Diabetes Diagnoser: Expert System for Diagnosis of Diabetes Type-II

Harsha Sethi
Department of Electronics & Communication Engineering
Rayat Bahra College of Engineering and Technology,
Ropar Campus,
Punjab Technical University, Punjab

ABSTRACT
The paper presents an expert system for prognosis of type II diabetes using artificial intelligent techniques. Diabetes is a chronic condition that occurs when the body cannot produce enough or cannot effectively use insulin which is a hormone produced by pancreas and is needed by our body to metabolize glucose. According to International diabetes federation (IDF) in 2015 around 415 million people suffered from diabetes and it is estimated that the prevalence of this disease will increase in the years to come. The expert system presented in this work can be used as a tool for initial screening of people who suffer from this disease and as such can be an effective aid for mitigating the mortality due to this disease. In addition to this we have also performed a comparative analysis of multiple machine learning based techniques for prognosis of diabetes. We have tested this expert system using ten-fold cross validation and comparing its results with the actual diagnosis of the patients. The results obtained indicate that diabetes diagnoser is proficient tool for prognosis of diabetes type II and can be used as an effective aid in primary level screening of diabetes. Among all the algorithms implemented for the problem under consideration, the artificial neural network outperformed & expressed an efficiency of about 96% for followed by J-48 Graft, END, and Decorate, Random forest, Bagging, Multi class classifier, Multi-boost Classifier, User Classifier, Decision Stump and Random tree with their efficiencies in the range of about 91% to 88%.

Keywords
Diabetes diagnoser, Artificial intelligence, Artificial neural networks, Diabetes type-2, Machine learning.

1. INTRODUCTION
Diabetes is a very common problem in today's advancing age group. With increasing life expectancy, sedentary lifestyle and high incidence of obesity its prevalence in the society is further increasing. Approximately 66 million people above 50 to 60 years suffer from this disease [1]. It is the most common disease which occurs to every 2 out of 3 persons and mostly searched on World Wide Web, more than three billion people search it each day. It is kind of slow poison which is broadening day by day. Diabetes is basically a chronic condition that affects the way the body processes glucose. In this type of disease human body is not able to produce sufficient insulin that is required by the body or is not able to utilize the insulin produced [2]. The frequent symptoms of diabetes are increased thirst, frequent urination, increased Hunger, Blurred visions, Fatigue etc. The diabetes being a lifestyle disorder is usually controllable by medicines and by managing the eating habits and regular physical activity. The conventional way of diagnosis of diabetes is by pathological test of checking the glucose level in the blood. If the glucose level in the human body remains continuously above the normal range as it is required, it is confirmed that the person is suffering from this disease.

Artificial intelligence which is sometimes also referred as "synthetic intelligence" brings out the branch of electronics and science together to make the machines as intelligent as humans. It is the name of the academic field of study which studies how to create computers and computer software that are capable of exhibiting intelligent behavior. It is usually defined as "the study and design of intelligent agents", in which an intelligent agent is a system that perceives its environment and takes actions that maximize its chances of success [3]. Over the last few decades artificial intelligence techniques have been increasingly used in solving problems in medical domains such as in Oncology, Urology, Liver Pathology, Cardiology, Gynecology, Thyroid disorders, Perinatology etc. The primary concern of AI in medicine is construction of artificial intelligent systems that can assist a medical doctor in performing expert diagnosis. These artificial intelligent systems support the clinical decision making by anticipating the diagnostic results, after being trained using previously acquired training data followed by expending specific information of some patient case. The use of Artificial intelligence in medicine has shown substantial progress in achieving timely, reliable diagnosis and more precise treatment of many diseases.

An expert system is a computer application that applies artificial intelligence (AI) to simulate the judgment and behavior of a human that has expert knowledge and experience in a particular field. It is a branch of applied artificial intelligence (AI), developed by the artificial intelligence (AI) community in the mid-1960s with an aim to transfer the expertise of a human into a computer. Characteristically, an expert system integrates an inference engine i.e. a set of rules in the form of a program for applying the knowledge obtained from a knowledge base that contains the accumulated experience. The expert systems these days are embedded with machine learning algorithms that allow then to learn from past experience just as is done by humans and thus improve their working efficiency with time [4].

