Expert Systems With Applications 130 (2019) 188–205

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A rule extraction approach from support vector machines for diagnosing hypertension among diabetics

Namrata Singh[a,*], Pradeep Singh[a], Deepika Bhagat[b]

[a] Department of Computer Science and Engineering, National Institute of Technology, Raipur 492001, Chhattisgarh, India
[b] Department of Medicine, Dr. Bhim Rao Ambedkar Memorial Hospital, Pt. JNM, Medical College, Raipur 492001, Chhattisgarh, India

ARTICLE INFO

Article history:
Received 12 June 2018
Revised 16 April 2019
Accepted 16 April 2019
Available online 17 April 2019

Keywords:
Diabetes
Extreme gradient boosting
Hypertension
Medical diagnosis
Rule extraction
Support vector machine

ABSTRACT

Diabetes mellitus is a major non-communicable disease ever rising as an epidemic and a public health crisis worldwide. One of the several life-threatening complications of diabetes is hypertension or high blood pressure which mostly remains undiagnosed and untreated until symptoms become severe. Diabetic complications can be greatly reduced or prevented by early detection of individuals at risk. In recent past, several machine learning classification algorithms have been widely applied for diagnosing diabetes but very few studies have been conducted for detecting hypertension among diabetic subjects. Specifically, existing rule-based models fail to produce comprehensible rule sets. To resolve this limitation, this paper endeavours to develop a hybrid approach for extracting rules from support vector machines. A feature selection mechanism is introduced for selecting significantly associated features from the dataset. XGBoost has been utilized to convert SVM black box model into an apprehensible decision-making tool. A new dataset has been obtained from Pt. JNM, Medical College, Raipur, India comprising of 300 diabetic subjects with 108 hypertensives and 192 normotensives. In addition, five public diabetes-related datasets have been taken for generalization of the results. Experiments reveal that the proposed model outperforms ten other benchmark classifiers. Friedman rank and post hoc Bonferroni-Dunn tests demonstrate the significance of the proposed method over others.

1. Introduction

1.1. Background

Diabetes mellitus has emerged as one of the largest global health problems of the 21st century. According to the International Diabetes Federation (IDF), 415 million adults are currently affected from diabetes which is expected to rise to 642 million by 2040, posting a huge hike of 10.4% during next two decades. Additionally, the prevalence of impaired glucose tolerance or pre-diabetes is also increasing at a fast pace. The estimates reveal that over 481 million persons will suffer from pre-diabetes by 2040 (Ogurtsova et al., 2017). Globally, over a million children and adolescents are suffering from T1DM, and one in six live births is being affected by gestational diabetes ("International Diabetes Federation IDF Diabetes Atlas-8th Edition," 2017). It is estimated that almost half (49.7%) of all people living with diabetes are undiagnosed. The highest percentage of undiagnosed diabetes was found in Africa (~70%), South-East Asia (~60%) and Western Pacific (54%) regions (Nanditha et al., 2016). Moreover, each year approximately 5 million deaths across the globe are attributable to diabetes in the age range of 20–99 years (Cho et al., 2018).

Diabetes mellitus is a group of metabolic disorders characterized by hyperglycemia resulting in elevated blood sugar. It is caused due to impaired insulin secretion, defective insulin action, or both. The majority of cases of diabetes mellitus are extensively classified into two groups: Type I (T1DM) and Type II (T2DM). T1DM occurs when the pancreas are unable to produce insulin due to autoimmune destruction of pancreatic beta cells resulting in insulin deficiency. Another most common form of diabetes is T2DM where the body does not effectively use the insulin produced. One more category of diabetes is gestational diabetes that develops during pregnancy. The complications related to diabetes are liable for the considerable morbidity and mortality. Chronic hyperglycemia of diabetes is correlated with microvascular complications such as retinopathy, nephropathy, and neuropathy. Therefore, diabetes is the leading cause of blindness, renal failure, impotence and diabetic foot disorders whose severity can cause lower-limb amputations. The macrovascular complications of diabetes include cardiovascular diseases such as heart attacks, strokes and cerebrovascu-

* Corresponding author.
E-mail addresses: nsingh.phd2016.cs@nitrr.ac.in (N. Singh), psingh.cs@nitrr.ac.in (P. Singh).

https://doi.org/10.1016/j.eswa.2019.04.029
0957-4174/

lar disease. Consequently, considering the severity of these health complications, early detection and prevention of diabetes and its associated risk factors can reduce its incidence and related healthcare expenditures.

