IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

A Novel Rule Pruning Ensemble Learning Approach to Recognize the Risk of Impaired Glucose Tolerance

K. Suganya[1], Dr. L. Sankari[2]

Research Scholar, Department of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, India[1]

Associate Professor, Department of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, India[2]

Abstract: Objective: Prediction of diabetes in patients by reducing the size of rule set. Accurate prediction of diabetes risk level Methods: Early detection of diabetes can help the patients by providing information about treatment and clinical suggestions to prevent the particular individual from dangerous condition. Machine learning techniques are thus developed for accurate diagnosis of diabetes. Support vector machine (SVM) with ensemble learning approach is used for rule extraction. A novel rule pruning ensemble learning approach using frequent patterns is implemented to reduce the rule sets and improve the diagnostic performance by recognizing the risk of impaired glucose tolerance. Finding: Diabetes is a chronic condition causes high blood sugar levels. The diabetic patients are classified into type 1 and type 2diabetes. Diabetes mellitus of type 2 is considered to be the most critical worldwide public health problems that increase the level of sugar in the blood. Application/improvements: this approach accuracy, precession, recall, F-measure of the prediction of diabetes is increased.

Keywords: Diabetes mellitus, Ensemble Learning, Rule pruning, Impaired glucose tolerance.

1. INTRODUCTION

Data mining is the process of mining the patterns form data. Generally, data mining is the search for hidden patterns that could be present in huge databases. Data mining scans via a huge volume of data to find out the patterns and correlations between patterns. Data mining consists of more than gathering and running data also contains analysis and prediction. Data mining application could use several parameters to inspect the data. Diabetes, a chronic condition that is incorporated with severe consequences affects the economic and social life of the particular patient. This diabetes is categorized based on blood glucose levels and divided into type 1 and type 2 diabetes. Impaired glucose tolerance means that blood glucose is raised beyond normal levels, but not high enough to warrant a diagnosis. With impaired glucose tolerance you face a much greater risk of developing diabetes and cardiovascular disease.

In this paper support vector machine (SVM) with ensemble learning approach is used for rule extraction. RF (Random Forest) rule induction technique is used to develop assessment rules to diagnose the diabetes. Initially, SVM model is constructed by training data. The trained SVM model provides class label. Then Support vectors (SV) are predicted by the trained SVM model and original labels of SV are replaced by predicted labels of SV. The purpose of changing labels is to remove some noise from the data. Then the artificial data is applied to RF algorithm and best rules sets are obtained by adjusting the rule induction method's parameters. Finally, the rules obtained are estimated based on the test data. The dataset used in this work used 90% for rule extraction, and last 10% of data is used as test set. Here for data preparation, tenfold cross validation (CV) is used as the training method to generate optimal parameters and incorporates tenfold results to estimate averaged accuracy of tenfold CV. Then based on best precision and recall rate, best fold is considered as the chosen set to generate rule sets and estimated by the test data. However Extraction of more rules from SVM model, risk of impaired glucose tolerance is not considered and additional techniques are required to monitor the glucose level. To overcome this problem we proposed a novel rule pruning ensemble learning approach.

2. LITERATURE SURVEY

L.C. Hay et al (2003) [1] diagnosed the elder type 2 diabetes patients using Continuous Glucose Monitor System (CGMS). This system allows intensive interstitial glucose monitoring for determining prevalence of unrecognized hypoglycemia and hyperglycemia in elderly (>65 years old) patients with type 2 diabetes. Here a sensor is placed in

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

subcutaneous tissue of the abdominal wall which samples the interstitial fluid and then electrons are produced based on concentration of glucose in the fluid. This current is detected by the monitor, and the corresponding glucose value is recorded. CGMS provides good tolerability and accurate results on glycemic profile in elderly patients with type 2 diabetes. However, obtained results are only for short duration study. This paper used the method for Continuous Glucose Monitor System (CGMS) is used to monitor the continuous glucose profile of elder patients.