2. RELATED WORK
Kemal Polat et.al. have proposed a system for diagnosis of diabetes using principal component analysis and adaptive neuro fuzzy inference system [12]. The author's have used the dataset obtained from UCI machine learning data repository. The system works in two phases, in first phase feature selection algorithm is used to reduce the size of the data set

for analysis from eight attributes to four attributes; in the second phase the data set with four attributes is fed to neuro fuzzy inference system for artificial intelligence based diagnosis of diabetes. They have reported that their system has presented an efficiency of 89.47% which is better than the other systems reported in the literature. Chad Ton Su et.al. have proposed a system for prediction of diabetes type-II using anthropometrical body[13]. They have used four data mining approaches including dignitary, logistic regression, neural networks and rough set for feature selection and reducing the size of data base for analysis. The results of their study have indicated that the volume of trunk, left and right thigh circumference, waist circumference, volume of right leg and subjects age are the factors /parameters that play key role in the manipulation of disease. They have also reported that the classification by dignitary and rough set is much better as compared to the classification of back propagation and logical regression. Mohammad Amine chikh has proposed a modified version of Artificial Immune Recognition system-2 (AIRS-2) which they have called MAIRS-2[14]. In this modified version they have replaced the KNN nearest algorithm with fuzzy k nearest neighbor algorithm with an aim to improve the diagnostic accuracy of diabetes. They have used the pima Indian diabetes dataset obtained from UCI machine learning repository. They have evaluated the performance of MAIRS-2 using ten-fold cross validation on the performance parameters like sensitivity, accuracy and specific values. The authors have reported that the MAIRS-2 presented an accuracy of 89.10% as compared to AIRS-2 which presented an accuracy of 82.69%. T Jayalaxmi has stressed on management of missing values in various dataset used for analysis of medical disorders[15]. She has also investigated the impact of data pre-processing techniques on various classification algorithms. She has used some pre-processing techniques on pima Indian diabetes dataset and reported that the performance of algorithms has improved considerably after pre- processing. Abid sarwar et.al have performed a comparative analysis of machine learning techniques for prognosis of type-II diabetes [16]. Authors have identified multiple features, after consultation with medical experts of the concerned domain, which play a important role in manipulation of diabetes .On the basis of these parameters they prepared a primary database for analysis. They applied three machine learning algorithms and reported that among all the algorithms artificial neural network outperformed. Seral sahan et.al. have tested the performance of a novel artificial immune system proposed by them named as attribute weighted artificial immune system (AWAIS) using multiple performance metrics[17]. Authors have used two dataset like statlog – heart disease dataset and pima Indian diabetes dataset obtained from UCI machine learning repository and evaluated the proposed system through ten-fold cross validation method. They have reported that AWAIS reached an accuracy of 82.59% for statlog heart disease dataset while as 75.87% for pima Indian diabetes dataset. R Bharat Rao et al**.** have proposed an artificially intelligent system (LungCAD) that helps in the detection of Lung cancer[18]. They have applied a classification algorithm for detecting solid pulmonary nodules from CT thorax studies. The LungCAD system was clinically tested by a number of radiologists and was found to deliver significantly greater accuracy both in detecting the affected nodules and in identifying the potentially actionable nodules. LungCAD was approved by FDA in 2006. Zhi-Hua Zhou et. al. have proposed an automatic pathological diagnostic procedure based on the ensemble of neural networks for identification of lung cancer cells in the images of the specimen of needle biopsies obtained from the bodies of the subjects to be diagnosed [19]. They named this system as neural ensemble based detection (NED). They have build the ensemble on a two level based architecture among which the first level ensemble is used to judge whether the cell is normal or cancerous with high level of confidence. The second level deals with the cells which are marked as cancerous by the level one ensemble. The second level ensemble has five outputs viz adenocarcinoma, squamous cell carcinoma, small cell carcinoma, large cell carcinoma and normal. The authors have reported that the neural ensemble based detection (NED) has not only shown a high rate of overall identification but also a very low rate of false negative identification. Abid Sarwar et.al. presented a novel hybrid ensemble technique i.e. ensemble of ensemble methods for improving the predictive performance of Artificial intelligence based system for screening of cervical cancer by characterization and classification of Pap smear images [20]. Using such a technique, the classification potentials of individual algorithms are fused together to gain greater classification accuracy. In addition to this they have also presented a comparative analysis of various artificial intelligence based algorithms for screening of cervical cancer [21]. For evaluation of the predictive performance of the said technique, authors have used root mean square error and percentage of correct classification as the measures. The results indicate that hybrid ensemble technique is an efficient method for classification of Pap smear images and hence can be effectively used for diagnosis of cervical cancer. Among all the algorithms implemented, the hybrid ensemble approach outperformed & expressed an efficiency of about 96% for 2-class problem and about 78% for 7-class problem. The results when compared with the all the standalone classifiers were significantly better for both 2-class and 7-class problems.