Hypertension is the leading cause of deaths and cardiovascular diseases worldwide (Lopez-jaramillo, Lopez-Lopez, Lopez-Lopez, & Rodriguez-Alvarez, 2014), accounting for about 12.8% (7.5 million) deaths per year (World Health Organization, 2013). Hypertension, also known as high blood pressure (systolic BP $\geq$ 140 mmHg and/or diastolic BP $\geq$ 90 mmHg) accelerates the progression of all-cause mortality, stroke, coronary artery disease, and diabetes mellitus (Sowers, Epstein, & Frohlich, 2001). The development of diabetes among hypertensive patients is approximately three times more frequent than in normotensive individuals. Both hypertension as well as diabetes are common diseases which occur at a high frequency and share common causes, therefore they are liable to develop in combination (Katayama, Hatano, & Issiki, 2018; Liu et al., 2013). As a result, there arises need for development of suitable models which can perform accurate prediction of hypertension among diabetic patients. Recently, systematic reviews of various risk prediction models developed for detecting hypertension among patients with T2DM were published (van Dieren et al., 2012; Liu et al., 2013; Teramukai et al., 2016).

Various factors causing high prevalence of hypertension among diabetic patients are inadequate physical activity, unhealthy dietary habits, obesity, in particular abdominal obesity, smoking etc., which if not controlled can produce long-term effects. Recent researches have shown that at least 80% of diabetes-related complications can be reduced or prevented by early identification and intervention among high-risk groups, achieved by incorporating appropriate lifestyle changes or by medications and other therapeutic techniques. Thus, there arises a need for intelligent data analysis techniques such as data mining and machine learning to identify people suffering from diabetes with hypertension. In the last few years, several novel learning methods have been proposed for the effective diagnosis, prognosis, management and treatment of diabetes (Kavakiotis et al., 2017; Jayanthi, Babu, & Rao, 2017; Maniruzzaman et al., 2017; Marling, Wiley, Bunescu, Shubrook, & Schwartz, 2012). Computer-Aided Diagnosis (CAD) systems have been employed for enhancing the diagnostic capabilities of physicians and reducing the time required for accurate diagnosis of diabetes (Chen & Pan, 2018). Additionally, clinical decision support systems (CDSS) aid clinicians in managing diabetes by providing enhanced patient safety and quality of health-care services (Rodbard & Vigersky, 2011; Sim, Ban, Tan, Sethi, & Loh, 2017).

In this context, several rule-based classification approaches such as C4.5, CART, RF, XGBoost and non-rule-based approaches such as k-NN, SVM, Naïve Bayes, ANN, etc., have been applied by various researchers for diagnosis and treatment of diseases. Although many rule-based algorithms exist, yet majority of them fail to generate rules which are more expressive, easily interpretable, comprehensible, balanced and optimal. It still remains a challenging task to generate an optimal ruleset with less number of rules and shorter antecedents that produces higher classification accuracy while maintaining a balance between both sensitivity and specificity. Hence, this paper aims at constructing a hybrid rule-based model for diagnosis of hypertension among diabetic individuals by integration of a non-rule-based classification algorithm with a rule-based algorithm. The advantages of the non-rule-based algorithm are incorporated into the rule-based approach.

1.2. Research motivations

Rule extraction from SVMs follow the earlier methodology to acquire human-understandable rules from ANNs for explaining "how" a decision was formed or "why" a certain result was attained. SVM rule extraction is a natural variant of the ANN rule extraction domain. Therefore, much of the motivation for rule extraction from SVMs is carried out from the well-established area of rule extraction from NNs (Diederich, 2008). The main motivation behind rule extraction is due to following two reasons: (1) To interpret the classifications performed by the underlying nonlinear black-box model; and (2) To improve the performance of the rule induction techniques. One of the most common motivation for extraction of rules is to obtain a set of classification rules that can best explain the black-box model. The set of obtained rules which mimic the predictions of SVM helps in attaining insights into the logical workings of SVMs. Rule extraction approaches provide the integration of an explanation capability within a trained SVM. Therefore, utilizing the trained black-box model for training the rule-based classifier can lead to a much better performance.

