A Rule-Based Monitoring System for Accurate Prediction of Diabetes
Monitoring System for Diabetes

Anand Kumar Srivastava, ABES Engineering College, India

Yugal Kumar, Jaypee University of Information Technology, India

Pradeep Kumar Singh, Jaypee University of Information Technology, India

https://orcid.org/0000-0002-7676-9014

ABSTRACT

Diabetes is a chronic disease that can affect the life of people due to high sugar level in their blood. The sugar level is increased due to a lack of production of insulin in the human body. Large numbers of people are affected with diabetes and it can increase tremendously due life style behavior. Diabetes can also affect the other human organs, like kidneys, hearts, retinas and lead to the failure of these organs. This article presents a diabetic monitoring system to determine the risk of diabetes based on the personal health record of patients. In this work, several rules are designed based on the clinical as well as non-clinical symptoms. The effectiveness of the diabetes monitoring system is tested on a set of two hundred forty people. The simulation results are also compared with well-known techniques available for diabetes prediction. It is stated that proposed monitoring system obtains 90.41% accuracy rate as compared with other techniques.

KEYWORDS

Diabetes, Machine Learning, Monitoring System, PHR

1. INTRODUCTION

Diabetes s is one of popular chronic disease that can affect the life of people throughout the world. The main reason of diabetes is lack of production of insulin in human body. It can also characterize as metabolic disease and having high blood sugar rate. The blood sugar level varies due to insulin emission. Further, it can also affect the other body parts of human and becomes worse, if cannot treat timely. The reduced level of insulin can also communicate blood glucose in blood platelets. In turn, the chance of several other diseases like heart disease, paralysis, strokes can also increase in diabetic people. In developed countries, diabetes become well-known reason of death and it one of most common non-communicable diseases throughout world (World Health Organization, n.d.). Till 2025, 300 million people will be either diabetic or pre-diabetic in world. A study reveals that fifty million people in India are either pre-diabetic or diabetic (Alice & Balachandran, 2015). Further, this study also highlights that nearly forty-four people in India are not know, they are diabetes affected. Diabetes can occur at any stage in human being. Clinically, diabetes can be categorized in type 1, type 2, and gestational diabetes (Zheng et al., 2007). Type 1 diabetes can be described as juvenile diabetes and it can occur in children frequently, but it also occurs in adults. In Type 1 diabetes, the body is unable

to produce insulin in appropriate amount and also destroy the cell which is responsible to produce insulin. Type 2 diabetes can occur any stage of human life, but it mainly occurs in fat, middle-aged, and elderly people. Gestational diabetes occurs in women due to hormonal and other changes. In the literature, numerous algorithms are presented for predicting diabetes. These algorithms include various conventional machine learning methods like decision tree, logistic regression, neural network, etc. (Kavakiotis et al., 2017; Yadav et al., 2012; Kumar & Sahoo, 2013). Some meta-heuristic algorithms and feature selection methods are also reported to determine the risk factors of diabetes disease (Yue et al., 2008; Duygu & Esin, 2011; Sahoo & Kumar, 2014; Gambhir et al., 2016).

It is also noticed that some ensemble methods are also developed for diabetes diagnosis (Kavakiotis et al., 2017; Ozcift & Gulten, 2011; Gambhir et al., 2017; Kumar et al., 2019; Gambhir et al., 2018). So, it is concluded that there is not a single system available in literature that can analyze the personal health record of patients. Hence, there is need of a diagnostic system that can effectively diagnose patients with diabetes. In this work, a diabetes monitoring system is proposed based on personal health record of patients. Further, several machine learning techniques are also used to compare the results of proposed system.

1.2 Statement of the Problem

The aim of this work is to develop a system based on personal health record of users for monitoring diabetes. Further, several rules are designed to accurate classification of diabetes disease. Moreover, a diabetes application is also developed to collect the data form users and this app is deployed in mobile environment.

1.3 Contribution of the Work

The aim of this work is to develop a diabetes monitoring system based on personal health record to determine the risk of diabetes. An online diabetes application is developed to keep record of the user information and also to collect the data. Through this app, several suggestions can also send to the registered user regarding diabetes prevention and also enables user to consult with doctor and medical staff. The aim of the monitoring system is to overcome the risk of diabetes and also to aware the people regarding diabetes. The contribution of this work is highlighted below.

1. A diabetic monitoring system is developed to monitor the health of users.
2. Several rules are designed for accurate classification of diabetes disease.
3. The developed decision system integrates with mobile environment.

Rest of paper is organized as. Section 2 highlights related work on diabetes prediction. Section 3 presents PHR based monitoring systems for diabetes. The implementation and simulation results of proposed monitoring system are discussed in section 4. The work is summarized in section 5.

2. RELATED WORKS

Nilashi et al. (2017) developed an intelligent system based on machine learning techniques to improve diabetes classification accuracy. The proposed system is the combination of clustering, noise removal and classification techniques. In this system, SOM is adopted for clustering, whereas, PCA technique is applied for noise removal. Further, classification task is performed using neural network. The Pima diabetes dataset is used to evaluate the performance of the proposed system. Authors claimed that the proposed system improves the classification accuracy.

Priyadarshini et al. (2017) presented a hybrid classifier to determine diabetic affected people. The proposed hybrid classifier is the combination of K-Means and GSA algorithm. In proposed classifier, GSA algorithm is applied to optimize the initial centroid for K-means algorithm. Moreover, K-means

algorithm is used to identify diabetic affected people correctly. It can be stated that proposed classifier obtains good classification results.

Cui et al. (2018) designed a decision support system for the management of patients with diabetes, especially for clinicians. In this model, a combination of support vector machine and genetic algorithm is applied to determine the risk of diabetes. The imbalance data is handled using SMOTE based method. The results of decision model are compared with other existing classifiers like naive bayes, DT, LR and NN. It is observed that the proposed decision model achieves 81.02% accuracy rate.

Aljumah et al. (2013) presented a study on diabetes affected people and divided in two group i.e. young age group and old age group. The objective of this work is to examine the effects of drug treatment especially for controlling diabetes. The oracle data miner tool is applied for predicting the treatment of diabetes. Further, SVM technique is used for analysis purpose. This study revealed that the drug treatment for young age group can delay side effects, but this study also revealed that the older age group requires immediate treatment.

Abdullah and Selvakumar (2013) developed PSO based decision tree algorithm to determine risk factor of Type-2 diabetes. Moreover, concept of self-adaptive inertia weight is also integrated into PSO algorithm for accelerating the search process. Simulation results of proposed algorithm are compared with several PSO and decision tree variants. It is stated that proposed algorithm effectively identifies the risk factors associated with diabetes.

Dalakleidi et al. (2017) adopted several machine learning approaches for estimating the risk of Type 2 diabetes. In their work, authors applied ensemble neural network to determine long term diabetes risks. Some popular machine learning approaches like logistic regression, Bayesian approaches and decision tree are also considered in this work. Simulation results stated that the ensemble neural network achieves 80.20% accuracy rate as compared to other approaches.

