttheshapeofthepartitionbetween classes. In artificial neural networks, computational units called neurons replace the nerve cells and the strengths of the interconnections are represented by weights, in which the learned information is stored. This unique arrangement can acquire some of the neurological processing ability of the biological brain such as learning and drawing conclusions from experience. Artificial neural networks combine the flexibility of the boundary shape found in K-nearest neighbor with the efficiency and low storage requirements of discriminant functions. Like the K-nearest neighbor, artificial neural networks are data driven; there are no assumed model characteristics or distributions, as is the case with discriminant analysis. Multi-layer perceptrons (MLPs) are one of the most important and widely used forms of artificial neural networks for modelling, forecasting, and classification [16]. These models are characterized by the network of three layers of simple processing units connected by acyclic links. Data enters the network through the inputlayer, moves through hidden layer, and exits through the output layer. Each hidden layer and output layer node collects data from the nodes above it (either the input layer or hidden layer) and applies an activation function. Activation functions can take several forms. The type of activation function is indicated by the situation of the neuron within the network. In the majority of cases input layer neurons do not have an activation function, astheir role is to transfer the inputs to the hidden layer. In practice, simple network structure that has asmall number of hidden nodes often works well in out-of-sample forecasting. This may be due to the over-fitting effect typically found in the neural network modelling process. An over-fitted model has a good fit to the sample used for model building but has poor generalizability to data out of the sample. There exist many different approaches such as the pruning algorithm, the polynomial time algorithm, the canonical decomposition technique, and the network information criterion for finding the optimal architecture of an artificial neural network. These approaches can be generally categorized as follows [17]:

1) Empirical or statistical methods that are used to study the effect of internal parameters and choose appropriate values for them based on the performance of model. The most systematic and general of these methods utilizes the principles from Taguchis design of experiments.
2) Hybrid methods such as fuzzy inference where the artificialneuralnetworkcanbeinterpretedasanadaptive fuzzy system or it can operate on fuzzy instead of real numbers.
3) Constructive and/or pruning algorithms that, respectively, add and/or remove neurons from an initial architecture using a previously specified criterion to indicate how artificial neural network performance is affected by the changes.

Evolutionary strategies that search over topology space by varying the number of hidden layers and hidden neurons through application of genetic operators andevaluation of the different architectures according to an objective function [18].

Although many different approaches exist in order to find the optimal architecture of an artificial neural network,these methods are usually quite complex in nature and are difficult to implement. Furthermore, none of these methods can guarantee the optimal solution for all real forecasting problems. To date, there isno simple clearcut method for determination of these parameters and the usual procedure is to test numerous networks with varying numbers ofhidden units, estimate generalization error for each and select thenetwork with the lowestgeneralization error. Once a network structure is specified, the network is ready for training a process of parameter estimation. The parameters are estimated such that the cost function of neural network is minimized.

2.2. K-Nearest Neighbor (KNN)

The K-nearest neighbor (KNN) model is a well known supervised learning algorithm for pattern recognition that first introduced by Fix and Hodges in 1951, and is still one of the most popularnon parametric models for classification problems [19].K-nearest neighbor assumes that observations, which are close together, are likely to have the

same classification. The probability that a point x belongs to a class can be estimated by the proportion of training points in a specified neighborhood of x that belong to that class. The point may either be classified by majority vote or by a similarity degree sum of the specified number (k) of nearest points. In majority voting, the number of points in the neighborhood belonging to each class is counted, and the class to which the highest proportion of points belongs is the 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. Due to its lower sensitivity to outliers, majority voting is more commonly used than the similarity degree sum [20]. In order to determine which points belong in the neighborhood, the 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 employed [19]. In general thefollowing steps are performed for the K-nearest neighbor model:

i) Chosen of k value.
ii) Distance calculation.
iii) Distance sort in ascending order.
iv) Finding k class values.
v) Finding dominant class.

One challenge to use the K-nearest neighbour is to determine the optimal size of k,which acts as a smoothing parameter. A small k will not be sufficient to accurately estimate the populationproportions around the testpoint. A larger k will result in less variance in probability estimates but the risk of introducing more bias. K should be large enough to minimize the probability of a non-Bayes decision, but small enough that the points included give an accurate estimate of the trueclass. Authors in [21] found that the optimal value of k depends upon the sample size and covariance structures in each population, as well as the proportions for each population in the total sample. This model presents several advantages:

- Its mathematical simplicity, which does not prevent it from achieving classification results as good as (or even better than) other more complex pattern recognition techniques.
- It is free from statistical assumptions, such as the normal distribution of the variables.
- Its effectiveness does not depend on the space distribution of the classes.

