AIRO JOURNALS

Free / Unpaid
Peer Reviewed
Multidisciplinary
International

ISSN: 2320-3714
Volume:2 Issue: 2
May 2022
Impact Factor: 5.7
Subject Computer & Science

Classification and Prediction of Diabetes Using Machine Learning for Healthcare Applications

Ajay Vyas
Research Scholar
P HD comp sc
Sri Satya Sai Universitu Bhopal,M.P.
Dr. Jitendra Sheethlani
Guide Name

Abstract

Biotechnology and public medical care foundation enhancements have brought about a monstrous convergence of indispensable and delicate medical care information. Many captivating patterns for ahead of schedule and beginning ID and counteraction of various lethal infections have been revealed utilizing clever information examination procedures. Diabetes mellitus is a life threatening illness since it makes harm the heart, kidneys, and nerves, in addition to other things. A machine learning based method for diabetes arrangement, early stage distinguishing proof, and forecast is proposed in this work. It likewise shows a potential IoT based diabetes observing gadget for a sound and impacted individual to follow their blood glucose (BG) levels. Three unique classifiers were utilized to characterize diabetes: Random Forest (RF), Multilayer Perception (MLP), and Logistic Regression (LR). We utilized long transient memory (LSTM), moving normal (MA), and direct relapse for prescient examination (LR). The PIMA Indian Diabetes benchmark dataset is utilized for exploratory assessment. During the review, we found that MLP beat different classifiers with 86.08 percent exactness and LSTM expanded significant diabetes expectations with 87.26 percent precision. Also, an examination of the proposed procedure with existing cutting edge strategies has been performed, exhibiting the adaptability of the proposed approach in different general wellbeing applications.

Keywords: *Healthcare, Prediction, Diabetes, Machine Learning, Data mining*

1. Introduction

ISSN: 2320-3714
Volume:2 Issue:2
May 2022
Impact Factor: 5.7
Subject Computer & Science

Information mining joins factual examination, AI, and data set innovation to reveal examples and connections stowed away from huge datasets. [1]

Diabetes is a persistent sickness that can cause significant medical issues like coronary illness, kidney disappointment and visual deficiency. [4], [5] It is a significant reason for cardiovascular sickness (heart and cardiovascular illness). Diabetes carries the risk of micro vascular and macro vascular problems. As a result, diabetes is one of the major causes of death worldwide. According to statistics from 2011, almost 366 million people worldwide have diabetes. Diabetes is expected to climb to over 552 million people by 2030, according to projections.

Heart sickness refers to a huge variety of situations that have an effect on the coronary heart, such as chest pain, shortness of breath, coronary heart attack, and different symptoms. It covers a huge variety of coronary heart-associated illnesses. [6] When the blood deliver to the coronary heart muscular tissues is insufficient, chest pains occur. Heart sickness refers to quite a few troubles that have an effect on the coronary heart and its blood arteries. The phrase `cardiovascular sickness,' which refers to a sort of coronary heart illness, refers to a huge variety of problems that have an effect on the coronary heart, blood arteries, and the manner blood is pumped and circulated all through the body [7]. Over the ultimate ten years, coronary heart sickness has end up the pinnacle reason of dying worldwide. According to the World Health Organization, coronary heart sickness is the pinnacle reason of dying in each prosperous and low-earnings countries [8]. Individually, there are various processes within side the literature for diagnosing diabetes or coronary heart sickness. To our knowledge, there's no automatic prognosis method for diagnosing coronary heart sickness in diabetic sufferers primarily based totally on diabetes prognosis features.

This studies paper builds on our preceding work; Diagnosing Vulnerability of Diabetic Patients to Heart Diseases Using Support Vector Machines [24] and Diagnosing Heart Disease for Diabetic Patients Using Nave Bayes Method [23] to are expecting coronary heart sickness for diabetic sufferers the use of diabetic prognosis attributes.

2. Data Mining Technology

You can utilize information mining procedures to separate significant information from your clinical data set. It is an insightful computational investigation of a lot of information assortment that utilizes a mix of AI, measurable examination, and data set innovation to find examples and decides that can be utilized to direct future action choices. [2], [3]. The motivation behind information mining is to foresee designs and sum them up to different information. In the clinical field, clinical information mining is turning out to be increasingly significant. Information mining is an integral asset that assists organizations with zeroing in on the main information in their information distribution center [9].Data mining techniques enable organizations to proactively make knowledge-based decisions by predicting future trends and behaviors [10]. A variety of data

ISSN: 2320-3714
Volume: 2 Issue: 2
May 2022
Impact Factor: 5.7
Subject Computer & Science

mining techniques are available, each suitable for a particular domain application. The use of data mining in health care is very promising and useful. Automate the process of retrieving predictive data on large datasets.