3. MATERIAL & METHODS

3.1 Database used for study

In order to perform the research reported in this manuscript, we first studied the medical literature of diabetes and the related research work being carried out in this domain[5]. We also consulted the medical experts of the concerned domain and discussed with them the problem under consideration. After a detailed conversation, we identified ten physiological parameters that play a pivotal role in the manipulation of diabetes .The parameters selected are illustrated in table 1, which also presents a interval-frequency data about the various parameter values in the database. On the basis of these parameters authors prepared a rich database of about 400 people from across a wide geographical region. During the preparation of database it was ensured to have diversity and variety in the database in terms of all the parameters under consideration. The database included both categories of people i.e. diabetic and non-diabetic. To populate the database with variety of instances we collected data from different sections of society i.e people residing in urban areas, rural areas, upper class, lower class, people from different age groups, people with different eating habits, cultures, smokers, non-smokers, drinkers, non-drinkers etc. The minimum and maximum age of subjects in the data set is five and 78 years, respectively. While constructing the database care was taken to assign discrete values to the parameters so as to maintain consistence in the database. Fig. 1 shows the screen shot of sample of the database used for study.

Table 1: Details of the various parameters used and their analysis

Physiological parameter	Description	Range of values	Analysis of data
Age	Age of the Person	5 to 78	Age 5 to age 20 : 20 Age 21 to age 35 : 131 Age 36 to age 50 : 142 Age 51 to age 78 : 95
Gender	Gender of the person	0 or 1	Male: 183 (represented by 1) Female: 205 (represented by 0)
Smoking	Whether the person is smoker or not	0 or 1	Smokers: 58 Non-Smokers: 330
Drinking	Drinker or non drinker	0 or 1	Drinkers: 79 Non-Drinkers: 319
Urination	How many times person urinates in day	2- 10	1-5 : 194 6-10: 153 11-15: 41
Thirst	How many times person drinks	5 – 15	1-5: 93 6-10: 195
Height	Height of a person	60-185	60 – 95: 2 96 – 125: 4 126 – 155: 118 156 – 185: 264
Fatigue	Healthy levels of fat mass for a fit person	0 or 1	Fatigue(Yes): 269 Fatigue(No): 119 Min-5% in men, 12%in women Max-25% in men ,32%in women Average-15 to 18% in men, 22 to 25% in women
Weight	Weight of the person	15 to 96	15 – 36: 11 37 – 56 : 110 57 – 76 : 244 77 – 96: 23 Average weight-62 kg Overweight-34.7 %
Family History	Any person in family is diabetic or not	0 or 1	Family History(Yes): 110 Family History(No): 278
Diabetic	If a person is diabetic or not	0 or 1	Diabetic: 295 Non-Diabetic: 193