Over the past few years, many researchers have applied machine learning techniques for assessing diabetes and related complications. (LaFreniere, Zulkernine, Barber, & Martin, 2016) applied ANNs to predict hypertension using CPCSSN dataset; (Farran, Channanath, Behbehani, & Thanaraj, 2013) implemented four machine learning techniques: logistic regression, k-NN, multifactor dimensionality reduction, SVMs and developed risk assessment tools for T2DM, hypertension and comorbidity; (Luo, 2016) used rule-based associative classifier to develop a predictive model for T2DM diagnosis; (Hayashi & Yukita, 2016) proposed a novel recursive-rule extraction algorithm, *Re*-RX with J48graft, integrated with sampling selection techniques (sampling *Re*-RX with J48graft) for generation of classification rules on T2DM dataset; (Sakr et al., 2018) compared six machine learning classification algorithms namely LogitBoost (LB), Bayesian Network (BN), Locally Weighted Naïve Bayes (LWB), ANNs, SVMs and Random Tree Forest (RTF) for predicting risk of hypertension on cardiorespiratory fitness data; (Guo et al., 2017) performed cluster analysis using k-means clustering algorithm for investigating heterogeneity of hypertension and assessing cardiovascular risk; (Melin, Miramontes, & Prado-Arechiga, 2018) developed a hybrid model using modular neural networks and fuzzy logic for identifying risk of hypertension.

Previous studies have utilized numerous approaches for diagnosis and risk evaluation of hypertension by utilizing hybrid intelligent systems (Miramontes, Martínez, Melin, & Prado-Arechiga, 2017, 2018). (Guzmán, Melin, & Prado-Arechiga, 2018) proposed a study that aims to model 24 h blood pressure level and acquire the trend of each patient, which is then provided to the fuzzy system containing rules given by experts. These fuzzy rules are then optimized using genetic algorithms to find the best rules and obtain a better classification of blood pressure levels. A new model is introduced in Pulido, Melin, and Prado-Arechiga (2018) based on fuzzy system for arterial hypertension classification. Mamdani fuzzy inference system is utilized with Gaussian membership functions to classify patient's systolic and diastolic blood pressure levels. Another study (Guzman, Melin, & Prado-Arechiga, 2017) developed a new Neuro-Fuzzy Hybrid Model (NFHM) according to the definitions of the European Guidelines for the classification of Hypertension Blood pressure (HBP). The blood pressure levels were categorized into hypotension, optimal, normal, normal high, hypertension grade 1, hypertension grade 2, hypertension grade 3 and isolated systolic hypertension. The approach utilizes the 24 h patient monitoring database to obtain the blood pressure trend, that is classified using fuzzy system consisting of expert-based rules, which are further optimized with genetic algorithms to obtain the optimal number of rules.

For the classification problem, feature selection aims at selecting a subset of highly discriminant features from the dataset. Irrelevant features not only lead to insufficient classification accuracy, but also increases the algorithmic computational cost and complexity. The main goal of feature selection is to search for an op-

timal feature subset from the initial feature set that leads to improved classification performance and efficiency in generating classification model. Based on the interaction between feature selection and classification model, numerous feature selection methods have been developed which can be divided into three categories: (1) filter methods, (2) wrapper methods, and (3) embedded methods. Feature selection is one of the main issues in biomedical data classification wherein appropriate selection provides the significant biomarkers and risk factors for disease diagnosis and finally leads to improved predictive performance.

This paper introduces a hybrid learning approach for generation of classification rules to diagnose hypertension among diabetic individuals. The methodology utilizes an ensemble technique XGBoost for extracting rules from SVMs. At the initial stage, support vectors (SVs) are extracted from the SVM model with reasonable accuracy. Then, the trained SVM model is used for predicting the actual class labels of SVs, where a modified artificial dataset is formed by the SVs and their predicted labels. Lastly, the modified training dataset is given to XGBoost for generating rules and determining the performance of the proposed classification model. The XGBoost rule induction method together with SVs develops economical and beneficial evaluative rules for detecting hypertension.