In additional, when the boundaries between classes cannot be described as hyper-linear or hyper-conic, K-nearest neighbour performs better than the (LDA) and (QDA) functions. Authors in [21]found that the linear discriminant performs slightly better than K-nearest neighbour when population covariance matrices are equal, a condition that suggests a linear boundary. As the differences in the covariance matrices increases, K-nearest neighbor performs increasingly better than the linear discriminant function. However, despite of the all advantages cited for the K-nearestneighbour models, they also have some disadvantages. K-nearest neighbour model cannot work well if large differences are present in the number of samples in each class. K-nearest neighbor provides poor information about the structure of the classes and of the relative importance of each variable in the classification. Furthermore, it does not allow a graphical representation of the results, and in the case of large number of samples, the computation can become excessively slow. In addition, Knearestneighbour model much higher memory and processing requirementsthan other methods. All prototypes in the training set must be stored in memory and used to calculate the Euclidean distance from every test sample. The computationalcomplexity grows exponentially as the number of prototypes increases.

2.3 Support Vector Machines (SVMs)

Support vector machines (SVMs) are a new pattern recognition tool theoretically founded on Vapniks statistical learning theory [22]. Support vector machines, originally designed for binary classification, employs supervised learning to find the optimal separating hyper-plane between the two groups of data. Having found such a plane, support vector machines can then predict the classification of an unlabeled example by asking on which side of the separating plane the example lays. Support vector machine acts as a linear classifier in a high dimensional feature space originated by a projection of the original input space, the resulting classifier is in general nonlinear in the input space and it achieves good generalization performances by maximizing the margin between the two classes. Support vector machines differ from discriminant analysisin two significant 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) isused to represent the data in higher dimensions where the boundary between classes is assumed to be linear [23]. 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 [24].

2.4. Decision tree

Decision tree is a supervised approach to classify a large number of datasets that make up the structure of the rules are simple, clear and easy to understand [25]. The decision tree is used to examine the data and form a rules in a tree that will be used for forecasting needs. There are many types of algorithms in the decision tree that can be utilized for a variety ofneeds [26], some of them and used in this study are J48 which is the developmentof C4.5, Naive Bayess Tree (NBTree) and Reduced Error Pruning Tree (REPTree).

2.5. Feed Forward Back propagation Neural Network

Neural systems are prescient model that have capacity to learn, examine, sort out the information and foresee test outcomesinlikemanner.Amongafewsortsofneuralsystems, Feed forward neural system is generally utilized in therapeutic conclusion applications and others. These systems are prepared by a situated of examples called preparing set, whose result is as of now known. In our study, Multilayer Perceptron Feed forward back spread Neural Network prepared with Levenberg Marquardt (LM) calculation is requisitioned arrangement. LM preparing

calculation does not get stuck in nearby minima and produces a superior expense capacity [27]. Feed Forward NN comprises of info, shrouded and a yield layer, and the information works in forward course, and the lapse is back spread to upgrade the weights at each age to lessen blunders [28]. Extensive studies by many researchers have demonstrated higher performance and accuracy in predicting clinical outcomes of diabetes diagnosis using neural network strategies (Table 3).

Table I: Artificial Intelligence Approaches For Early Diabetes Detection

Author	Algorithm	Accuracy
Kayaer, ildirim	MLP	77.08
Kayaer, ildirim	RBF	68.23
Kordos et al	k-nearest-neighbor	77
Barakat et al	SVM	94
Ster, Dobnikar	k-NN	71.9
Ster, Dobnikar	CART	72.8
Ster, Dobnikar	MLP	75.2
Ster, Dobnikar	LVQ	75.8
Ster, Dobnikar	LDA	77.5
Polat et al	GDA and LS-SVM	78.21
Polat, Gunes	PCA-ANFIS	89.47
Dogantekin et al	LDA-ANFIS	84.61
Ubeyli	MLPNN	91.53
Ubeyli	MME	99.17