The classification model and the evaluation model are two models used in data mining classification technology. In order to develop a classification predictive model, the classification model uses training data. The categorization efficiency is tested using the testing data set. Patients with heart disease symptoms are recruited from a diabetes healthcare institute. Classification algorithms such as Naive bayes and Support vector machine are used to predict whether a diabetic patient has heart disease by indicating levels.

3. Literature Review

Dilip Singh Sisodia, Deepti Sisodia [1] (2018): They used classification techniques such as nave Bayes, decision trees, and SVM to predict diabetes. The highest level of accuracy was 76.3 percent, as determined using the naive Bayes algorithm.

Prof Pramila N. Chawan, Tejas N. Joshi [2] (2018): They proposed an effective approach for diabetes early detection using algorithms such as SVM, ANN, and Logistic Regression.

Dr. Y. Jeevan Nagendra Kumar et al. [3] (2017): Hybrid machine learning approaches were used to do a map-centered spatial analysis of rainfall data from the AP and TS states.

Yogesh Kumar Rathore and Piranha Indore [4] devised a technique for diabetes prediction. For their system, they used methods like artificial neural networks and Bayesian networks. Their research focuses on recent advances in machine learning that have had a substantial influence on diabetes detection and diagnosis.

Dr. Y. Jeevan Nagendra Kumaret al. [5] (2016): For knowledge discretion supervision and grain access control.

For the investigation and prediction of diabetes, Akansha Rathore and Simran Chauhan [6] used decision tree and SVM classification methods. Their system is based on the R Framework.

A new symbol-based tree traverse searching strategy was proposed by Dr. Y. Jeevan Nagendra Kumaret al [7] (2014).

Hrushikesh N. Mhaskar et al [8] used the MATLAB data mining programmed to classify diabetic patients and forecast glucose levels.

Aishwariya R, Gayathri P, and Jaysankar N [9] presented diabetes prediction methods. Support Vector Machine was the promising machine learning algorithm they utilized (SVM).

ISSN: 2320-3714
Volume: 2 Issue: 2
May 2022
Impact Factor: 5.7
Subject Computer & Science

Dr. Y. Jeevan Nagendra Kumar and colleagues [10] (2013): Our world is progressing with a lot of concerns on bio-diversity, science and technology, and many more, and this article provided a detailed overview of the efforts in this field.

For diabetes prediction, Abdullah A. Aljumah et al [11] (2013) used regression-based machine learning approaches. They used the Oracle Data Mining programmed.

SVM was the analysis algorithm employed.

For their suggested work, IoannisKavakiotis et al [12] (2017) used the support vector machine (SVM) classification algorithm.

Y. Jeevan, M.D. Nagendra Kumar et al. [13] (2013) devised a greedy strategy for decreasing the number of points and employed spatial mining to find a solution.

NongyaoNai-aruna and RungruttikarnMoungmaia [14] (2015) created a technique to detect diabetes.

They employed the DT, ANN, Random Forest, LR, and NB classification algorithms.

Sajida Perveena et al[15] (2016) used three classifiers for diabetes prediction: bagging, adaboost, and J48 decision tree. J48 yielded superior results.

D. Asir Machine learning algorithms such as nave Bayes, random forest, and MPL are employed for diabetes analysis and prediction, according to Antony Gnana Singh et al[16] (2017). The random forest classification algorithm produced better results.

4. Physical Activity's Role in Diabetes Mellitus Prevention and Control

Physical activity is usually the first preventive and therapeutic method recommended by diabetics or pre-diabetic patients [47]. Exercise is an important part of the programmer for lifestyle changes in diabetes, cardiovascular disease, obesity, nutrition and medicine. Still, dealing with all the deadly illnesses is expensive. Diabetes mellitus, on the other hand, has developed as a devastating burden for a country's health and economy in this century.