Both the ensemble classifiers, RF and XGBoost utilize decision trees as base learners. XGBoost possesses a large number of hyperparameters which are immensely related to model performance. In RF, trees are built independent of each another while XGBoost builds trees to complement the already built ones. Moreover, XGBoost provides better accuracy with lesser number of trees and is less likely to overfit. The rule sets obtained from SVM+XGBoost are more representative than other hybrid approaches such as SVM+C4.5 and SVM+RF. The proposed model produces more concise and accurate rules which can help in much wider assessment of the subjects having hypertension. Moreover, the results depict that the proposed hybrid approach can be utilized as a promising tool for detection of diabetes and related risk factors.

Overall, the main contributions of this work can be summarized as follows:

- Introduction of feature selection as a pre-processing step to choose significant biomarkers and risk-factors from the six diabetes-related datasets.
- Developing a hybrid model based on rule extraction approach from SVMs using XGBoost, an implementation of gradient boosted decision trees.
- Comparison of the proposed SVM+XGBoost approach with two hybrid and eight existing ensemble, individual and rule-based approaches.
- Generation of cross-validated rulesets with computation of average rule length
- Computation of probability-based objective interestingness measures such as support and confidence which provide much extensive interpretation of the obtained rules.

The organization of the paper is as follows. Section 2 presents the background of rule extraction techniques from SVMs. Section 3 introduces the proposed rule extraction methodology. Section 4 describes the dataset and empirical setup of our study with their results and discussions in Section 5. Section 6 represents the conclusions, limitations drawn from this study along with future research directions.

2. Related work

In medical diagnosis, explanation of decision for classification of the disease is an important requirement for acceptance of black-box models like SVMs by final users (Fung, Sandilya, & Rao, 2005). SVMs suffer from a significant drawback of their inability to provide comprehensible justifications for the classification outcomes they make. This is the reason why during the recent past approaches for rule-extraction are being proposed (Barakat, Bradley, & Barakat, 2010) enabling SVMs to be more understandable. However, not only SVMs but ANNs also produce black-box models (Andrews, Diederich, & Tickle, 1995) due to their inherent incapability to explain models and their results with lack of clarity and intelligibility to individuals. This argument becomes the main driving force and motivation behind the rule extraction studies from both SVMs and ANNs particularly in the area of medical diagnosis (Barakat & Diederich, 2005).

Over the past few decades, there has been a rapid increase in proposed approaches related to rule-extraction techniques from SVMs. Altogether, these methods can be classified into three major groups: “pedagogical” or “learning-based”, “decompositional” or “transparent”, and “eclectic” or “hybrid” (Han, Luo, Yu, Pan, & Chen, 2015). The pedagogical rule-extraction techniques consider the underlying classifier such as SVM as a black box model and employ a learning algorithm to provide rule-based interpretations. The main focus is to utilize the trained SVM classifier for generation of instances to be used as input to the learning algorithm which produces the rules as output. The Genetic Rule Extraction (G-REX) algorithm (Johansson, König, & Niklasson, 2003) uses genetic programming to produce an optimal set of rules which can be of boolean, fuzzy, m-of-n type. The versatility of G-REX (Johansson, König, & Niklasson, 2004) has been established by its application to different types of models and generation of multitude of representations. However, the decompositional approaches use SVM as an oracle and extract the rules by employing SVs of the constructed hyperplane. In case of ANNs, the rules are extracted at the level of individual units within the trained ANN and focuses on the symbolic representation of hidden and output associations in the network. Zhang, Su, Jia, and Chu (2005) presented a hyperrectangle rule extraction (HRE) algorithm for extraction of rules from the trained SVMs. Support Vector Clustering (SVC) algorithm was used to find prototype vectors for each class and then using those vectors with SVs for generation of hyperrectangles. Based on the similar notion of producing hyperrectangle rules, Fu, Ong, Keerthi, Hung, and Goh (2004) proposed RulExSVM for extraction of rules from nonlinear SVMs with kernel function RBF. This technique utilizes the SVs of SVM decision boundary and executes in three steps: initial, tuning and pruning. For extracting rules from linear kernel SVMs, a linear programming formulation was introduced by Fung et al. (2005). This technique considers the rule extraction process as a multiple constraint optimization problem. Another decision function based decompositional study, suggested by Zhang, Li, Tang, and Cui (2004) known as “DRC-BK”, trained with Disjunctive Normal Form (DNF) boolean kernel extracted rules from SVMs.