Arvind Gupta et al [2] proposed Prevalence of diabetes, impaired fasting glucose and insulin resistance syndrome in an urban Indian population. In this paper, Epidemiological study among urban subjects in western India is used to determine prevalence of diabetes, insulin resistance syndrome (IRS) and their risk factors. The study was designed to investigate people at random and to cover large and varied areas of Jaipur with a view to include persons from all walks of urban life. The study thus shows that the simple clinical and biochemical measurements can predict the presence of IRS and these subjects have risk factors similar to those with diabetes. This paper used the method for Epidemiological Study. Jian-Jun Wang et al (2004) [3] showed the Effects of impaired fasting glucose and impaired glucose tolerance on predicting incident type 2 diabetes. The main purpose of this effect is to which impaired glucose homeostasis at baseline is predictive of conversion to type 2 diabetes. First, capillary glucose pre-screening is done for OGTT at baseline. It combined the IFG and IGT at baseline to strongly predict the risk of diabetes. However, it does not provide in baseline characteristics was found between the participants and non-participants in the follow-up study. This paper used the method for impaired fasting glucose and impaired glucose tolerance records are used for predicting type 2 diabetes.

Chao-Ton Su et al (2006) [4] constructed a prediction model to predict Type II diabetes using anthropometrical body surface scanning data. This model includes back propagation neural network, decision tree, logistic regression, and rough set to select the relevant features from data. Back propagation neural network is useful for selecting features. Decision tree such as CART and C4.5 is used to reduce misclassification. Logistic regression is a binary data analysis which improves accuracy of classification. Rough set theory provides information about vagueness and uncertainty and gave better results on data deduction. Finally, the result shows that classification accuracy of decision tree and rough set is better than logistic regression and back propagation neural network. This paper used the method for back propagation neural network, decision tree, logistic regression are used to select the relevant features from data to predict the diabetes.

H.J. Yoo et al (2008) [5] used a real time continuous glucose monitoring system (RT-CGM) in patients with type 2 diabetes. This system controls the glucose level in blood by modifying patient's diet and exercise habits after applying RT-CGM. Here the patients are randomly allocated to the Guardian RT and SMBG groups. Sensors are placed and thresholds are set for both hyperglycemic and hypoglycemic Guardian RT group underwent real time continuous glucose monitoring once a month for 3 days Sensors. SMBG group was instructed to check their blood glucose level at least four times a week. This paper used the method for real time continuous glucose monitoring system (RT-CGM) is used patient's diet and exercise habits and control glucose level in diabetes patients.

Masoud Amini& Mohsen Janghorbani (2009) [6] presented the Comparison of metabolic syndrome with glucose measurement for prediction of type 2 diabetes. In this comparison, the ability of the metabolic syndrome (MetS) and fasting and 2-h glucose to predict progression to diabetes in non-diabetic first-degree relatives (FDRs) of patients with type 2 diabetes is compared. By analysis, FPG measurement is a much better predictor of progression to diabetes than MetS. This paper used the method for OGTT is used to observe metabolic syndrome.

Mehdi Rambod et al (2009) [7] proposed a cross-sectional study to predict the isolated impaired glucose tolerance. This study estimated the result of oral glucose tolerance test (OGTT) to determine the diagnostic value of isolated impaired glucose tolerance (isolated-IGT). For this purpose, author used the characteristics in multivariate analyses and then used a quantitative approach to calculate positive and negative likelihood ratios (LR) for each component and different constellations of them to facilitate opportunistic case finding in daily clinical practice. However, it is cost-effective. This paper used the method for oral glucose tolerance test (OGTT) to determine the diagnostic value in clinical prediction model. TsvetalinaTankova et al (2011) [8] evaluated the performance of Finnish Diabetes Risk Score (FINDRISC) on screening the impaired fasting glucose (IFG) and impaired glucose tolerance (IGT) and undetected diabetes (UDD) to find the risk level of diabetes. This test uses age, body mass index, family history of diabetes, waist circumference, use of anti-hypertensive medication, history of elevated blood glucose, meeting the criterion for daily physical activity and daily consumption of fruit and vegetables and identifies the risk of developing diabetes. It is a simple, safe and practical way to detect pre-diabetes and individuals at high risk for diabetes. However, it is costly and time consuming tool. This paper used the method used the Finnish Diabetes Risk Score (FINDRISC) on screening the impaired fasting glucose and impaired glucose tolerance to find the risk level of diabetes.