Meng et al. (2013) considered LR, ANN, and DT methods to predict pre-diabetes and diabetes affected patients. This study identified that seven hundred thirty-five people are either pre-diabetes or diabetes and seven hundred fifty two people are normal. Further, it is noticed that decision tree method obtains 77.52% accuracy rate in comparison to logistic regression and artificial neural network.

Yuvaraj and SriPreethaa (2017) implemented several machine learning algorithms in Hadoop environment for diabetes prediction. These machine learning algorithms are neural network, SVM, decision tree, Naïve Bayes and random forest. The performance of these algorithms is tested on Pima Indian diabetes dataset. Authors claimed that random forest algorithm obtains higher accuracy rate.

Pimentel et al. (2018) designed a screening method to determine the symptoms of Type -2 diabetes. In this work, a machine learning framework based on electronic health record is reported to determine the risk of Type-2 diabetes. Further, temporal features are also considered in this work. It is seen that the proposed screening method provides higher accuracy rate and accurate prediction of T2DM.

Wu et al. (2018) developed a prediction model based on improved K-mean and logistic regression for Type 2 diabetes prediction. The performance of prediction model is tested on Pima Indian diabetes dataset. Authors claimed that the proposed prediction model achieves 3.04% higher accuracy rate in comparison to existing model reported in literature.

Varma et al. (2014) applied a decision tree-based approach. Further, it is seen that to overcome crisp boundaries problem of decision tree, fuzzy based decision boundaries are integrated to achieve better accuracy rate. The effectiveness of the proposed approach is evaluated on Pima Indian Diabetes dataset. Author claimed that the proposed approach outperforms than others. In diabetes patients, diabetic retinopathy is one common causes for vision loss.

For accurate prediction of diabetic retinopathy, a clinical decision support system (CDSS) is reported in (Piri et al., 2017). The proposed system requires demographic and lab data for detecting susceptibility of retinopathy in patients rather than ophthalmologists and expensive equipment. It is noticed that CDSS provides higher accuracy rate.

A machine learning based framework is presented for detecting Type 2 diabetes (Zheng et al., 2017). This study explores genome-wide association to find relation between genotype-phenotype

associations. The framework includes k-NN, NB, DT, RF, SVM, and LR methods to detect T2DM. It is observed that proposed framework is more accurate and efficient for identifying T2DM.

Ruiz et al. (2017) presented a web based CDSS for the treatment of patients with gestational diabetes. The objective of CDSS is to improve specialized health assistance and prevent patients from unnecessary displacements. It is claimed that the proposed CDSS is capable to avoid gestational diabetes with adverse outcomes and provides remarkable results.

To provide an efficient framework for diabetes prediction, Dwivedi (2017) adopted six computational intelligence techniques. These techniques are DT, SVM, LR, NB, and ANN. The performances of these techniques are measured using different performance metrics. It is reported that ANN and LR give better results than other techniques.

Heart rate information based automated system to detect normal and diabetes patients is presented in (Acharya et al., 2015). The Electrocardiogram signals are used to determine heart rate information. Results stated that the proposed system obtained an average accuracy rate more than 92%.

Sappagh et al. (2015) developed a fuzzy ontology-based CBR framework for semantic diabetes diagnosis. The fuzzy ontology is applied on sixty real diabetic cases. It is claimed that the proposed system achieves an accuracy rate of 97.67%.

To generate the optimal rule set for diagnosis of diabetes, Cheruku et al. (2017) proposed a SM-RuleMiner for diabetes prediction. The Pima Indians Diabetes dataset is adopted to measure the performance of SM-RuleMiner. The proposed rule miner gives better performance in terms of rule length and rule set size.

A machine learning based approach for classification of healthcare data is reported in (Srivastava et al. 2019). In this work, RF, SVM and multi-layer NN are applied to accurate classification of healthcare data. It is claimed that SVM based system provides more accurate results.

A fog-based healthcare system developed for providing support to the remote patients (Devarajan et al., 2019). This system comprises of IOT and cloud computing techniques. The simulation results stated that proposed system provides better performance in terms of energy efficiency, accuracy and latency.

A mobile cloud computing-based decision system developed to monitor the progress of stroke patients (Karaca et al., 2019). The proposed system consists of two main components, i.e. mobile application and server application. ANN technique is used to classifying the data. The aim of this work is to develop user friendly stroke management system.

3. PROPOSED DIABETES MONITORING SYSTEM

This section presents diabetes monitoring system for determine the risk of diabetes in human being. To design the proposed monitoring system, a set of rule are designed for accurate prediction of diabetes. This rule set consists of twelve rules that are described in subsection 3.1. These rules are designed on the basis of several standards which are mentioned in Tables 1-5. Further, these rules are also certified through a panel of expert in the field of diabetes and this panel recommends these rules for diabetes identification. After that, a diabetes application is developed to collect the information of patients. This application is deployed in online environment and data is collected through online mode. This application is easily downloaded on user mobile phone and user can provide the details through this particular application. The aim of this application to connect with the user all time and also provides the feedback to users immediately. Some suggestions are provided to the user regarding diabetes prevention. Some of snapshots this application is illustrated in Figures 1-9. The major attributes of collected information is highlighted in Table 1. In this work, diabetes can divide in three categories- Pre-diabetes Type-1 diabetes and Type -2 diabetes. The pseudo code of diabetes is given in subsection 3.2. Table 2 illustrates cholesterol standards. Table 3 shows blood pressure standards. Table 4 shows standards for Type-1 diabetes. Table 5 displays standards for Type-2 diabetes.

Table 1. Standards for pre diabetic screening

Sr. No.	Standards	Range
1	BMI	> 25
2	BP	> 120/80
3	CL	High/Low
4	HD	Yes/No
5	Lifestyle behavior	Yes/No
6	Prenatal DB	Yes/No
7	Age	> 40

3.1 Rule Set for diabetes Identification

This subsection describes different rules designed for accurate prediction of diabetes disease. The rules are given as

Rule 1: If ((BMI>25) && (BP==Yes) && (CL==high) && (HD==Yes) && (Prenatal DB==Yes) && (lifestyle behavior==Yes) && (Age >40)); Pre-diabetic

Rule 2: If ((BMI>25) && (BP==Yes) && (CL==high) && (History HD==No) && (Prenatal DB==Yes) && (Lifestyle behavior ==Yes) && (Age<40)); Pre-diabetic

Rule 3: If ((BMI<25) && (BP==Yes) && (CL==high) && (History HD==No) && (Prenatal DB==Yes) && (Lifestyle behavior ==Yes) && (Age<40)); Pre-diabetic

Rule 4: If ((BMI<25) && (BP==Yes) && (CL== lower) && (HD==No) && (Prenatal DB==Yes) && (Lifestyle behavior ==Yes) && (Age<40)); Go for diabetic test

Rule 5: If ((BMI<25) && (BP==NO) && (CL== lower) && (HD==No) && (Prenatal DB==Yes) && (Lifestyle behavior ==Yes) && (Age<40)); Go for diabetic test

Rule 6: If ((BMI<25) && (BP==NO) && (CL== lower) && (HD==No) && (Prenatal DB==No) && (Lifestyle behavior ==No) && (Age>40)); Normal

Rule 7: If ((BMI<25) && (BP==NO) && (CL== lower) && (HD==No) && (Prenatal DB==No) && (Sedentary lifestyle==No) && (Age<40)); Normal

Rule 8: If ((LDL>100) && (HDL<40) && (Triglycerides>150)); High Cholesterol

Rule 9: If ((LDL>100) || (HDL<40) || (Triglycerides>150)); High Cholesterol

Rule 10: If ((LDL<100) && (HDL>40) && (Triglycerides<150)); Low Cholesterol

Rule 11: If ((Random test >100) && (Fasting test>126) && (Glycated hemoglobin>140)); Type-1 Diabetes

Rule 12: If ((Two-hour test >180) && (Oral glucose tolerance test >200)); Type-2 Diabetes

Figure 10 shows a Diabetes Monitoring System.