Further, the eclectic approaches are the combination of both decompositional and pedagogical methods (Tickle, Andrews, Golea, & Diederich, 1998). This approach uses the knowledge gained by the SVM model that is represented in the form of SVs and the parameters related with them. Extensive work has been performed for development of hybrid rule extraction methods for ANNs but less work has been accomplished for extraction of rules from SVMs. A rule extraction method for RBFNN was developed which utilizes the algorithm for construction of ellipsoids and hyperrectangles proposed for SVM (Núñez, Angulo, & Català, 2004). Farquad, Ravi, and Bapi (2008) proposed a hybrid approach for extracting fuzzy rules from SVMs and evaluated the generated rules against the DT classifier built from same SVM model. FRBSs were developed for medical data classification (Gorzałczany & Rudziński, 2017;Malmir, Amini, & Chang, 2017), where hybrid approach SVM+FRBS (Farquad, Ravi, & Bapi, 2009) was compared with other approaches such as SVM + ANFIS,

Table 1
Overview of the existing rule-extraction algorithms.

Algorithm	Type	BB method	Format of Rules	Explanation	EvaluationMeasures	Ref.
Trepan	P	ANN	m-of-n splits of tree	decision tree induction	A, C, F	Craven and Shavlik (1996)
SQRex-SVM	D	SVM	propositional rules	modified sequential covering algorithm	A, C, F	Barakat and Bradley, (2007)
Eclectic Rule Ext.	E	SVM	decision tree	SVs+C5.0	A, C, F	Barakat and Diederich (2005)
Eclectic Rule Ext. (ROC)	E	SVM	decision tree	SVs+C5.0	A, F, AUC	Barakat and Bradley (2006)
Hybrid Rule Ext.	E	SVM	decision tree	SVs+C5.0	A, C, F	Diederich and Barakat (2004)
DEDEC	E	ANN	propositional rules	ranking of ANN inputs	A, C, F	Tickle, Orlowski, and Diederich (1994) Andrews et al. (1995)
SVM+Prototypes	D	SVM	ellipsoid, interval type equation rules	clustering using RBF	A, C, F, Cv, Ov	Núñez, Angulo, and Català (2002)
ALBA	D	SVM	open	SVs+active learning	A, C, F	Martens, Baesens, and Van Gestel (2009)
ALPA	P	open	open	(SVM, ANN, RF)+(C4.5, Ripper)+active learning	A, C, F	de Fortuny and Martens (2015)
ANN-DT	P	ANN	decision tree	extraction of binary decision tree	A, C, F	Schmitz, Aldrich, and Gouws (1999)

Note: P = Pedagogical; D = Decompositional; E = Eclectic; BB = Black Box; ANN = Artificial Neural Network; SVM = Support Vector Machine; RBF = Radial Basis Function; RF = Random Forest; A = Accuracy; C = Comprehensibility; F = Fidelity, Cv = Coverage; Ov = Overlapping; AUC = Area Under Curve.

SVM+DT and SVM+RBF. Furthermore, a hybrid rule-extraction technique (Farquad, Ravi, & Raju, 2010) was developed to extract rules from SVMs and Support Vector Regression (SVR) for solving classification and regression problems. Artificial Immune System (AIS) based rule-extraction technique (Kahramanli & Allahverdi, 2009) and other hybrid approaches such as Pourpanah, Lim, and Saleh (2016), Stoean and Stoean (2013), Seera and Lim (2014), generate efficient rulesets for medical data interpretation. An overview of the existing rule-extraction algorithms is shown in Table 1.

3. Proposed rule-extraction approach

The proposed rule-extraction technique is divided into two major steps. In the first step, the SVM model is constructed utilizing the training data by tuning the hyper-parameters within the search space to obtain an acceptable accuracy. The instances which lie close to the decision boundary known as support vectors (SVs) are extracted from the SVM model. Henceforth, this model is used for predicting class labels of the SVs. An artificial dataset is generated by substituting actual labels with the predicted labels of SVs. Since SVs contain noisy class labels, therefore replacing the original labels with those predicted by the SVM eliminates the label noise from the artificial dataset. In addition, the generated rules not only imitate the predictions of SVMs but also provide a better understanding of their internal workings. In the second step, the artificial dataset known as modified training dataset is given to XGBoost algorithm with tuned hyper-parameters for generation of best rule sets. Finally, the evaluation of the rules is performed on testing data. The proposed SVM+XGBoost technique is shown in Fig. 1.