Age	Sex	Family	Smoking	Drinking	Thirst	Urination	Height	Weight	Fatuge	Diabetic
41	1	1	1	1	8	10	173	55	1	1
68	1	0	0	0	4	3	172	80	1	0
35	0	0	0	0	3	3	162	70	1	0
40	0	0	0	0	4	3	170	49	1	0
70	0	0	0	0	10	10	185	65	1	0
27	0	0	0	0	4	3	154	48	0	0
16	0	0	0	0	6	3	167	47	1	0
26	0	1	0	0	5	3	160	56	0	0
36	1	0	0	1	8	12	170	85	1	1
45	1	0	1	1	7	10	172	69	1	1
12	0	1	0	0	5	3	147	34	0	0
38	1	1	0	1	15	10	172	70	1	1
46	1	0	0	1	7	5	170	80	1	1
46	1	0	0	1	7	5	170	80	1	1
30	1	0	1	1	4	4	185	80	1	0
49	1	0	1	1	5	7	170	70	1	1
54	0	1	0	0	6	9	154	59	1	1
44	1	0	0	0	5	4	162	55	0	0
36	0	0	0	0	8	10	144	63	1	1
36	1	0	1	1	5	4	167	55	0	0
33	1	0	1	0	5	9	173	63	1	1
44	0	0	0	0	5	13	157	80	1	1
66	0	0	0	0	2	3	157	40	1	0
53	1	0	1	1	14	3	171	63	1	1
16	0	0	0	0	6	4	157	64	0	0
22	0	0	0	0	5	4	154	41	0	0
24	1	1	0	0	6	6	167	70	1	1

Fig 1: Screen shot of a sample of database used for training of Algorithms

In order to perform the research which is presented in this manuscript, we prepared the data containing sets of 400 different people from the society so that variety of data could be ensured. The data set we choosen involves 10 pysiological parameters which descripts indispensible role in the declaration of diabetes. The values contain in data is taken after the detailed discussion with expert consultants we come to an end with this statistical analysis given above.

Among the various parameters selected for study; Age has a pivotal role to play. According to analysis type II diabetes typically accounts for more than 90-95% after the age of 35 but in the case of type I diabetes, it has been observed that, it is more prevalent in children than in adults. It has been observed that the people whose sibling or parents has/had suffered from diabetes, are 30-35 % times more susceptible to diabetes than a person without such family history. People who suffer from diabetes are not able to manage and maintain the levels of the sugar in their blood, as such usually report an increased desire for thirst. There is also a change in the body weight of the persons suffering from diabetes[6]. According to a report of the American diabetes association around 50% of men and 70% of women become obese at onset of diabetes after bariatric surgery[7]. It has been observed that women are more at risk of developing diabetes as compared to men; this is attributed to usually lesser amount of the physical activity in women than in men. Some studies have also revealed that smoking habits or consumption of tobacco may be an independent risk factor for diabetes type II. It has been observed that people who do not smoke or use tobacco are at lower risk as compared to the other counterparts. Many studies have reported that moderated consumption of alcohol is associated with an increased sensitivity towards insulin and as such lowers down the risk of developing diabetes. India being one of the countries sharing a highest amount world's diabetic population, so the collection of data from the subjects for study was easy.

3.2 Algorithms used for study

For designing the diabetes diagnoser we have used three algorithms namely Artificial Neural Networks, K-nearest neighbor and Naïve Bayes. These algorithms were selected for inclusions in this study because of their popularity in the recently published literature as well as their good performance even in fewer amounts of training data[8]. Below given is a short description about the selected algorithms for study.

3.2.1 K-nearest neighbor

K nearest neighbor is used for competitive processing and uses topologies which are used in efficient lazy learning algorithms. In this algorithm the units are located next to each other and respond to input vector. For one or two dimensional it is easy to visualize the data using Self organizing maps that groups the input data into clusters. Given N training vectors, K-NN algorithm identifies the K nearest neighbors of the test data regardless of labels, and classifies the test data into one of the possible classes by taking the votes from all the k-nearest neighbors. It has been observed that K-NN is being increasingly used in estimating statistical data and recognizing the data patterns. K-NN being a lazy learner, trains the dataset that is stored on querying similar data between test data and training data records. K-NN is termed as lazy and non parametric learner [9] because it only stores the trained database and no general model from training dataset is

constructed as is in the case of eager learners like decision trees.