3.2 Pseudo Code of Proposed Decision System

This subsection summarizes the pseudo code of proposed decision system. Fig. 10 illustrates the proposed diabetes monitoring system. The pseudo code of proposed system is divided into three phases.

***Phase 1*:** This phase contains the personal health report of individual patients. The steps of this phase are listed below.
1. If user is already registered with decision system, then user click on login option, otherwise click on registration link.

**Figure 1. LOGIN details**
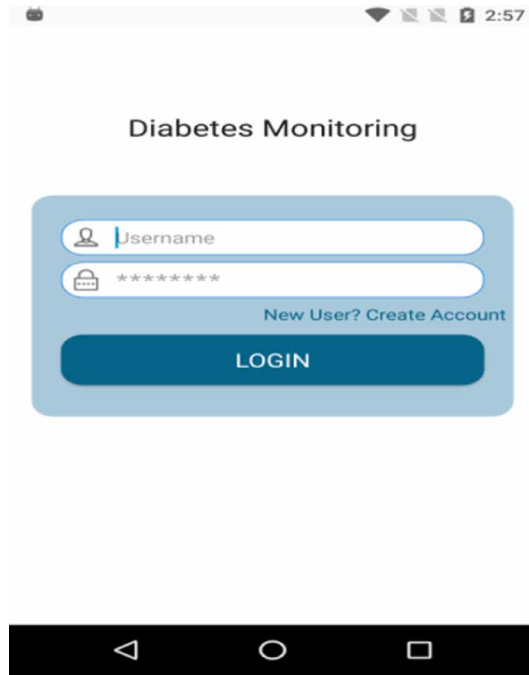
2. After successful login, User enter their personal health details like BMI, blood pressure, cholesterol, history of diabetes etc. to detect pre-diabetic or normal condition.
3. If user is pre-diabetic, then it is suggested for predetermined laboratory tests to identify Type-1 or Type-2 diabetes, otherwise, suggested for workout and control BMI, etc.
4. For Type-1 diabetes, some non-clinical parameters like thirst level, urination level, loss in weight, tiredness and vision bullring are described. If existence of these parameters is found, then go for clinical tests for confirmation of Type-1 diabetes.
5. If clinical tests, confirm Type-1 diabetes, then patients take recommended dose of insulin and also go for Type-2 diabetes.
6. In Type-2 diabetes, go for post prandial test and OGTT, if test positive, patients is affected with Type-2 diabetes and precaution can be taken.
7. If any user needs doctor consultation, then go for doctor consultation link.

Phase 2: This section deals with PHR data which is stored on server. Further, the stored data is classified using rules described in sub section 3.1. The main task of these rules is to divide data in two class normal and diabetes. The diabetes is further classified in pre diabetes, Type-1 diabetes and Type-2 diabetes. Some suggestions and recommendations are also given in case of presence of diabetes.

Phase 3: This section deals with the doctor opinion on clinical reports.

1. The information of patients is available for doctors through PHR system.
2. Doctors can see the information after login to the system.
3. Patient report online and prescription is sent to user through mail.

Figure 2. User registration
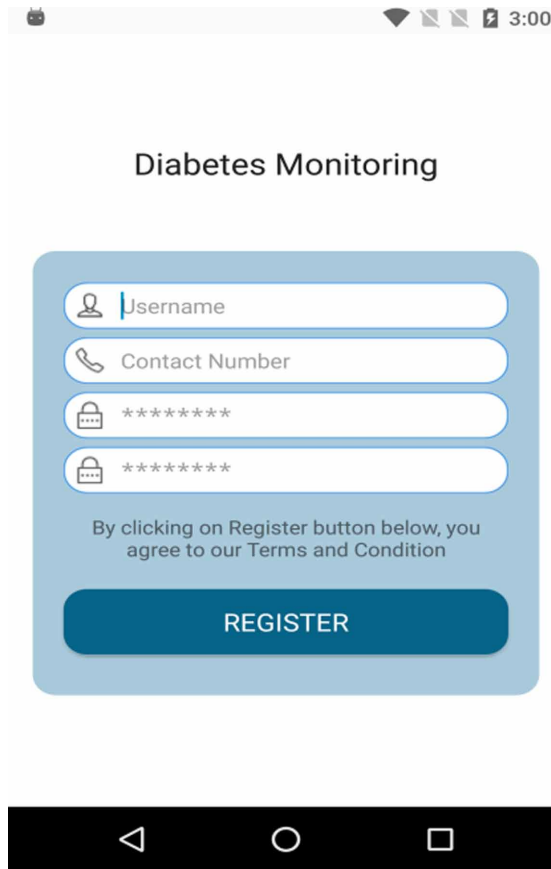
4. EXPERIMENTAL SECTION

To validate the effectiveness of proposed monitoring system, information of two hundred forty people were collected using a diabetic app. The data collection app is developed in android environment. This information comprises of several attributes. The information of these attributes is summarized in Table 1. Out of two hundred forty, one hundred fifty are diabetes patients and rest of are non-diabetic or normal. The simulation results of the proposed system are compared with artificial neural network (ANN), decision tree (DT), support vector machine (SVM), logistic regression (LR) and RBF-SVM. These are well defined and popular machine learning methods used in healthcare applications. In this work, accuracy, sensitivity, specificity and error rate are taken as performance parameters to assess the performance of machine learning methods. A confusion matrix is designed to derive aforementioned parameters. Table 6 illustrates the confusion matrix of ANN, DT and NB techniques. The confusion matrix of SVM, LR, and RBF-SVM is presented in Table 7. The confusion matrix of proposed monitoring system is mentioned in Table 8. It is observed that the proposed approach successfully detects diabetes-affected patients more efficiently than other techniques. It is also seen that decision tree-based approach exhibits worst results among all techniques. Table 9 consists of simulation results of proposed system and other classification techniques.

It is seen that proposed system obtains higher accuracy rate as compared to other techniques. Further, analysis of sensitivity and specificity parameters, it is stated that proposed system exhibits better performance than others. It is observed that DT approach exhibits worst performance using

Figure 3. Pre-diabetic interface

all parameters. The performance comparison of proposed system and other methods using accuracy, sensitivity, specificity and error rate parameters is shown in Figure 11. It can be concluded that proposed system outperforms than others.