3.1. Data preparation methodology

The diabetic hypertension (JNMMC) dataset is used for predicting the presence of positive versus negative class. The entire dataset is initially divided into two parts, 90% of the training data is utilized for generation of rules from SVMs and remaining 10% is used as testing data. During the rule-extraction process, the optimal tuning hyper-parameters for the models are obtained using 10-fold CV. Then these hyper-parameters are utilized for obtaining the trained model. Further, this model is used for generation of rules and evaluation by the testing data. The average results of performance evaluation measures such as accuracy, precision, recall, F-measure and AUC for the model are obtained by conducting 10 runs of 10-fold CV. Accordingly, the rules generated for each of the 10 folds are assessed by ruleset size and mean rule length.

3.2. Support vector machines (SVMs)

SVM is a learning procedure that involves the structural risk minimization (Cortes & Vapnik, 1995) principle and statistical learning theory (Vapnik, 1995). SVMs possess the robust capabilities of non-linear processing, generalization and handling small-samples due to which they are mostly applied in classification decision making and medical diagnosis problems. The SVM model uses a training dataset $\{x_i, y_i\}_{i=1}^{l}$ where l denotes the number of samples and $x_i \in R^p$ is the input training samples with corresponding class labels $y_i \in \{-1, +1\}$. The goal of SVM is to maximize the generalization capability of the model by widening the margin between the decision hyperplane and datapoints. The SVM model tries to find an optimal separating hyperplane by maximizing the margin $d = 2/||w||$. This hyperplane is defined as $w^T x + b = 0$ where w and b depict the weight vector and bias term

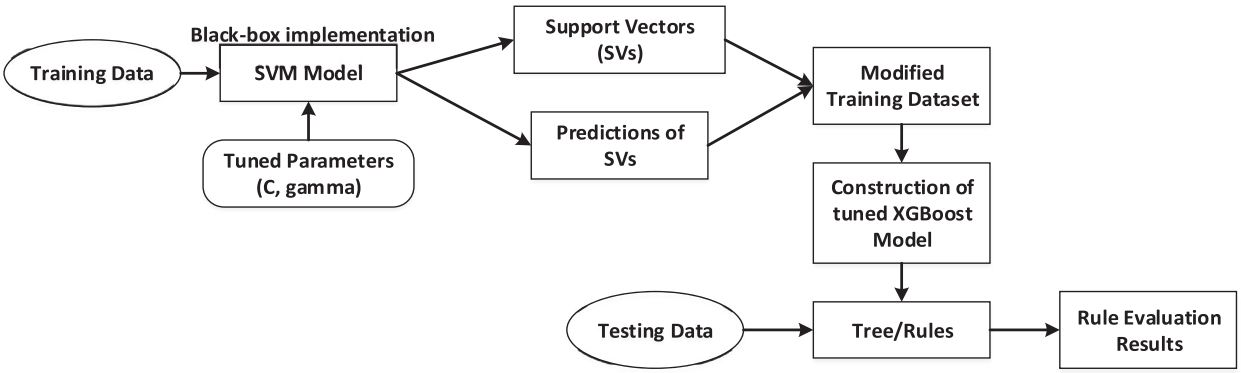

Fig. 1. Schematic representation of the proposed rule extraction system.

respectively. Further, for maximizing the margin, $||w||$ needs to be minimized. This leads to formulation of a convex optimization problem with quadratic criterion and linear inequality constraints:

$$\min \frac{1}{2}||w||^2 \text{ subject to } y_i(w^T x_i + b) \geq 1,\ i = 1, \ldots, N \tag{1}$$

To deal with the problem of overlapping classes, SVM formulation uses non-negative slack variables for tolerating misclassifications in training data. After introduction of slack-variable (ξ) into the constraint of Eq. (1), it changes to:

$$w^T x_i + b \geq 1 - \xi_i, \text{ if } y_i = +1 \tag{2}$$

$$w^T x_i + b \leq -1 + \xi_i, \text{ if } y_i = -1 \tag{3}$$

where $\forall_i$: $\xi_i \geq 0$.