3.2.2 Naïve Bayes Algorithm

Naive bayes algorithm is based on three concepts-Antecedent, feasibility and prediction where antecedent means some past information about the incident, feasibility means chances of that event to occur in future and prediction means some forecasting made about the occurrence of that event on basis of first two concepts. The relation between the three variables is given by:-

Prediction= Antecedent * Feasibility / Corroboration

Mathematically the above relation can be represented as:-

Probability (B Given A) = Prior_Probability * Probability (A And B)/ Probability (A)

Naı¨ve Bayes is considered to be one of the most efficient and effective inductive learning algorithms for machine learning and data mining[10]. Naı¨ve Bayes classifier is known for its exceptional performance even in less amounts of training data. In 2006, Rich Caruana and Alexandru Niculescu-Mizil did a comprehensive comparison of many classification algorithms and showed that Naı¨ve Bayes outperformed many of the counterparts such as boosted trees or random forests. Its competitive performance is attributed to its principle of conditional independence assumption, which means that it assumes that the presence or absence of some parameters of a class to be independent to the presence or absence of some other parameters. Thus, each parameter has independent contribution to the prediction of the final result.

3.2.3 Artificial neural networks

Artificial neural network is a machine learning method inspired and evolved from the structure, function and working of human brain. ANN can be visualized as a data processing system consisting of a large number of simple, highly interconnected processing elements in an architecture inspired by the cerebral cortex of brain. The output of a neural network relies on the cooperation of the individual neurons within the network to operate. Well designed neural networks are trainable systems that can often "learn" to solve complex problems from a set of examples and generalize the "acquired knowledge" to solve unforeseen problems, i.e. they are self-adaptive systems. These processing elements are usually organized into a sequence of layers with full or random connections between various layers. Each neural network has three critical components; nodal character, network topology and learning rules. Nodal character determines how signals are processed by the node, such as the number of inputs and outputs associated with the node, weights associated with each input and output and the activation function. Network topology determines the fashion in which the nodes are connected. Learning rules determine how the weights are initialized and adjusted[11]. The input layer is not the neural computing layer because the node doesn't have the input weights and also they don't have any activation function. The top layer is the output layer that presents the response for the input fed to the network. The other layers are called the hidden or intermediate layers as they don't have any connection with inputs and outputs of the network.

4 DEVELOPMENT OF DIABETES DIAGNOSER AND COMPARISON WITH OTHER ALGORITHMS

For realization of algorithms used for the development of Diabetes diagnoser, author wrote program in MATLAB (Matrix Laboratory) in such a way that all the three algorithms work on a centralized primary database prepared for study. The interface of the diabetes diagnoser has been so developed, as it is easy and intuitive to use. The interface of diabetes diagnoser contains multiple drop down menus and text boxes for facilitating the entry of ten physiological parameters that have been considered for the study. There is a sub-module for conversion of the height of the subject from foot to centimeter. The text data fed by the user are converted into corresponding coded integer values for all the possible values in the domain. Once all the values are fed in the interface of the expert system, they are converted into a matrices-vector representing the subject in physiological characteristics. This vector is fed to the trained algorithms that work as the backbone of the expert system for the prognosis of diabetes. The diabetes diagnoser presents the result to the user in terms of Yes / No in the interface of the system. Figure 2 shows the screenshot of the matlab program of diabetes diagnoser in execution.