Moreover, the effectiveness of the diabetes monitoring system is also evaluated on Pima Indian diabetes dataset. Table 10 summarizes the results of the proposed system and other classification techniques using pima Indian diabetes dataset. The results are evaluated in terms of accuracy, sensitivity, specificity and error rate. It is revealed that proposed system obtains more accurate results than other classification algorithms. Further, a rank parameter is also designed to obtain rank of each technique. It is seen that proposed system having first rank among all techniques, whereas, ANN techniques having last rank. It is also noticed that RBF-SVM obtains second rank among all techniques. Figure 12 illustrates the comparison of all techniques using accuracy, sensitivity, specificity and error rate. It is clearly stated that proposed system attains better results than compared algorithm and determines risk of diabetes in people more accurately.

5. CONCLUSION

This paper presents a diabetes monitoring system for prediction of diabetes affected people. The proposed system comprises of several rules for the identification of diabetes. Moreover, a diabetes app is developed to collect the patient's information and this app is deployed in online environment. This

Figure 4. Type- 1 symptoms non clinical
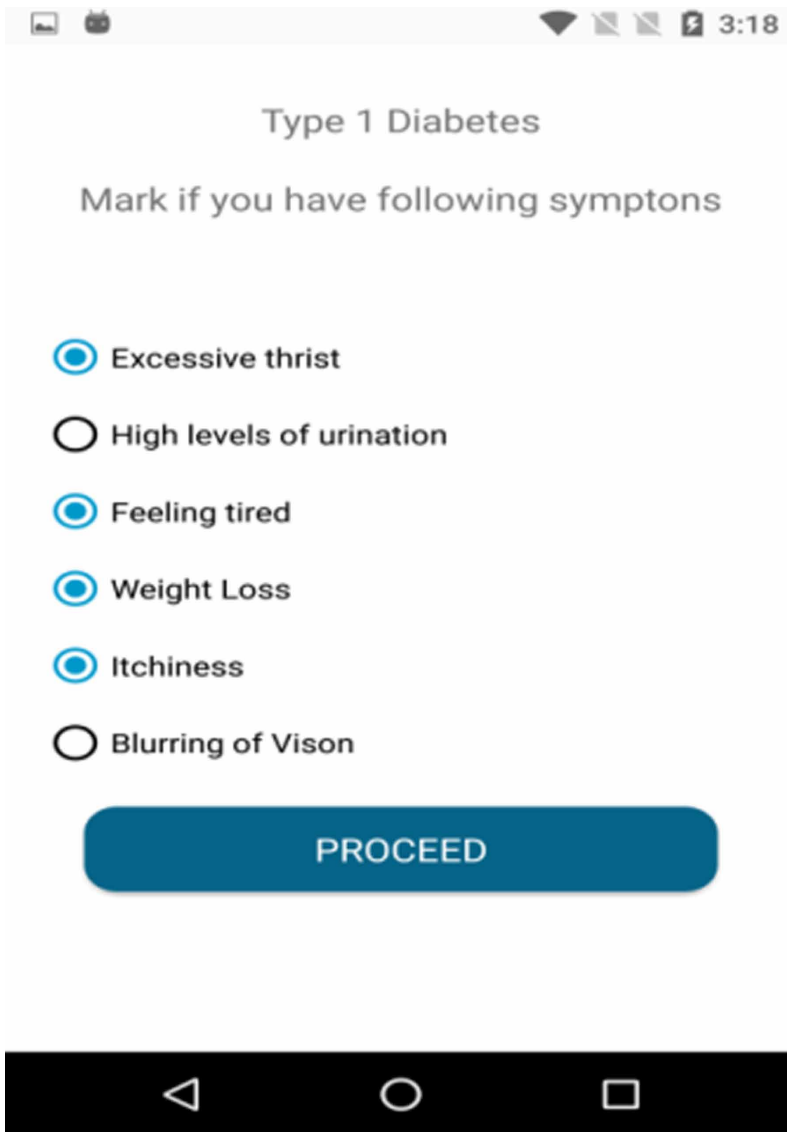
app also provides some suggestion regarding diabetes diseases. Through this app, information of two hundred forty patients is collected and this information is processed using the proposed monitoring system. The simulation results of proposed monitoring system are also compared with well-known approaches such as NN, LR, SVM, RBF-SVM and DT. The results are evaluated using specificity, accuracy, error control and sensitivity parameters. It is observed that proposed monitoring system 90.41 accuracy rates. The sensitivity rate of proposed system is 93.19, whereas, specificity rate of proposed system is 86.02. Moreover, it is also noticed that DT approach exhibits worst performance among all approaches. It is also stated that proposed system also obtains higher accuracy rate for Pima diabetes dataset as compared to other classification techniques. It can be concluded that proposed monitoring system is an effective solution for the identification of diabetes affected patients.
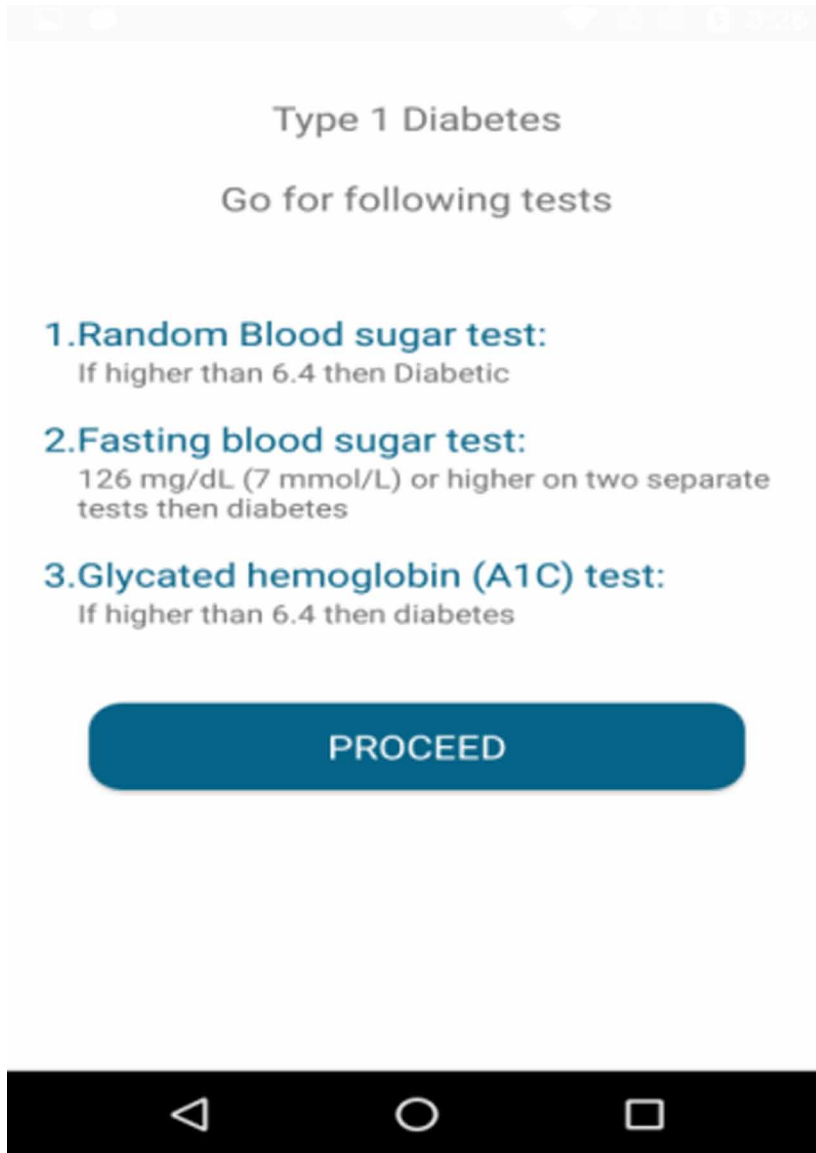
Figure 5. Type-1 clinical symptoms