C. Bianchi et al (2011) [9] evaluated the high risk of type 2 diabetes using metabolic syndrome (MS). The author used the data from genetics, pathophysiology and evolution of type 2 diabetes (GENFIEV) for finding the prevalence of MS, identifies role of Insulin Resistance and insulin secretion and cardiovascular risk profile associated with MS. The result shows that, MS prevalence is better in impaired glucose tolerance (IGT) and impaired fasting glucose (IFG). This paper used the method for OGTT is used for the estimation of plasma levels of glucose.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

Laura N. McEwen et al (2013) [10] used available health plan data to screen the impaired fasting glucose and type 2 diabetes. Here, the author uses demographic, claims, pharmacy data, laboratory data and clinical data. And also uses Equation development and Equation validation sets for screening process. This provides high specificity and improves detection of at-risk populations. However, it shows low sensitivity and positive predictive value (PPV). This paper used the method for The diabetes prediction is done by using Wald chi-square test which provides predictive equation.

3. MATERIALS METHODS

In this paper uses a novel rule pruning ensemble learning approach to reduce the large rule sets by removing rules that are helpful for diagnosing procedure. This approach uses pruning algorithm with SVM and RF ensemble learning method to generate small and essential rule set and diagnoses the diabetes patients.
In addition to this approach, to determine the status of diabetes risk in patients, amount of blood glucose level under various conditions should be analyzed from Diabetes patient records.

3.1 rule extraction from SVM

In this module, the unbalanced dataset is handled and data is used for training SVMs with RBF kernel. SVM is based on the principle of structural risk minimization and it belongs to the supervised learning models for nonlinear classification analysis. The SVM model is achieved by finding the optimal separating hyper plane (w.x+b=0) with maximizing the margin d, which is defined as d=2/||w||. This optimal hyper plane can be represented as a convex optimization problem:

$$\text{Minimize } \tfrac{1}{2} \|w\|^2 \text{ subject to } y_i (wx_i + b) \geq 1 \quad (1)$$

$$\text{Minimize } \tfrac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} \varepsilon_i \quad (2)$$

In the nonlinear classification problem, the SVM uses kernel functions to map the examples into the high-dimensional feature space and separates categories by a clear linear margin. Usually, radial basis function (RBF) is used as the kernel function to map the data

$$K(x, x') = \exp\left(\frac{||x-x'||^2}{2\sigma^2}\right) \quad (3)$$

Where $||x - x'||^2$ the squared Euclidean distance between two is vectors and σ^2 is a free parameter. The kennel function becomes

$$\text{maximize } w(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1,j=1}^{n} \alpha_i y_i \alpha_j y_j K(x_i, x_j)$$

$$\text{subject to } C \geq \alpha_i \geq 0 \quad \forall i, \sum_{i=1}^{n} \alpha_i y_i = 0. \quad (4)$$

Hence, solving for α by the gradient decent algorithm, the SVs can be obtained by the examples of training data which have nonzero Lagrange multiplier. The hyper plane is completely defined by SVs as

$$f(x) = \text{sign}\left(\sum_{s=1}^{sv} \alpha_s y_s K(x_s, x) + b\right).$$

SVs are the only examples that make contribution to the classification of the SVM.
Then, the SVM model in the CV was constructed by the best fold, which was defined as the fold gave the best classification rate with the particular fold's test set, and finally the SVM model was used to test on the remained 10% dataset.This is a particularly pressing problem for small test datasets. In addition, if the approaches were applied to the datasets on which rule induction techniques perform better than SVM, the rule extraction from SVM would seem illogical.The average accuracy of these models was calculated with precision, recall, F score, and AU

3.2 Rule extraction by novel pruning method

This module uses a novel rule pruning ensemble learning approach to reduce the large rule sets by removing rules that are not helpful for diagnosing procedure. This approach uses pruning algorithm with SVM and RF ensemble learning method to generate small and essential rule set and diagnoses the diabetes patients.