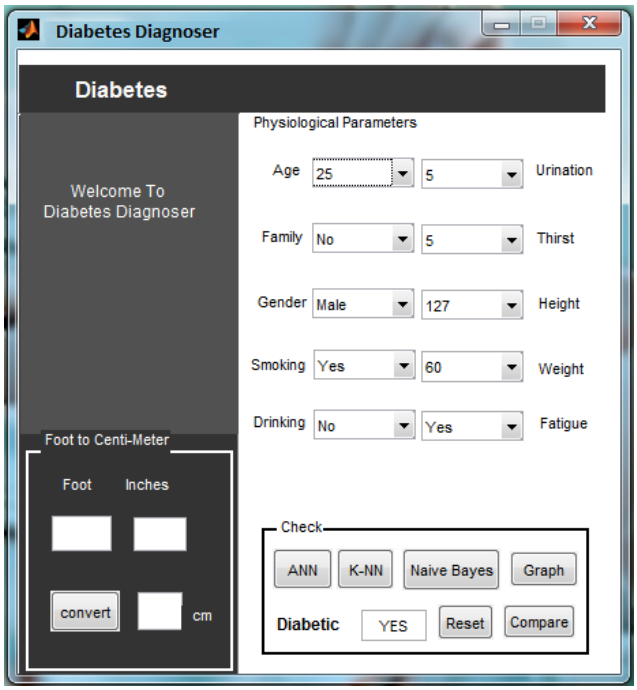

Fig 2.: Diabetes diagnoser in execution

Authors have also compared the working efficiency of the Diabetes Diagnoser with ten other well known classifiers in the domain of artificial intelligence. For realization of these algorithms WEKA workbench (version 3.6.10) was used. WEKA is an application written in Java & used machine learning community for various research goals. The tool facilitates the application of artificial intelligence based algorithms for analysis of data and designing of various predictive models as solution for some real world problems. The dataset was first converted into the required Attribute-Relation File Format (ARFF) format, for feeding into the WEKA workbench. ARFF format is an ASCII text file format that describes a list of cases sharing a set of attributes. For training and testing of multiple algorithms the dataset was divided in 80:20 ratio (311 instances for training and 77

instances for testing). The final results were evaluated on the basis of ten fold cross validation.

5 RESULTS AND DISCUSSION

In this paper author have presented an expert system for diagnosis of diabetes type-II and compared its efficiently with multiple machine learning techniques. For evaluation the efficiency of the system author used various performance metrics viz accuracy, true positive rate, false positive rate, precession and ROC curve analysis across ten fold cross validation. The reliability of the system was evaluated by computing the mean absolute error between the predicted values and actual values in the test data. The results indicate that the Diabetes diagnoser is an efficient tool for diagnosis of diabetes type-II. Among all the algorithms implemented Artificial Neural Networks performed the best prediction with an accuracy of about 96 % followed by J-48 Graft, END, and Decorate, Random forest, Bagging, Multi class classifier, Multi-boost Classifier, User Classifier, Decision Stump and Random tree. The working efficiencies of these algorithms can be further increased by increasing the number of instances in the database and by including various clinical features that play some role in diagnosis of diabetes. The results obtained are illustrated in table 2.

Table 2: Results obtained from various algorithms in terms of performance metrics

ALOGRITHMS USED	CORRECTLY CLASSIFIED	INCORRECTLY CLASSIFIED	TRUE POSITIVE RATE	FALSE POSITIVE RATE	ROC	PRECISION
Artificial neural Network (Implemented in diabetes diagnoser)	96.02%	3.98%	0.96	0.031	0.971	0.931
Bagging	89.69%	10.30%	0.86	0.067	0.941	0.927
Decision stump	88.65%	11.34%	0.891	0.118	0.858	0.882
Decorate	91.23%	8.76%	0.881	0.056	0.962	0.939
END	91.23%	8.76%	0.896	0.072	0.921	0.925
User classifier	88.65%	11.34%	0.891	0.118	0.858	0.882
J48 Graft	91.49%	8.50%	0.896	0.067	0.921	0.93
Multi class classifier	89.69%	10.30%	0.907	0.113	0.954	0.888
Multi boost	88.65%	11.34%	0.896	0.123	0.932	0.878
Random Tree	88.40%	11.59%	0.87	0.103	0.888	0.894
Random forest	90.97%	9.02%	0.896	0.077	0.955	0.92

6 REFERENCES

[1] Tommy LS Visscher and Jacob C Seidell, ”The Public Health Impact of Obesity”, Annual Review of Public Health Vol. 22: 355-375,

[2] Abdulfatai B. Olokoba,Olusegun A.Obateru and Lateefat B.Olokoba, ‘Type II Diabetes Mellitus:A review of current trends’;Oman Medical Journal.2012;27(4):269-27.

[3] Jorg P.Muller, ‘The Design of intelligent agents-A layered approach’;Book lecture notes in computer science volume 1177 1996;ISBN:978-3-540-62003-7(print) 978-3-540-49590-1

[4] R.S. Michalski,J.G Cabonell,T.M. Mitchell; ‘Machine Learning: An Artificial Intelligence Approach’;Springer Science & Business Media,17-April-2013-Computers-572 pages.