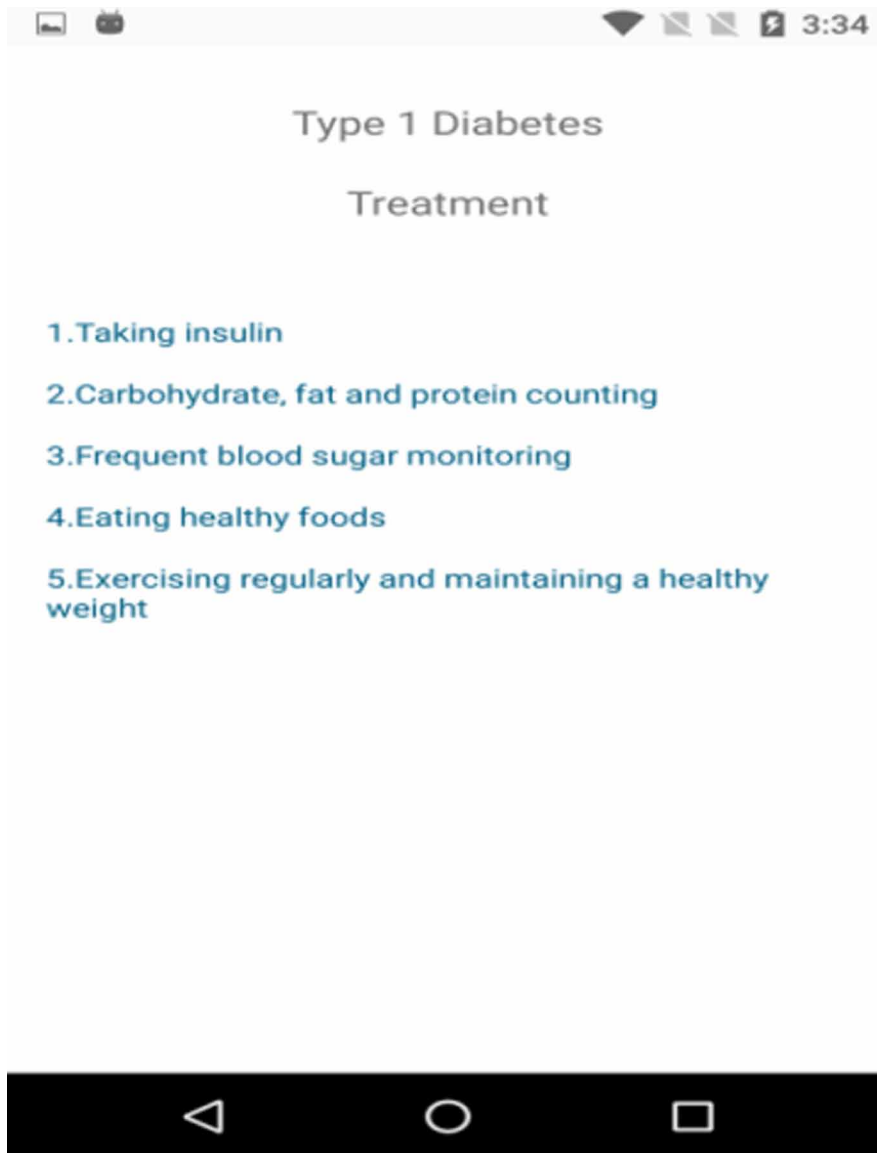Figure 6. Suggestions for Type- 1 diabetes

Figure 7. Type- 2 diabetes symptoms

Figure 8. Type- 2 diabetes clinical symptoms
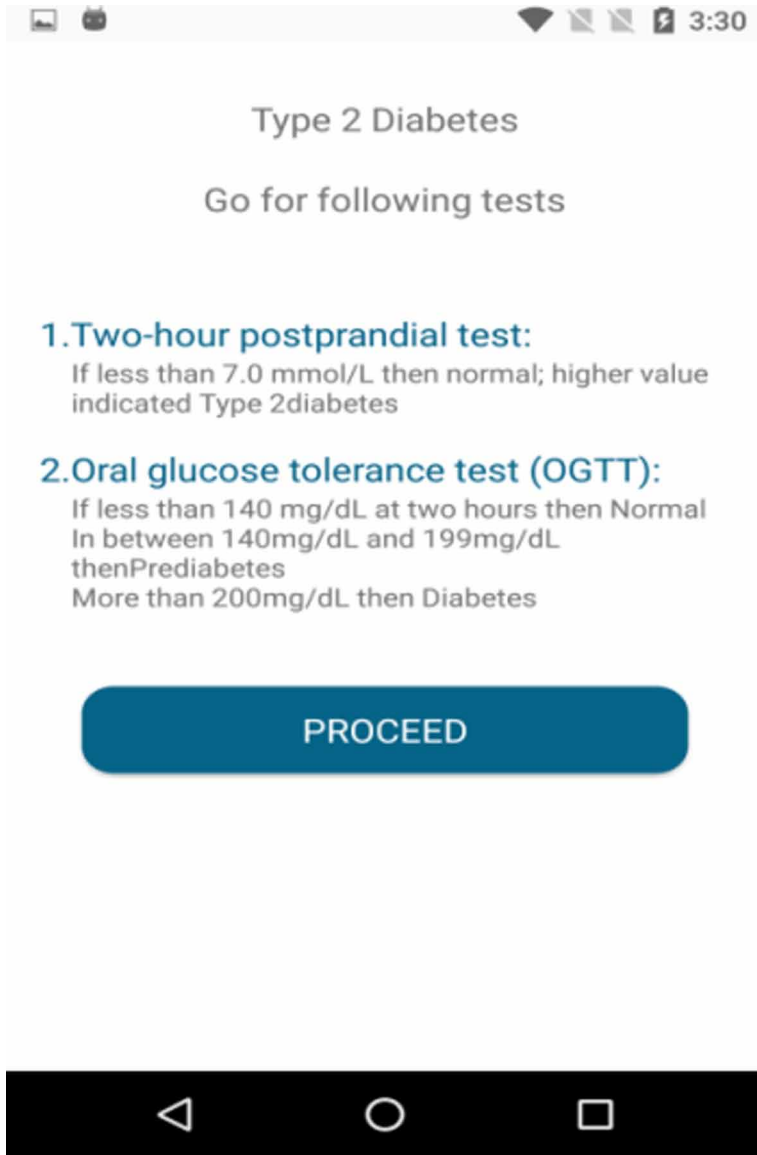
Figure 9. Suggestions for Type- 2 diabetes