3.3 Ensemble Pruning based on Frequent Patterns (EP-FP)

In EP-FP, classification results obtained from SVM and RF ensemble are used for constructing the special Boolean matrix called classification matrix. Candidate ensembles are then built by iteratively extracting the frequent base

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

classifiers. Based on the major voting, each instance is categorized by an ensemble $\{e_1, e_2, \dots, e_m\}$ and the subset with the size less than that of m/2. The size of final ensemble should satisfy the condition $|F_S| = \{F_s | 0 < F_s \leq m\}$. Then candidate ensembles are searched for every iteration by EP-FP. After revealing the majority votes, the generated ensemble of candidate consists of $F_S/2$ base classifiers and its size should satisfy, $z[F_S/2] \leq z \leq F_S$, where z represents the candidate size. Instance that is classified less than $F_S/2$, is considered as a useless instance for searching and evaluating candidate ensemble.

The process of eliminating the useless instance is known as refining classification matrix. After refining process, mining of frequent base classifiers is done by taking instances of matrix.The performance of candidate's ensemble can be evaluated using number of right-classified instances, prediction accuracy of base classifiers and candidate size. Final ensemble with optimal evaluation is given by,

$$E_{Feval}(C) = \frac{supp(C) \times \sum_{i \in C} Acc(B_{ci})}{\left|\frac{C}{B_c}\right|} \qquad (5)$$

Where $supp(C)$ represents the support of base classifier in candidate C and it is given by,

$$supp(C) = \frac{|R(C)|}{|I|} \qquad (6)$$

C represents the base classifiers in candidate C
$|B_c|$ represents the base classifiers in pruned ensemble
|I| represents the number of instances in dataset.
$Acc(B_{ci})$Represents the prediction accuracy of base classifierB_{ci}, and it is denoted by,

$$Acc(B_{ci}) = stre(B_{ci}) \times supp(B_{ci}) \quad (7)$$

Where $stre(B_{ci})$ represents the level of classification of base classifier and it is represented by,

$$stre(B_{ci}) \frac{\sum_{i=1}^{|d|}(hard(I_i) \times \sigma)}{|I|} \qquad (8)$$

Algorithm1: Main Algorithm
Step1: Read the diabetes dataset
Step2: Pre-preprocessing of the diabetes dataset.
Step3: Reduce the size of dataset
Step4: Select and Extract the features from
SVM(Support Vector Machine)
Step5: Assign two different hyper planes
Step6: Compute kernel function by using preprocessing
Step7: Calculate the distance between the hyper planes
Step8: Apply Algorithm 2 for defining the rules sets
Step9: Apply algorithm 3 for pruning the rules
Step10: Stop

Algorithm2: Random Forest
Step 1: Consider N is the number of training cases and M is the number of variables in classifiers.
Step2: The number of input attributes m used for determining decision at node of tree and$m < M$.
Step3: Select training set for this tree by choosing N times with replacement from all N available training cases.
Step4: Apply the remaining training cases to compute the error of the tree by means of predicting their classes.
Step5: For each node in tree, select m variables randomly.
Step6: Compute the best split according to the m variables in training set.
Step7: Each tree is completely grown and not pruned.

Algorithm 3: Rule Pruning
Step1: Support of base classifiers is calculated by $Supp(X) = \frac{|R(X)|}{|D|}$ //Where |D| is the number of instances in dataset D.
Step2: Construction of classification matrix and Row array
Step3: Get the refined classification matrix
Step4: Get the candidate ensemble
Step5: Get the evaluation of the candidate
Step6: Determine the final ensemble with the optimal evaluation

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

3.4 Rule generation and evaluation

The RF is an ensemble learning method for classification. RF constructs a multitude of decision trees and utilizes the mode of individual trees' output to classify the patterns. In the traditional decision tree method, it will be difficult to fit complex models (such as SVMs) if the tree is so large that each only has few examples. In other words, ensemble learning, such as bagging method, can produce a strong learner which has more flexibility and complexity than single model, for instance, decision tree.