According to the International Diabetes Prevention and Control Federation, diabetes will impact over 366 million people globally [49]. Diabetes affects more than 29 million people in the United States, according to the disease control and prevention centre [50]. While these worrying figures continue to rise, they will put a strain on economies all across the world. As a result, experts and healthcare professionals around the world are researching and developing guidelines to help prevent and control this deadly disease. Sato [51] conducted a comprehensive study in Japan on the importance of exercise prescription for diabetic patients. He recommended that people avoid

AIRO JOURNALS
Free / Unpaid
Peer Reviewed
Multidisciplinary
International
ISSN: 2320-3714
Volume: 2 Issue: 2
May 2022
Impact Factor: 5.7
Subject Computer & Science

sitting for lengthy periods of time and engage in physical activity every 30 minutes. Kirwan et al. [47] stressed the importance of regular exercise in the management and prevention of type 2 diabetes. They looked at the metabolic effect on diabetic patients' tissues in particular and discovered that people who exercise regularly have significant improvements. Moser et al. [48] also stressed the importance of regular exercise in enhancing the functionality of numerous body organs, as seen in Figure 1.

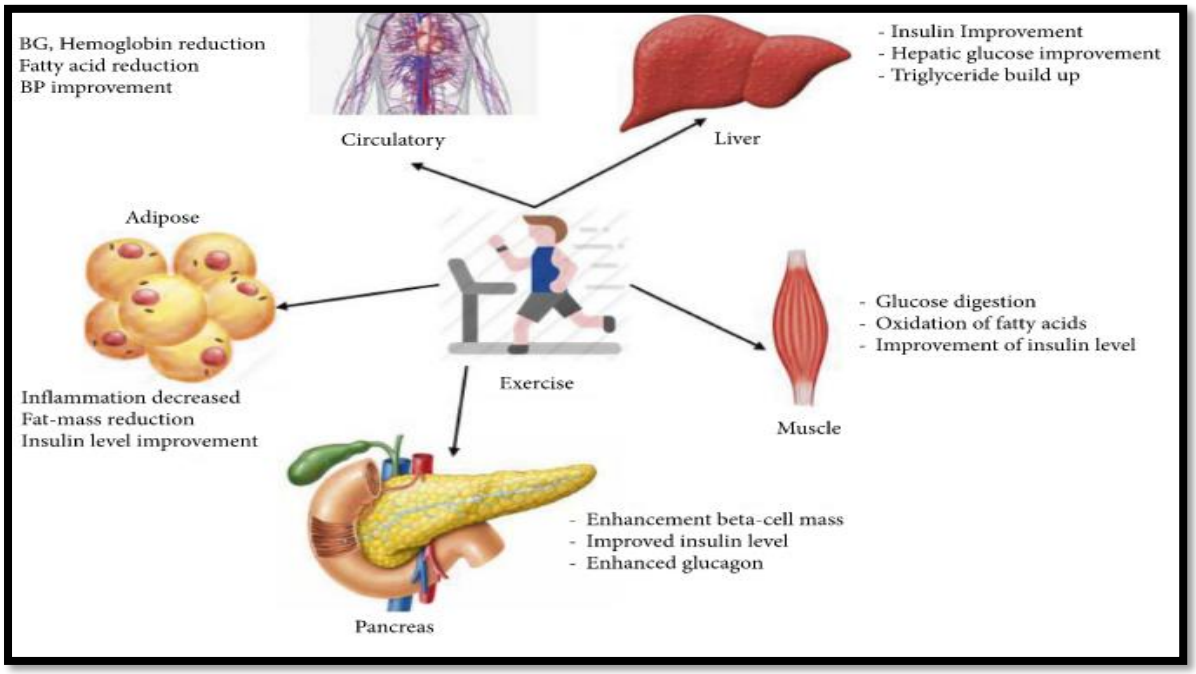

Figure: 1 Effect of regular exercise on metabolism in diabetics

Yang et al. [52] focused on exercise therapies that are important in the management of diabetes and its complications. They have discovered cytokines that provide new insights into the regulation of diabetes, but their sequences are still under investigation. Kim and John [53] provided a systematic review of the effects of various activities on improving metabolism in young diabetic patients. They cited various studies showing the effects of exercise on insulin, blood pressure, and blood sugar levels. However, none of these studies have addressed beta cell enhancement. As a result, many problems remain in the prevention and management of diabetes that require urgent attention from researchers around the world.