Table 2. Illustrate cholesterol standards

Standard	Classification	Range
CL	LDL	> 100 mg/dl
	HDL	< 40 mg/dl
	Triglycerides	> 150 mg/dl

Table 3. Blood pressure standards

Standard	Classification	Range
BP	Healthy BP	120/80 mmHg
	Early High BP	In range of 120/80-140/90 mmHg
	High BP	> 140/90 mmHg

Table 4. Standards for Type-1 diabetes

Diabetes	Tests	Range
Type-1	Random test	> 100mg/dl
	Fasting test	> 126 mg/dl
	Glycated hemoglobin	> 140 mg/dl

Table 5. Standards for Type-2 diabetes

Diabetes Classification	Tests	Range
Type-2	Two-hour test	> 180 mg/dl
	Glucose test	> 200 mg/dl

Figure 10. Diabetes monitoring system
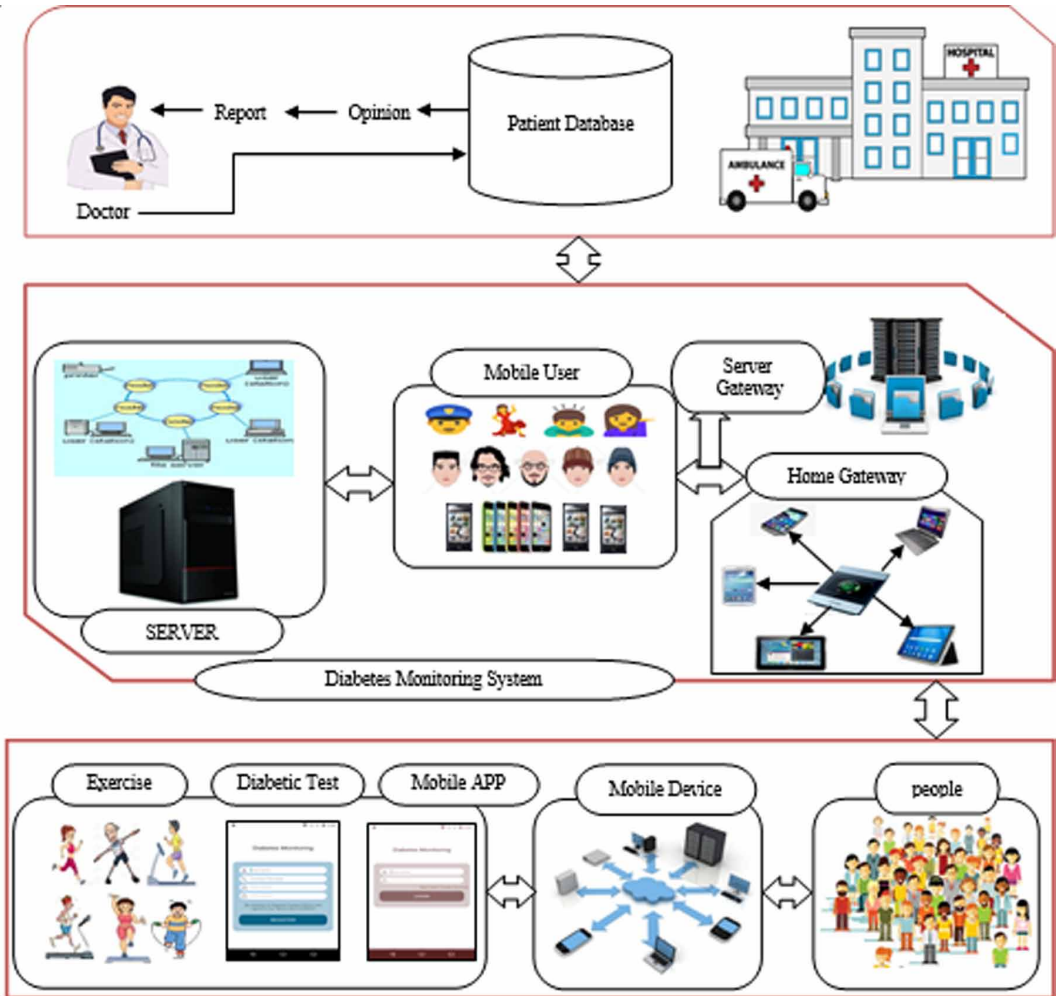
Table 6. Confusion matrix of ANN, DT and NB methods using tenfold cross validation method

Confusion Matrix		ANN		DT		NB	
		Predicted					
		P	N	P	N	P	N
Actual	P	116	34	113	37	112	28
	N	18	72	20	21	21	69

Table 7. Confusion matrix of SVM, LR and RBF-SVM methods using tenfold cross validation method

Confusion Matrix		SVM		LR		RBF-SVM	
		Predicted					
		P	N	P	N	P	N
Actual	P	124	26	125	25	137	21
	N	18	72	16	74	14	80

Table 8. Confusion matrix of proposed system using tenfold cross validation method

Confusion Matrix		Proposed System	
		Predicted	
		Positive	Negative
Actual	Positive	137	13
	Negative	10	80

Table 9. Performance comparison of proposed approach and other classification techniques

Parameters	Techniques						
	ANN	DT	NB	SVM	LR	RBF SVM	Proposed System
Accuracy	78.33	76.25	79.5	81.6	82.9	85.83	90.41
Rank	6	7	5	4	3	2	1
Sensitivity	86.56	84.96	84.21	87.32	88.65	90.72	93.19
Rank	5	7	3	6	4	2	1
Specificity	69.23	65.42	71.13	73.46	74.74	78.35	86.02
Rank	6	7	5	4	3	2	1
Error Rate	21.66	23.75	20.5	18.4	17.1	14.17	9.59
Rank	6	7	5	4	3	2	1
Overall Rank	5.75	7	4.5	4.5	3.25	2	1

Figure 11. Performance comparison of proposed system and other methods using different performance measures
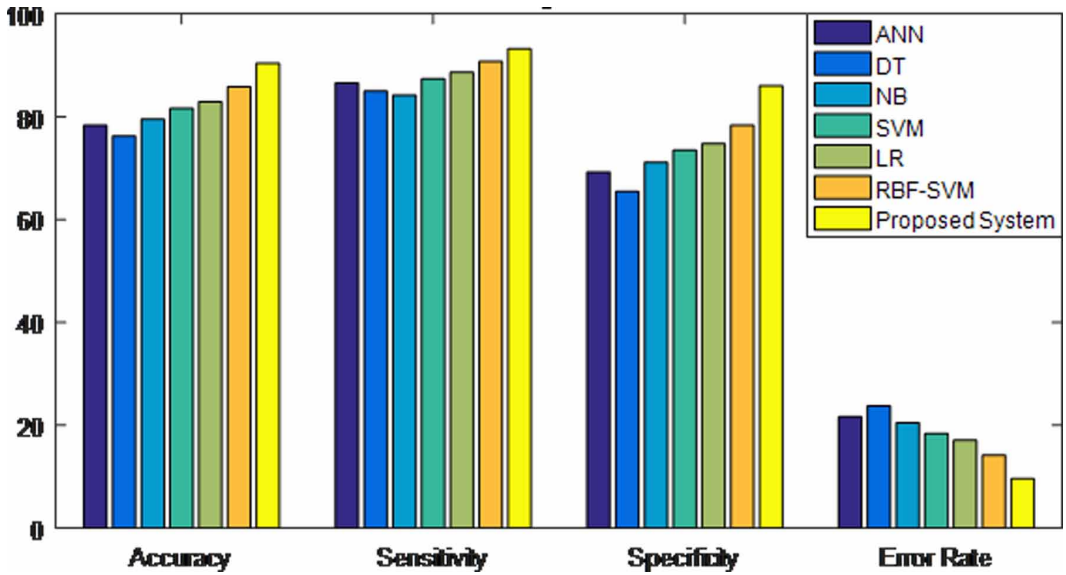