3. Features Selection

The greater part of true grouping issues require directed realizing where the basic class probabilities and class-contingent probabilities are obscure, and each occasion is connectedwith a class name [8].In genuinecircumstances, we regularly have little informationabout relevant features. In this manner, to better represent the domain, numerous candidate features are presented, bringing about the presence of unimportant/excess features to the objective idea. A candidate features is neither immaterial nor excess to the objective idea; an insignificant candidate is not specifically take up with the objective idea but rather influence the learning process, and a repetitive features does not add anything new to the objective idea [8]. In numerous order issues, it is hard to learn great classifiers some time recently expelling these undesirable features because of the tremendous size of the information. Decreasing the number of superfluous/repetitive elements can radically diminish the running time of the learning calculations and yields a more broad classifier. This aides in showing signs of improvement understanding into the fundamental idea of a true characterization issue. A general feature selection for classification framework is shown in Figure1. Highlight choice chiefly influences the preparation period of order [29]. In the wake of producing components,rather than handling information with the entire elements to the learning calculation straightforwardly, include determination for characterization will firstperform highlight choice to choose a subset of elements and afterward prepare the information with the chose components to the learning calculation. The element choice stage may be autonomous of the learning calculation, similar to channel models, or it may iteratively use the execution of the learning calculations to assess the nature of the chose highlights, similar to wrapper models. With the at last chose highlights, a classifier is instigated for the forecast stage.Typically highlight determination for arrangement endeavors to choose the insignificantly estimated subset of components as indicated by the accompanying criteria. Feature selection is the process ofselecting a subset of the terms occurring in the training setand using only this subset as features in text classification. Feature selection serves two main purposes. First, it makes training and applying a classifiermore efficient by decreasing the size of the effective vocabulary.

This is of particular importance for classifiers that, unlike NB, are expensive to train. Second, feature selection often increases classification accuracy by eliminating noise features. A noise feature is one that, when added to the document representation, increases the classification error on new data. Suppose a rare term, say arachnocentric, has no information about a class, say China, but all instances of arachnocentric happen to occur in China documents in our training set. Then the learning method might produce a classifier that misassigns test documents containing arachnocentric to China. Such an incorrect generalization from an accidental property of the training set is called overfitting.

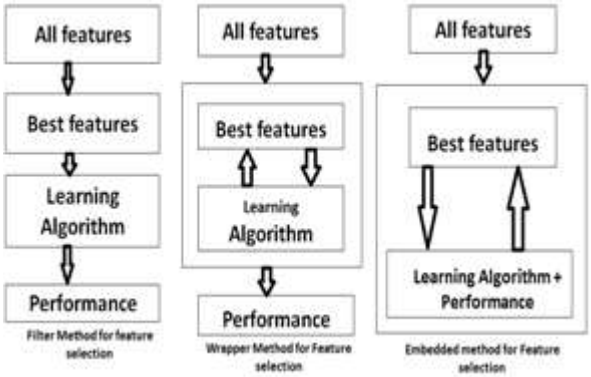

Figure 1: Basic feature selection algorithm for selecting the best features

4. Performance Criterion

Number of evaluation criterion was computed to evaluate the performance of the developed models. The set of criterion used are given as follows:

• Sensitivity (Sens):
Sens = TP/(TP + FN) (1)
• Specificity (Spec):
Spec = TN/(FP + TN) (2)
• Positive Predicted Value (PPV):
PPV = TP/(TP + FP) (3)
• Negative Predicted Value (NPV):
NPV = TN/(FN + TN) (4)
• Accuracy (Acc):
Acc = (TP + TN)/(TP + TN + FP + FN) (5) • True Positive (TP):
Sick people correctly diagnosed as sick.
• False Positive (FP):
Healthy people incorrectly identified as sick.
• True Negative (TN):
Healthy people correctly identified as healthy.
• False Negative (FN):
Sick people incorrectly identified as healthy.

5. Area of Study and Data

Pima Indian is a homogeneous group that inhabits the area around American, but they are popular for being the most infected group with type II diabetes. Pima Indians diabetes data can even be retrieved from UCI Machine Learning Repository's web site [30]. So, they are subject of intense studies in type II diabetes. The data consist of eight input variables and one output(0,1).

Table 2: Inputs and Output For Diabetic Prediction Model

Inputs	The number of times pregnant	X1
	The results of an oral glucose tolerance test	X2
	Diastolic blood pressure (mm/Hg)	X3
	E-Triceps skin fold thickness (mm)	X4
	2-h serum insulin (micro U/ml)	X5
	Body mass index	X6
	Diabetes pedigree function	X7
	Age (year)	X8
Output	Predicted class	X9

5.1. Experimental Result

In this paper many classifier are used for classification of Pima Indian diabetes data set. As in [4] the best feature selected are (The results of an oral glucose tolerance test,Diastolic blood pressure (mm/Hg),Bodymass index,Age (year)). Table (3) show the aalgorithm and the accuracy obtain with 8 feature and Table (4) show the result obtain with 4 feature (the best features).