5. Healthcare Diabetic Classification and Prediction System Proposed

Different AI procedures are utilized in the proposed diabetes order and expectation framework. To begin with, we utilized calculated relapse, irregular woodland, and MLP to classify diabetes. MLPs have also performed well in health care, especially in the prediction of diabetes, and have been

AIRO JOURNALS
Free / Unpaid
Peer Reviewed
Multidisciplinary
International
ISSN: 2320-3714
Volume: 2 Issue: 2
May 2022
Impact Factor: 5.7
Subject Computer & Science

improved for classification [20, 21, 35, and 36]. Figure 2 and Method 1 show the proposed MLP architecture and algorithm, respectively.

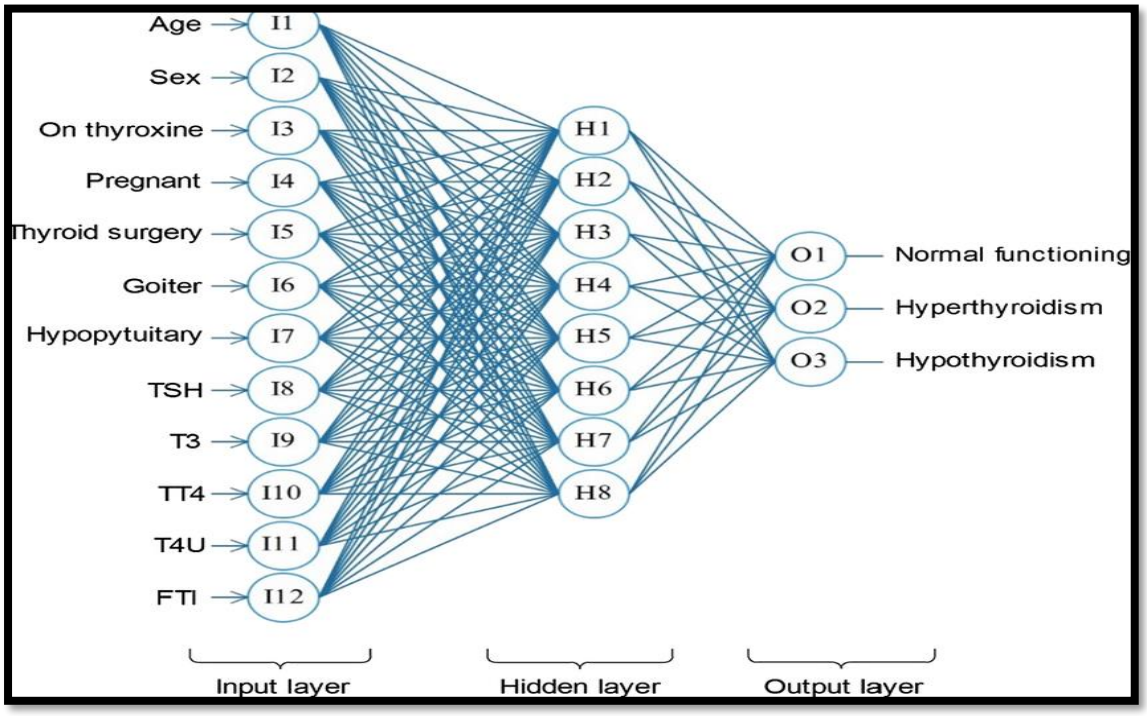

Figure: 2 Proposed MLP architecture for diabetes classification using eight factors as input.

Then, we will utilize moving midpoints, direct relapse, and LSTMs to foster three generally involved AI procedures for diabetes expectation. LSTMs center basically around wrongdoing expectation because of their outstanding presentation in true applications, particularly medical services [53]. The execution items in the proposed calculation are as per the following.

6. Research Methodology

The methodology discussed in this work is for detecting diabetic patients' vulnerability to heart disease, and we used 500 diabetic patient records to conduct the experiment. Table 1 lists the attributes that make up each record.

AIRO JOURNALS
Free / Unpaid
Peer Reviewed
Multidisciplinary
International
ISSN: 2320-3714
Volume: 2 Issue: 2
May 2022
Impact Factor: 5.7
Subject Computer & Science

Attribute Role	Attribute Name	Attribute Type	Description
Regular	Sex	binominal	Sex of the patient. Takes the following values: Male, Female
Regular	Age	integer	Age of the patient
Regular	Fam/Heri	polynomial	Indicates whether the patient's parents were affected by diabetes. Takes the following values: Father, Mother, Both
Regular	Weight	numeric	Weight of the patient
Regular	BP	polynomial	Blood Pressure of the patient
Regular	Fasting	integer	Fasting Blood Sugar
Regular	PP	integer	Post Prondial Blood Glucose
Regular	A1C	numeric	Glycosylated Hemoglobin Test
Regular	LDL	integer	Low Density Lipoprotein
Regular	VLDL	integer	Very Low Density Lipoprotein
Label	Vulnerability	nominal	Indicates the vulnerability of the patients to heart disease. Takes the following values: High, Low

Table: 1 Attributes used for the diagnosis.