The rule generation stage proceeds in two steps: In first step, the SVM model, which is constructed by best fold of CV, is applied to predict the labels of SVs, and the original labels of SVs are discarded. Hence, the artificial synthetic data are generated. During second step, the artificial data are used to train an RF model, and all decision trees of RF are the generated rule sets. Finally, the performance of the rule sets are evaluated on 10% remained test data, the precision, recall, and F-measure are used to estimate the accuracy of the rule sets.

4. RESULT AND DISCUSION

This section present the experiment result are Performance evaluation of existing SVM + C4.5 rule extraction, SVM + Random Forest rule extraction and proposed Rule pruning with SVM + Random Forest rule extraction are done by using performance metrics such as Accuracy, Precision, Recall and F-Measure.

Accuracy

The accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined.

$$\text{Accuracy}= \frac{\text{True positive} + \text{True negative}}{\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative}}$$

Precision

Precision value is evaluated according to the relevant information at true positive prediction, false positive.

$$\text{Precision} = \frac{\text{Truepositive}}{(\text{Truepositive} + \text{Falsepositive})}$$

Recall

The Recall value is evaluated according to the classification of data at true positive prediction, false negative.

$$\text{Recall} = \frac{\text{Truepositive}}{(\text{Truepositive} + \text{Falsenegative})}$$

F-Measure

F-Measure is defined as the harmonic mean of precision and recall. It is given by,

$$\text{F} - \text{Measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{(\text{Precision} + \text{Recall})}$$

PERFOMANCE MEASURE USED IN THE STUDY

The performance measure used in the study listed below.

Accuracy

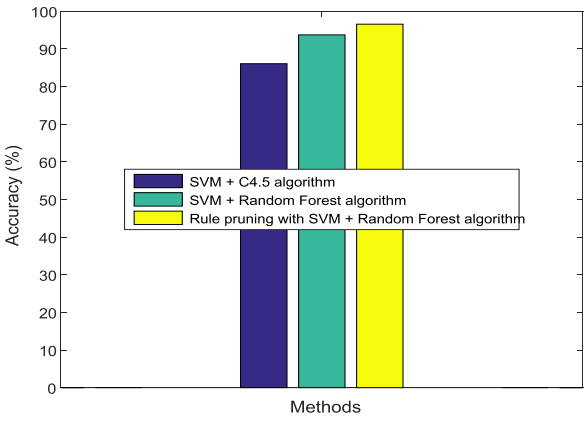

Fig 4.1 comparing the accuracy using existing and proposed method

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

Figure 4.1 shows that the comparison result of the existing SVM + C4.5 rule extraction, SVM + Random Forest rule extraction and proposed Rule pruning with SVM + Random Forest rule extraction labeled in x-axis. Accuracy values are taken in y-axis. Due to the removal of unnecessary rules from generated rule set by effective pruning algorithm, prediction accuracy of diabetes gets increases.

Table 4.1 comparing the accuracy among existing and proposed method

Algorithm	Accuracy
Support vector machine+C4.5 algorithm (Existing)	**81.6406**
Support vector machine + Random forest (Existing)	**89.7135**
Rule pruning with SVM+Random Forest (Proposed)	**94.6615**