Table 10. Performance comparison of proposed approach and other classification techniques using Pima Indian diabetes dataset

Parameters	Techniques						
	ANN	**DT**	**NB**	**SVM**	**LR**	**RBF SVM**	**Proposed System**
Accuracy	65.51	62.45	68.19	72.06	62.83	76.56	85.23
Rank	5	7	4	3	6	2	1
Sensitivity	65.23	85.14	84.21	77.02	65.38	86.14	90.16
Rank	7	3	4	5	6	2	1
Specificity	64.48	65.56	71.13	74.56	72.84	81.64	87.34
Rank	7	6	5	3	4	2	1
Error Rate	34.59	37.55	31.81	27.94	37.17	23.44	14.77
Rank	5	7	4	3	6	2	1
Overall Rank	6	5.75	4.25	3.5	5.5	2	1

Figure 12. Performance comparison of proposed system and other methods using Pima Indian diabetes dataset
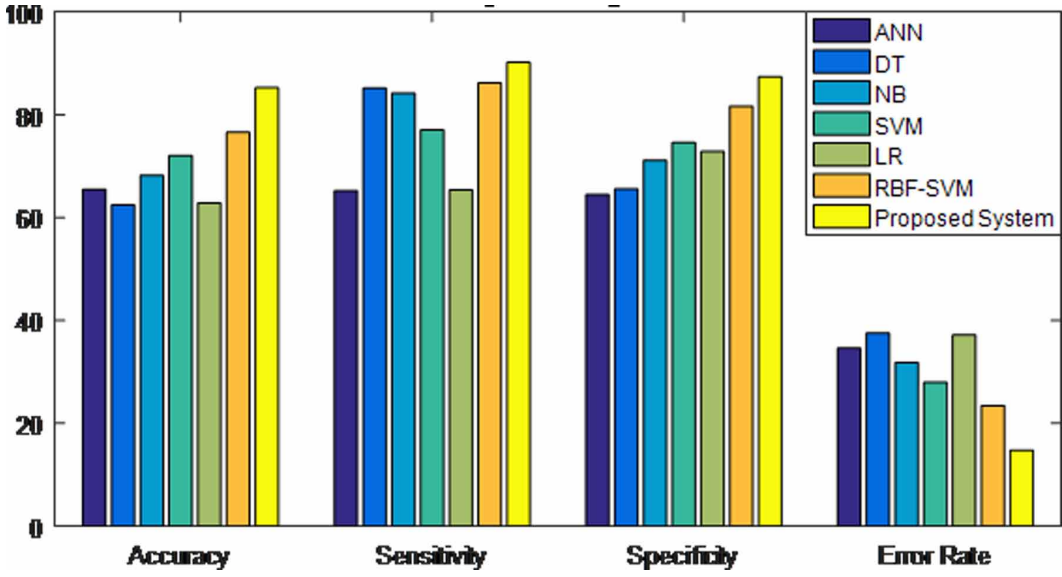
REFERENCES

Abdullah, A. S., & Selvakumar, S. (2018). Assessment of the risk factors for type II diabetes using an improved combination of particle swarm optimization and decision trees by evaluation with Fisher's linear discriminant analysis. *Soft Computing*, 1–23.

Acharya, U. R., Vidya, K. S., Ghista, D. N., Lim, W. J. E., Molinari, F., & Sankaranarayanan, M. (2015). Computer-aided diagnosis of diabetic subjects by heart rate variability signals using discrete wavelet transform method. *Knowledge-Based Systems*, *81*, 56–64. doi:10.1016/j.knosys.2015.02.005

Alice, S., & Balachandran, S. (2015), Performance Analysis of Training Algorithms of Multilayer Perceptrons in Diabetes Prediction. *Proceedings of the International Conference on Advances in Computer Engineering and Applications (ICACEA)*, Ghaziabad, India. Academic Press.

Aljumah, A. A., Ahamad, M. G., & Siddiqui, M. K. (2013). Application of data mining: Diabetes health care in young and old patients. *Journal of King Saud University-Computer and Information Sciences*, *25*(2), 127–136. doi:10.1016/j.jksuci.2012.10.003

Caballero-Ruiz, E., García-Sáez, G., Rigla, M., Villaplana, M., Pons, B., & Hernando, M. E. (2017). A web-based clinical decision support system for gestational diabetes: Automatic diet prescription and detection of insulin needs. *International Journal of Medical Informatics*, *102*, 35–49. doi:10.1016/j.ijmedinf.2017.02.014 PMID:28495347

Cheruku, R., Edla, D. R., & Kuppili, V. (2017). SM-RuleMiner: Spider monkey based rule miner using novel fitness function for diabetes classification. *Computers in Biology and Medicine*, *81*, 79–92. doi:10.1016/j.compbiomed.2016.12.009 PMID:28027460

Cui, S., Wang, D., Wang, Y., Yu, P. W., & Jin, Y. (2018). An improved support vector machine-based diabetic readmission prediction. *Computer Methods and Programs in Biomedicine*, *166*, 123–135. doi:10.1016/j.cmpb.2018.10.012 PMID:30415712

Devarajan, M., Subramaniyaswamy, V., Vijayakumar, V., & Ravi, L. (2019). Fog-assisted personalized healthcare-support system for remote patients with diabetes. *Journal of Ambient Intelligence and Humanized Computing*, 1–14.

Duygu, Ç., & Esin, D. (2011). An automatic diabetes diagnosis system based on LDA-wavelet support vector machine classifier. *Expert Syst. Appl.*, *38*, 8311–8315.

Dwivedi, A. K. (2017). Analysis of computational intelligence techniques for diabetes mellitus prediction. *Neural Computing & Applications*, 1–9.