142 of the 500 data were for patients who were very susceptible to cardiac disease. The remaining 358 data were for patients who were less likely to have heart disease. Since SVM parses only numeric attributes, the nominal values are replaced with unique integers and converted to numeric attributes. For example, the value of the attribute Sex is converted to male-1 and female-0.

The property values were then normalized to a range of 0 to 1. The SVM classifier was then given these records as input.

SVM performs classification by using kernel functions to map the input set to a high-dimensional data space. The types of kernel functions are: [22].

7. Result Analysis

AIRO JOURNALS
Free / Unpaid
Peer Reviewed
Multidisciplinary
International
ISSN: 2320-3714
Volume:2 Issue: 2
May 2022
Impact Factor: 5.7
Subject Computer & Science

The informational collection used to prepare the classifier comprises of 500 diabetic patient records, with 142 records addressing those with coronary illness (positive cases) and 358 records addressing those without coronary illness (negative cases).

After proper pre-processing, these records were used to train the SVM classifier.

A different value C for the RBF kernel parameter was used to train the SVM classifier. Next, the accuracy of the model obtained for each value of C was tested. Successful classifiers need to be able to achieve high accuracy with unknown datasets rather than training the data.

As a result, we employed 10 fold cross validation to verify the classifier's accuracy.

We partition the preparation set into 10 equivalent estimated subsets in 10-overlap cross-approval. The classifier prepared on the excess 9 subsets is tried consecutively on one subset. Accordingly, the cross-approval precision is the level of information that is appropriately ordered for each example of the entire preparation set. Over fitting is tried not to by utilize cross approval tests. In light of our broad testing, we found that the classifier had the best exactness of 94.60 percent for C = 5.0 and = 1.0. Table 2 shows the exactness acquired in our preliminaries for maybe a couple upsides of C and.

C Value	γ Value	Accuracy of the
.	.	.
2	0.125	89.60%
2	0.75	92.40%
4	2.5	93.20%
4	2	93.60%
4	1.5	93.80%
4	1	94.20%
5	**1**	**94.60%**
6	1.25	94%
.	.	.

Table: 2 Partial results of the trials conducted

8. Discussion

The utilizations of AI could be applied for the analysis of different infections, their side effects, their objective, their treatment. The abrupt passing's happening because of kidney disappointment, respiratory failure, strokes and so forth went with diabetes can be forestalled through early treatment and finding. In the review we saw different calculations, for example, SVM, choice tree, KNN, gullible straights, and so forth making their utilization in the

ISSN: 2320-3714
Volume:2 Issue:2
May 2022
Impact Factor: 5.7
Subject Computer & Science

expectation of rate of diabetes. The grouping procedures give various outcomes when applied to various dataset. We have found that different classification techniques are useful for different datasets. Fluctuations in model performance can be seen in different datasets and the root cause can be predicted accordingly. Future research may focus on collecting new insights and new datasets that lead to insights for improving diabetes prediction using machine learning techniques. Based on parameters such as age, classification of obesity index, obesity level, and history of chronic illness, combined with various machine learning techniques leads to better prediction levels. A new dimension that expands usage is deep learning. It can have amazing results in terms of pattern recognition and better predictions when machine learning is used.

9. Conclusion

We have discussed a method to assisting the healthcare area in this paper. The fundamental goal of this research is twofold. To begin, we suggested a Mapbase diabetes classification method as well as a profound learning based LSTM for diabetes expectation. Second, we proposed a speculative ongoing diabetic checking framework in light of the Internet of Things. To conjecture BG levels and diabetes, the proposed hypothetical diabetic observing framework will join a Smartphone, a belated sensor gadget, and AI based approaches in a continuous information handling climate. The proposed framework's significant objective is to permit clients to screen their important bodily functions using belated sensor gadgets and their phones. Besides, the recommended approach will help clients in deciding their diabetes risk at a beginning phase and getting future projections of their BG increment levels. MLP and LSTM have been adjusted for diabetic classification and expectation. On the PIMA Indian Diabetes dataset, the suggested philosophies are tried. The two methodologies surpassed best in class draws near, with precision evaluations of 86.083 percent and 87.26 percent, separately. Later on, we might want to utilize the gave arrangement and forecast calculations to foster an Android application for the proposed virtual diabetes observing framework. For better observing, hereditary calculations can be utilized with the proposed forecast system [24, 64, 66-71].