Precision

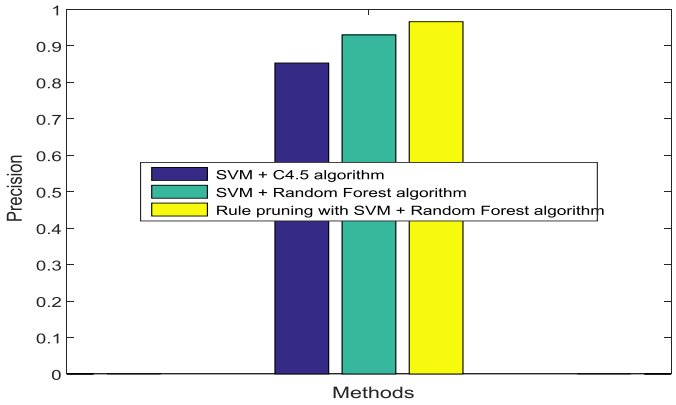

Fig.4.2 comparing the precision using existing and proposed method

Figure 4.2 shows the precision comparison result of existing SVM + C4.5 rule extraction, SVM + Random Forest rule extraction and proposed Rule pruning with SVM + Random Forest rule extraction labeled in x-axis. Precision values are taken in y-axis. Due to effective pruning of rule set, prediction accuracy increases by which precision result also get increases.

Table 4.2 comparing the precision among existing and proposed method

Algorithm	Precision
Support vector machine+C4.5 algorithm (Existing)	**0.7994**
Support vector machine + Random forest(Existing)	**0.8826**
Rule pruning with SVM+Random Forest(Proposed)	**0.9367**

Recall

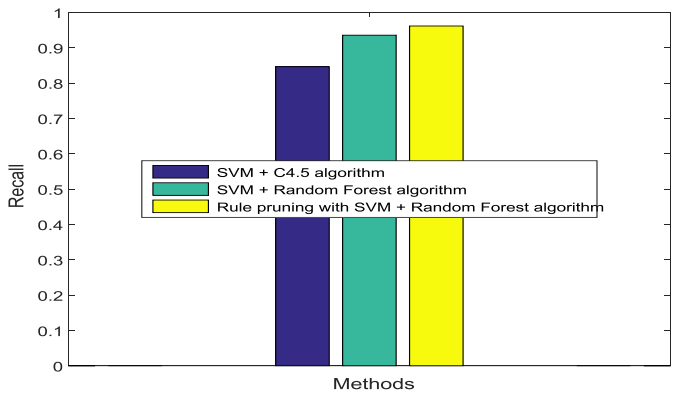

Fig4.3 comparing the recall using existing and proposed method

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

Figure 4.3 shows the comparison result of recall among existing SVM + C4.5 rule extraction, SVM + Random Forest rule extraction and proposed Rule pruning with SVM + Random Forest rule extraction labeled in x-axis. Recall values are taken in y-axis. Due to effective pruning of rule set and reduction in rule set size, recall value increases.

Table 4.3Comparing the recall among existing and proposed method

Algorithm	Recall
Support vector machine+C4.5 algorithm (Existing)	**0.8200**
Support vector machine + Random forest (Existing)	**0.8994**
Rule pruning with SVM+Random Forest(Proposed)	**0.9486**

F-Measure

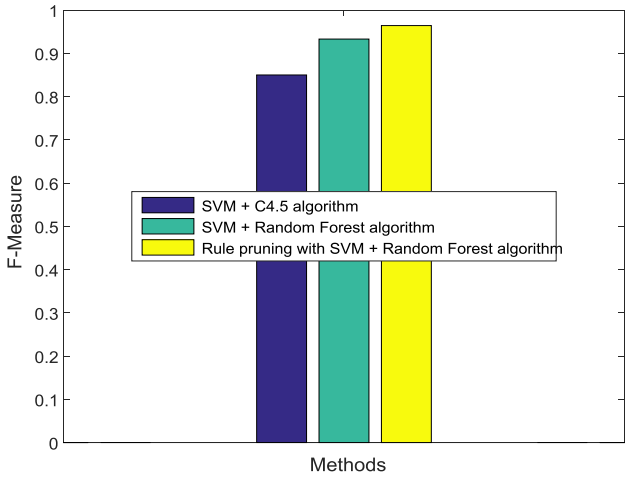

Fig4.4Comparing the F-measure using existing and proposed method

Figure 4.4 shows the F-measure result of existing SVM + C4.5 rule extraction, SVM + Random Forest rule extraction and proposed Rule pruning with SVM + Random Forest rule extraction labeled in x-axis. F-measure values are taken in y-axis. Due to effective pruning of rule set, F-measure increases with reduction in rule set size.