Table 3: Classification Of Pima Indian Diabetes Data Set Using 8 Features

Performance Criterion / Method	Sensitivity	Specificity	Positive Predicted Value	Negative Predicted Value	Accuracy
J48	0.627	0.834	0.669	0.807	76.1719
RANDOM TREE	0.563	0.786	0.585	0.771	70.8333
REP TREE	0.571	0.826	0.638	0.782	73.6979
Random Forest	0.586	0.840	0.662	0.791	75.1302
Hoeffding Tree	0.608	0.846	0.679	0.801	76.3021
Decision Stump	0.649	0.774	0.606	0.805	73.0469
Multilayer Perceptron	0.627	0.830	0.664	0.806	75.9115
support Vector Machine	0.537	0.902	0.746	0.784	77.474
Decision Table	0.560	0.848	0.664	0.782	74.7396
JRip	0.597	0.824	0.645	0.792	74.4792
OneR	0.440	0.868	0.641	0.743	71.875
PART	0.582	0.816	0.629	0.785	73.4375
ZeroR	0.000	1.000	0.0000	0.651	0.65.1042

Table 3: Classification of Pima Indian Diabetes Data Set Using 4 Features

Performance Criterion / Method	Sensitivity	Specificity	Positive Predicted Value	Negative Predicted Value	Accuracy
J48	0.590	0.858	0.690	0.796	76.4323
Random Tree	0.560	0.762	0.558	0.764	69.1406
REP Tree	0.578	0.856	0.683	0.791	75.9115
Random Forest	0.567	0.828	0.639	0.781	73.6979
Hoeffding Tree	0.582	0.864	0.696	0.794	76.5625
Decision Stump	0.653	0.762	0.595	0.804	72.3958
Multilayer Perceptron	0.616	0.840	0.673	0.803	76.1719
support Vector Machine	0.526	0.894	0.727	0.779	76.5625
Decision Table	0.586	0.828	0.646	0.789	74.349
JRip	0.582	0.848	0.672	0.791	75.5208
OneR	0.444	0.858	0.626	0.742	71.3542
PART	0.649	0.772	0.604	0.804	72.9167
ZeroR	0.000	1.000	0.000	0.651	65.1042

6. Conclusions and Future Work

Data mining and machine learning algorithms in the medical field extracts different hidden patterns from the medical data. They can be used for the analysis of important clinical parameters, prediction of various diseases, forecasting tasks in medicine, extraction of medical knowledge, therapy planning support and patient management. A number of algorithms were proposed for the prediction and diagnosis of diabetes. These algorithms provide more accuracy than the available traditional systems.

References

[1] M. Ahmed Awad, "Diabetes care in sudan emerging issues and acute needs," Diabetes Voice, vol. 51, no. 3, pp. 1–15, 2006.
[2] M. H. Crawford, Current diagnosis treatment in cardiology. New York: Lange Medical Books/McGraw-Hill, 2nd ed., 2003. "A Lange medical book".
[3] B. M. Patil, R. C. Joshi, and D. Toshniwal, "Hybrid prediction model for type-2 diabeticpatients," Expert Syst. Appl., vol. 37, no. 12, pp. 8102– 8108, 2010.
[4] A. S. Ahlam A. Sharief, "Developing a mathematical model to detect diabetes using multigene genetic programming," InternationalJournal of Advanced Research in Artificial Intelligence, vol. 3, no. 10, 2014.
[5] WHO, "World health organization," 2016.
[6] C. Velu and K. Kashwan, "Visual datamining techniques for classification of diabetic patients," in Advance Computing Conference (IACC), 2013 IEEE 3rd International, pp. 1070–1075, IEEE, 2013.
[7] V. VijayanV and A. Ravikumar, "Study of data mining algorithms for prediction and diagnosis of diabetes mellitus," International Journal of Computer Applications, vol. 95, no. 17, pp. 12–16, 2014.
[8] S. F. B. Jaafar and D. M. Ali, "Diabetes mellitus forecast using artificial neural network (ann)," in 2005 Asian Conference on Sensors and the International Conference on New Techniques in Pharmaceutical and Biomedical Research, pp. 135–139, Sept 2005.
[9] T. Jayalakshmi and A. Santhakumaran, "A novel classification method for diagnosis of diabetes mellitus using artificial neural networks," in 2010 International Conference on Data Storage and Data Engineering, pp. 159–163, Feb 2010.
[10] R. Dey, V. Bajpai, G. Gandhi, and B. Dey, "Application of artificial neural network (ann) technique for diagnosing diabetes mellitus," in 2008 IEEE Region 10