El-Sappagh, S., Elmogy, M., & Riad, A. M. (2015). A fuzzy-ontology-oriented case-based reasoning framework for semantic diabetes diagnosis. *Artificial Intelligence in Medicine*, *65*(3), 179–208. doi:10.1016/j.artmed.2015.08.003 PMID:26303105

Gambhir, S., Malik, S. K., & Kumar, Y. (2016). Role of soft computing approaches in healthcare domain: A mini review. *Journal of Medical Systems*, *40*(12), 287. doi:10.1007/s10916-016-0651-x PMID:27796841

Gambhir, S., Malik, S. K., & Kumar, Y. (2017). PSO-ANN based diagnostic model for the early detection of dengue disease. *New Horizons in Translational Medicine*, *4*(1-4), 1–8. doi:10.1016/j.nhtm.2017.10.001

Gambhir, S., Malik, S. K., & Kumar, Y. (2018). The diagnosis of dengue disease: An evaluation of three machine learning approaches. *International Journal of Healthcare Information Systems and Informatics*, *13*(3), 1–19. doi:10.4018/IJHISI.2018070101

35. Karaca, Y., Moonis, M., Zhang, Y. D., & Gezgez, C. (2019). Mobile cloud computing based stroke healthcare system. *International Journal of Information Management*, *45*, 250–261. doi:10.1016/j.ijinfomgt.2018.09.012

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal*, *15*, 104–116. doi:10.1016/j.csbj.2016.12.005 PMID:28138367

Kumar, Y., & Sahoo, G. (2013). Prediction of different types of liver diseases using rule based classification model. *Technology and Health Care*, *21*(5), 417–432. PMID:23963359

Kumar, Y., Yadav, G., Singh, P. K., & Arora, P. (2019). A PHR-Based System for Monitoring Diabetes in Mobile Environment. In *Mobile Solutions and Their Usefulness in Everyday Life* (pp. 129–144). Cham: Springer. doi:10.1007/978-3-319-93491-4_7

Meng, X. H., Huang, Y. X., Rao, D. P., Zhang, Q., & Liu, Q. (2013). Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. *The Kaohsiung Journal of Medical Sciences*, *29*(2), 93–99. doi:10.1016/j.kjms.2012.08.016 PMID:23347811

Nilashi, M., Ibrahim, O., Dalvi, M., Ahmadi, H., & Shahmoradi, L. (2017). Accuracy improvement for diabetes disease classification: A case on a public medical dataset. *Fuzzy Information and Engineering*, *9*(3), 345–357. doi:10.1016/j.fiae.2017.09.006

Ozcift, A., & Gulten, A. (2011). Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. *Computer Methods and Programs in Biomedicine*, *104*(3), 443–451. doi:10.1016/j.cmpb.2011.03.018 PMID:21531475

Pimentel, A., Carreiro, A. V., Ribeiro, R. T., & Gamboa, H. (2018). Screening diabetes mellitus 2 based on electronic health records using temporal features. *Health Informatics Journal*, *24*(2), 194–205. doi:10.1177/1460458216663023 PMID:27566751

Piri, S., Delen, D., Liu, T., & Zolbanin, H. M. (2017). A data analytics approach to building a clinical decision support system for diabetic retinopathy: Developing and deploying a model ensemble. *Decision Support Systems*, *101*, 12–27. doi:10.1016/j.dss.2017.05.012

Priyadarshini, R., Barik, R. K., Dash, N., Mishra, B. K., & Misra, R. (2017). A hybrid GSA-K-mean classifier algorithm to predict diabetes mellitus. *International Journal of Applied Metaheuristic Computing*, *8*(4), 99–112. doi:10.4018/IJAMC.2017100106

Sahoo, A.J., & Kumar, Y. (2014). Seminal quality prediction using data mining methods. *Technology and Health Care*, *22*(4), 531–545. PMID:24898862

Srivastava, S. K., Singh, S. K., & Suri, J. S. (2019). Effect of incremental feature enrichment on healthcare text classification system: A machine learning paradigm. *Computer Methods and Programs in Biomedicine*, *172*, 35–51. doi:10.1016/j.cmpb.2019.01.011 PMID:30902126

Varma, K. V., Rao, A. A., Lakshmi, T. S. M., & Rao, P. N. (2014). A computational intelligence approach for a better diagnosis of diabetic patients. *Computers & Electrical Engineering*, *40*(5), 1758–1765. doi:10.1016/j.compeleceng.2013.07.003

World Health Organization. (n.d.). Retrieved from http://www.who.int

Wu, H., Yang, S., Huang, Z., He, J., & Wang, X. (2018). Type 2 diabetes mellitus prediction model based on data mining. *Informatics in Medicine Unlocked*, *10*, 100–107. doi:10.1016/j.imu.2017.12.006

Yadav, G., Kumar, Y., & Sahoo, G. (2012). Predication of Parkinson's disease using data mining methods: A comparative analysis of tree, statistical and support vector machine classifiers. *Proceedings of the 2012 National Conference on Computing and Communication Systems (NCCCS)* (pp. 1-8). IEEE. doi:10.1109/NCCCS.2012.6413034

Yue, C., Xin, L., Kewen, X., & Chang, S. (2008). An intelligent diagnosis to type 2 diabetes based on QPSO algorithm and WLS-SVM. *Proceedings of the 2008 IEEE International Symposium on Intelligent Information Technology Application Workshops*, Washington, DC. IEEE Press. doi:10.1109/IITA.Workshops.2008.36

Yuvaraj, N., & SriPreethaa, K. R. (2017). Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster. *Cluster Computing*, 1–9.

Zheng, J. W., Zhang, Z. B., Wu, T. H., & Zhang, Y. (2007). A wearable mobihealth care system supporting real-time diagnosis and alarm. *Medical & Biological Engineering & Computing*, *45*(9), 877–885. doi:10.1007/s11517-007-0221-y PMID:17619091

Zheng, T., Xie, W., Xu, L., He, X., Zhang, Y., You, M., & Chen, Y. et al. (2017). A machine learning-based framework to identify type 2 diabetes through electronic health records. *International Journal of Medical Informatics*, *97*, 120–127. doi:10.1016/j.ijmedinf.2016.09.014 PMID:27919371

APPENDIX 1.

Table 11. List of Abbreviations

PHR	Personal health Record
SOM	Self-Organizing Maps
PCA	Principle Component Analysis
GSA	Gravitational Search algorithm
DT	Decision Tree
LR	Logistic Regression
NN	Neural Network
SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support Vector Machine
PSO	Particle Swarm Optimization
ANN	Artificial Neural Network
T2DM	Type 2 Diabetes Mellitus
NB	Naive Bayes
CDSS	clinical decision support system
k-NN	k-Nearest Neighbour
RF	Random Forest
CBR	Case Based Reasoning
BMI	Body Mass Index
BP	Blood Pressure
CL	Cholesterol Level
HD	Heart Disease
DB	Diabetes
LDL	Low Density Lipoproteins
HDL	High Density Lipoproteins
OGTT	Oral Glucose Tolerance Test
RBF	Radial Basis Function

Yugal Kumar is currently working as Assistant Professor (Senior Grade) in the Department of Computer Science & Engineering at Jaypee University of Information Technology (JUIT) Waknaghat, Solan, India. He has presented and published over 50 research papers in reputed journals and various national and international conferences. His important research contributions include developing data clustering algorithms, decision models for disease prediction, and methods for precision agriculture. His research interests include nature inspired algorithms, optimization problems like classification, clustering and function optimization, medical data mining, and precision agriculture.