Table 4.4 Comparing the F-measure among existing and proposed method

Algorithm	F-measure
Support vector machine+C4.5 algorithm (Existing)	**0.8096**
Support vector machine + Random forest (Existing)	**0.8909**
Rule pruning with SVM+ Random Forest (Proposed)	**0.9426**

5. CONCLUSION

In This thesis Ensemble Pruning based on Frequent Patterns (EP-FP) is utilized with SVM and RF ensemble learning method to reduce the size of generated rule set. The purpose of reducing rule set size is to improve the diagnostic accuracy of diabetes. Rule pruning algorithm with SVM and Random Forest rule extraction is also used find the risk level of diabetes by periodically monitor the glucose level of detected patients under various conditions using data files. Experiments are conducted and from the Performance comparison result, it is proved that, better results are achieved by Rule pruning with SVM + Random Forest rule extraction in terms of diagnostic accuracy, precision, recall and F-Measure. The future extensions of this research is to improve the diagnosis model which focuses the factors such as food, medicines and exercise activities which are taken by the patients. These factors will help us to recognize the risk level of glucose tolerance. Furthermore, it will help to determine the status of diabetes risk derived from both impaired fasting glucose and impaired glucose tolerance.

REFERENCES

1. Hay, L. C., Wilmshurst, E. G., &Fulcher, G. (2003). Unrecognized hypo-and hyperglycemia in well-controlled patients with type 2 diabetes mellitus: the results of continuous glucose monitoring. Diabetes technology & therapeutics, 5(1), 19-26.