10. References

[1] Thuraisingham, B.: “A Primer for Understanding and Applying Data Mining”, IT Professional, pp: 28-31, 2000

[2] J. Han Kamber, M. 2006. Data Mining: Concepts and Techniques, 2nd ed. San Francisco: Morgan Kaufman.

[3] U. Fayyad, G.Piatetsky-Shapiro, and P.Smyth, “From Data Mining to Knowledge Discovery in Databases”, AI Magazine, Vol.17, pp.37-54, 1996.

ISSN: 2320-3714
Volume:2 Issue:2
May 2022
Impact Factor: 5.7
Subject Computer & Science

[4] World Health Organization. Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia: Available: http://www.who. int/ diabetes/en

[5] World Health Organization. Available: http: // www.who.int/topics/ diabetes mellitus/en/

[6] K.Srinivas et al. / "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks", International Journal on Computer Science and Engineering (IJCSE) Vol..02, No.02, 2010, 250-255. Available: http://www.enggjournals.com/ijcse/doc /IJCSE10-02-02-25.pdf

[7] "Hospitalization for Heart Attack, Stroke, or Congestive Heart Failure among Persons with Diabetes", Special report: 2001 – 2003, New Mexico.

[8] World Health Organization. (July 2007-Febuary 2011). [Online]. Available:http://www.who.int/mediacentre/factsheets/fs3 10.pdf.

[9] Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02-5, Two Crows Corporation 10500 Falls Road, Potomac, MD 20854 (U.S.A.), 1999.

[10] Lacrosse, D.T., "Discovering Knowledge in Data: An Introduction to Data Mining", ISBN O-471-66657-2, on Wiley & Sons, Inc, 2005.

[11] Naïve bays classifier based on applying bays theorem: http://en.wikipedia.org/wiki/Naive bays classifier

[12] Weka Data mining software http://www.cs.waikato .ac.nz/ml/weka

[13] An Introduction to the WEKA Data mining system – http://www.cs.ccsu.edu/~markov/weka-tutorial.pdf

[14] Introduction to Support Vector Machine Available: http://en. wikipedia. Org/wiki/Support_vector_machine

[15] Colin Campbell and Liming Ying, Learning with Support Vector Machines, 2011, Morgan and Claypool. Available: http://www. morganclaypool.com/doi/abs/10.2200/S00324ED1V01Y201102 AIM010?journalCode=aim

[16] H.Barakat, Andrew P.Bradley and Mohammed Nabil H.Barakat (2009) "Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus", IEEE Transactions on Information Technology in Bio Medicine, Volume 14, Issue 4, pp 1-7, 2009. Available: http://ieeexplore. ieee.org /xpls/ abs_ all.jsp? Renumber=5378519 Digital Object Identifier: 10.1109 /TITB. 2009.2039485

ISSN: 2320-3714
Volume:2 Issue:2
May 2022
Impact Factor: 5.7
Subject Computer & Science

[17] Mertik M., Kokol P., Zalar B.Gaining Features in Medicine Using Various Data-Mining Techniques //Computational Cybernetics ICCC 2005, IEEE rdInternational Conference. – 2005. – P. 21–24.

[18] G.Suganya, D.Dhivya "Extracting Diagnostic rules from SVM" , Journal of Computer Applications (JCA), 2011.

[19] N. Barakat and A.P.Bradley, "Rule Extraction from Support Vector Machines: A Sequential Covering Approach " IEEE Transactions on Knowledge and Data Engineering, Volume 19,no.6,pp 729-741, 2007. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=04 161896 Digital Object Identifier no. 10.1109/TKDE.2007.1023.

[20] S.Balakrishnan, R.Narayanaswamy, N.Savarimuthu, R.Samikannu "SVM Ranking with Backward Search for Feature Selection in Type II Diabetes Databases" 2008 IEEE International Conference on Systems, Man and Cybernetics. Available: http://ieeexplore. ieee.org /xpls /abs_all.jsp? Arnumber=4811692 Digital Object Identifier: 10.1109 /IC SMC.2008.4811692

*****