[5] Subbiah, A., & Subbiah, G. (2002). Diabetes research in India and China today: from literature-based mapping to health-care policy.

[6] El-Khatib, F., Rauchenzauner, M., Lechleitner, M., Hoppichler, F., Naser, A., Waldmann, M., ... & Luef, G. J. (2007). Valproate, weight gain and carbohydrate craving: a gender study. *Seizure*, *16*(3), 226-232.

[7] Buchwald, H., Estok, R., Fahrbach, K., Banel, D., Jensen, M. D., Pories, W. J., ... & Sledge, I. (2009). Weight and type 2 diabetes after bariatric surgery: systematic review and meta-analysis. *The American journal of medicine*,*122*(3), 248-256..

[8] Vijiyarani, S., & Sudha, S. (2013). Disease prediction in data mining technique–a survey. *International Journal of Computer Applications & Information Technology*, *2*, 17-21.

[9] Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003, November). KNN model-based approach in classification. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"* (pp. 986-996). Springer Berlin Heidelberg.

[10] Sopharak, A., Dailey, M. N., Uyyanonvara, B., Barman, S., Williamson, T., Nwe, K. T., & Moe, Y. A. (2010). Machine learning approach to automatic exudate detection in retinal images from diabetic patients. *Journal of Modern optics*, *57*(2), 124-135.

[11] Zain, A. M., Haron, H., & Sharif, S. (2009, May). Review of ANN technique for modeling surface roughness performance measure in machining process. In *2009 Third Asia International Conference on Modelling & Simulation* (pp. 35-39). IEEE.

[12] Goodarzi, M., Olivieri, A. C., & Freitas, M. P. (2009). Principal component analysis-adaptive neuro-fuzzy inference systems (ANFISs) for the simultaneous spectrophotometric determination of three metals in water samples. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *73*(4), 608-614.

[13] Su, C. T., Yang, C. H., Hsu, K. H., & Chiu, W. K. (2006). Data mining for the diagnosis of type II diabetes from three-dimensional body surface anthropometrical scanning data. *Computers & Mathematics with Applications*, *51*(6), 1075-1092.

[14] Chikh, M. A., Saidi, M., & Settouti, N. (2012). Diagnosis of diabetes diseases using an artificial immune

recognition system2 (AIRS2) with fuzzy k-nearest neighbor. *Journal of medical systems*, *36*(5), 2721-2729.

[15] Jayalakshmi, T., & Santhakumaran, A. (2010, February). A novel classification method for diagnosis of diabetes mellitus using artificial neural networks. In *Data Storage and Data Engineering (DSDE), 2010 International Conference on* (pp. 159-163). IEEE..

[16] Sarwar, A., & Sharma, V. (2014). Comparative analysis of machine learning techniques in prognosis of type II diabetes. *AI & society*, *29*(1), 123-129..

[17] Şahan, S., Kodaz, H., Güneş, S., & Polat, K. (2004, October). A new classifier based on attribute weighted artificial immune system (AWAIS). In*International Symposium on Computer and Information Sciences* (pp. 11-20). Springer Berlin Heidelberg.

[18] Rao, R. B., Bi, J., Fung, G., Salganicoff, M., Obuchowski, N., & Naidich, D. (2007, August). LungCAD: a clinically approved, machine learning system for lung cancer detection. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1033-1037). ACM.

[19] Zhou, Z. H., Jiang, Y., Yang, Y. B., & Chen, S. F. (2002). Lung cancer cell identification based on artificial neural network ensembles. *Artificial Intelligence in Medicine*, *24*(1), 25-36.

[20] Sarwar, A., Sharma, V., Gupta, R., "Hybrid ensemble learning technique for screening of cervical cancer using Papanicolaou smear image analysis Hybrid ensemble learning technique for screening of cervical cancer using Papanicolaou smear image analysis", Personalized Medicine Universe 4 (2015) 54 – 62

[21] Sarwar,A., Ali, M., Suri, J., and Sharma, V., 2015 Performance evaluation of machine learning techniques for screening of cervical cancer, In Computing for Sustainable Global Development (INDIACom), March 2015 2nd International Conference on (pp. 880-886). IEEE.