and the Third international Conference on Industrial and Information Systems, pp. 1–4, Dec 2008.
[11] D. Michie, D. J. Spiegelhalter, C. C. Taylor, and J. Campbell, eds., Machine Learning, Neural and Statistical Classification. Upper Saddle River, NJ, USA: Ellis Horwood, 1994.
[12] H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases," Expert Syst. Appl., vol. 35, no. 1-2, pp. 82– 89, 2008.
[13] B. M. Patil, R. C. Joshi, and D. Toshniwal, "Hybrid prediction model for type-2 diabetic patients," Expert Syst. Appl., vol. 37, no. 12, pp. 8102– 8108, 2010.
[14] B. Prasad, P. Lingras, and A. Ram, eds., Proceedings of the 4th Indian International Conference on Artificial Intelligence, IICAI 2009, Tumkur, Karnataka, India, December 16-18, 2009, IICAI, 2009.
[15] G. P. Zhang, "An investigation of neural networks for linear time-series forecasting,"Computers & OR, vol. 28, no. 12, pp. 1183–1202, 2001.
[16] L. M. Silva, J. M. de S´a, and L. A. Alexandre, "Data classification with multilayerperceptrons using ageneralized error function," Neural Networks, vol. 21, no. 9, pp. 1302–1310, 2008.
[17] M. Khashei and M. Bijari, "An artificial neural network model for timeseries forecasting," Expert Syst. Appl., vol. 37, no. 1, pp. 479–489, 2010.
[18] P. G. Benardos and G.-C. Vosniakos, "Optimizing feedforward artificial neural network architecture," Eng. Appl. of AI, vol. 20, no. 3, pp. 365– 382, 2007.
[19] M. Tsypin and H. R¨oder, "Method for reliable classification of samples in clinical diagnostics using an improved method of classification," Sept. 20 2011. US Patent 8,024,282.
[20] W. A. Chaovalitwongse, Y.-J. Fan, and R. C. Sachdeo, "On the time series k-nearest neighbor classification of abnormal brain activity," IEEE Transactions on Systems, Man, and Cybernetics, Part A, vol. 37, no. 6, pp. 1005–1016, 2007.
[21] G. G. Enas and S. C. Choi, "Choice of the smoothing parameter and efficiency of k-nearest neighbor classification," Computers Mathematics with Applications, vol. 12, no. 2, pp. 235 – 244, 1986.
[22] V. Vapnik, Statistical learning theory. Wiley, 1998.
[23] F. Murtagh and M. M. Farid, "Pattern classification," J. Classification, vol. 18, no. 2, pp. 273–275, 2001.
[24] A. M. Andrew, "An introduction to support vector machines and other kernel-based learning methods by nellochristianini and john shawetaylor, cambridge university press, cambridge, 2000, 189 pp., isbn 0521-78019-5," Robotica, vol. 18, no. 6, pp. 687–689, 2000.
[25] T. A. Munandar and E. Winarko, "Article: Regional development classification model using decision tree approach," International Journal of Computer Applications, vol. 114, pp. 28–33, March 2015. Full text available.
[26] A. Iyer, S. Jeyalatha, and R. Sumbaly, "Diagnosis of diabetes using classification mining techniques," CoRR, vol. .abs/1502.03774, 2015.
[27] E. D. beyli, "Combined neural networks for diagnosis of erythematosquamous diseases.," Expert Syst. Appl., vol. 36, no. 3, pp. 5107–5112, 2009.
[28] A. Parashar, K. Burse, and K. Rawat, "A comparative approach for pima indians diabetes diagnosis using lda-support vector machine and feed forward neural network," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, pp. 378–383, 2014.
[29] S. Vanaja and K. R. Kumar, "Analysis of feature selection algorithms on classification:asurvey,"InternationalJournalofComputer Applications, vol. 96, no. 17, 2014.
[30] M. W. Aslam and A. K. Nandi, "Detection of diabetics using genetic programming," in European Signal Processing Conference, no. 18, (Aalborg, Denmark), August 2010.