2. Gupta, A., Gupta, R., Sarna, M., Rastogi, S., Gupta, V. P., & Kothari, K. (2003). Prevalence of diabetes, impaired fasting glucose and insulin resistance syndrome in an urban Indian population. Diabetes research and clinical practice, 61(1), 69-76.
3. Wang, J. J., Yuan, S. Y., Zhu, L. X., Fu, H. J., Li, H. B., Hu, G., &Tuomilehto, J. (2004). Effects of impaired fasting glucose and impaired glucose tolerance on predicting incident type 2 diabetes in a Chinese population with high post-prandial glucose. Diabetes research and clinical practice, 66(2), 183-191.
4. 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.
5. Yoo, H. J., An, H. G., Park, S. Y., Ryu, O. H., Kim, H. Y., Seo, J. A.& Choi, K. M. (2008). Use of a real time continuous glucose monitoring system as a motivational device for poorly controlled type 2 diabetes. Diabetes research and clinical practice, 82(1), 73-79.
6. Amini, M., &Janghorbani, M. (2009). Comparison of metabolic syndrome with glucose measurement for prediction of type 2 diabetes: The Isfahan Diabetes Prevention Study. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 3(2), 84-89.
7. Rambod, M., Hosseinpanah, F., Ardakani, E. M., Padyab, M., &Azizi, F. (2009). Fine-tuning of prediction of isolated impaired glucose tolerance: A quantitative clinical prediction model. diabetes research and clinical practice, 83(1), 61-68.
8. Tankova, T., Chakarova, N., Atanassova, I., &Dakovska, L. (2011). Evaluation of the Finnish Diabetes Risk Score as a screening tool for impaired fasting glucose, impaired glucose tolerance and undetected diabetes. Diabetes research and clinical practice, 92(1), 46-52.
9. Bianchi, C., Miccoli, R., Bonadonna, R. C., Giorgino, F., Frontoni, S., Faloia, E., ...&Consoli, A. (2011). Metabolic syndrome in subjects at high risk for type 2 diabetes: the genetic, physiopathology and evolution of type 2 diabetes (GENFIEV) study. Nutrition, Metabolism and Cardiovascular Diseases, 21(9), 699-705.
10. McEwen, L. N., Adams, S. R., Schmittdiel, J. A., Ferrara, A., Selby, J. V., & Herman, W. H. (2013). Screening for impaired fasting glucose and diabetes using available health plan data. Journal of diabetes and its complications, 27(6), 580-587.
11. Bethel, M. A., Chacra, A. R., Deedwania, P., Fulcher, G. R., Holman, R. R., Jenssen, T., ... &Raptis, S. A. (2013). A novel risk classification paradigm for patients with impaired glucose tolerance and high cardiovascular risk. The American journal of cardiology, 112(2), 231-237.
12. Fong, S., Mohammed, S., Fiaidhi, J., &Kwoh, C. K. (2013). Using causality modeling and Fuzzy Lattice Reasoning algorithm for predicting blood glucose. Expert Systems With Applications, 40(18), 7354-7366.
13. 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.
14. 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.
15. Fonville, S., den Hertog, H. M., Zandbergen, A. A., Koudstaal, P. J., &Lingsma, H. F. (2014). Occurrence and predictors of persistent impaired glucose tolerance after acute ischemic stroke or transient ischemic attack. Journal of Stroke and Cerebrovascular Diseases, 23(6), 1669-1675.
16. Franco, L. J., Dal Fabbro, A. L., Martinez, E. Z., Sartorelli, D. S., Silva, A. S., Soares, L. P., ... &Moisés, R. S. (2014). Performance of glycatedhaemoglobin (HbA1c) as a screening test for diabetes and impaired glucose tolerance (IGT) in a high risk population—The Brazilian Xavante Indians. Diabetes research and clinical practice, 106(2), 337-342.
17. Kim, Y. A., Ku, E. J., Khang, A. R., Hong, E. S., Kim, K. M., Moon, J. H., ... & Lim, S. (2014). Role of various indices derived from an oral glucose tolerance test in the prediction of conversion from prediabetes to type 2 diabetes. Diabetes research and clinical practice, 106(2), 351-359.
18. Lee, Y. B., Lee, J. H., Park, E. S., Kim, G. Y., &Leem, C. H. (2014). Personalized metabolic profile estimations using oral glucose tolerance tests. Progress in biophysics and molecular biology, 116(1), 25-32.
19. Aguiar, E. J., Morgan, P. J., Collins, C. E., Plotnikoff, R. C., &Callister, R. (2015). Characteristics of men classified at high-risk for type 2 diabetes mellitus using the AUSDRISK screening tool. Diabetes research and clinical practice, 108(1), 45-54.
20. Jelinek, H. F., Stranieri, A., Yatsko, A., &Venkatraman, S. (2016). Data analytics identify glycatedhaemoglobin co-markers for type 2 diabetes mellitus diagnosis. Computers in biology and medicine, 75, 90-97.
21. Crofts, C., Schofield, G., Zinn, C., Wheldon, M., & Kraft, J. (2016). Identifying hyperinsulinaemia in the absence of impaired glucose tolerance: An examination of the Kraft database. Diabetes Research and Clinical Practice, 118, 50-57.
22. Jeong, B., Jung, C. H., Lee, Y. H., Shin, I. H., Kim, H., Bae, S. J., ... & Park, J. Y. (2016). A novel imaging platform for non-invasive screening of abnormal glucose tolerance. Diabetes Research and Clinical Practice, 116, 83-85.
23. Meijnikman, A. S., De Block, C. E. M., Verrijken, A., Mertens, I., Corthouts, B., & Van Gaal, L. F. (2016). Screening for type 2 diabetes mellitus in overweight and obese subjects made easy by the FINDRISC score. Journal of diabetes and its complications.
24. Han, L., Luo, S., Yu, J., Pan, L., & Chen, S. (2015). Rule extraction from support vector machines using ensemble learning approach: an application for diagnosis of diabetes. IEEE journal of biomedical and health informatics, 19(2), 728-734.