By introduction of the regularization term $\frac{1}{2}||w||^2$ and slack variable (ξ), the quadratic optimization problem can be formulated as:

$$\min_{w,b,\xi} \delta(w, b, \xi) = \frac{1}{2}||w||^2 + C\sum_{i=1}^{n} \xi_i \tag{4}$$

subject to $y_i(w^T x_i + b) \geq 1 - \xi_i$, $\xi_i \geq 0$, $1 \leq i \leq n$.where $||.||^2$ represents the L_2-norm and C is the penalty parameter that balances the trade-off between margin maximization and minimization of misclassification error.

The dual formulation for the above soft-margin problem (or primal problem) given in Eq. (4) can be obtained by introducing Lagrange multiplier α. Thus, Eq. (4) becomes

$$\max D(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y_i \alpha_i y_j \alpha_j (x_i, x_j) \tag{5}$$

subject to $\sum_{i=1}^{n} \alpha_i y_i = 0$, $0 \leq \alpha_i \leq C$, $\forall i$.

After obtaining optimal α values, the decision function for a new input vector x is specified as:

$$f(x) = \text{sgn}\left(\sum_{i=1}^{n} \alpha_i y_i ({x_i}^T x) + b\right) \tag{6}$$

However, in real world applications, the data is rarely linearly inseparable. For solving the non-linearity problem, the original data is projected through non-linear mapping $\phi(x)$ to a higher dimensional feature space, where the data points are linearly separable. The input space X is mapped by a feature map ϕ: $X \rightarrow H$ (H defines the Hilbert space) onto new higher dimensions. A symmetric function expressed as an inner product in the new feature space satisfies the following Mercer's theorem, which is known as the kernel function.

$$K(x_i, x) = < \varphi^T(x_i), \varphi(x) > \tag{7}$$

With the help of kernel functions, SVMs can be applied to non-linear cases. Various types of kernels and their parameter configurations conform to different nonlinear problem-handling behaviours. One of the most popular and generally used kernel function is radial basis function (RBF) kernel:

$$K(x, y) = e^{-\frac{||x-y||^2}{2\sigma^2}},\ \sigma > 0 \tag{8}$$

where $||x - y||^2$ denotes the square of Euclidean distance between the two feature vectors and σ denotes a free parameter. Also, the parameter gamma is defined as:

$$\gamma = \frac{1}{2\sigma^2} \tag{9}$$

Thus, Eq. (8) can be represented as:

$$K(x, y) = e^{-\gamma||x-y||^2},\ \gamma > 0 \tag{10}$$

The computational performance of SVM highly depends upon the proper selection of hyperparameters. Selection of inappropriate values of hyper-parameters can cause underfitting or overfitting of models which not only affects the classification accuracy but also fails to provide an optimized model (Wang, Huang, & Cheng, 2014). The hyper-parameter γ controls the dispersion of the kernel in the input space. Higher values of gamma correspond to large number of support vectors (SVs) which overfits the model and leads to generalization error. C has an impact on the number of SVs, with increase in C the number of SVs gradually decrease (Hsu & Lin, 2002). The selection of an appropriate size penalty factor C can prevent SVM from falling into both the phenomena of underfitting as well as overfitting (Wang et al., 2014).

In a non-linear case, the dual form according to the Karush-Kuhn-Tucker (KKT) conditions (Boyd & Vandenberghe, 2004) is given as:

$$\max D(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y_i \alpha_i y_j \alpha_j K(x_i, x_j) \tag{11}$$

subject to $\sum_{i=1}^{n} \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, \forall i$.

Subsequently, the final decision function becomes:

$$f(x) = \text{sgn}\left(\sum_{i=1}^{S} \alpha_i y_i K({x_i}^T x) + b\right) \tag{12}$$

However, the evaluation of a kernel SVM is costlier than the evaluation of a linear SVM. The reason being that, additionally S number of support vectors x_i and corresponding coefficients $\alpha_i \in$

R are needed. Theoretically, the number of support vectors S can increase linearly with the number of training samples (Steinwart, 2003). Furthermore, these SVs play a significant role in determining the classification hyperplane (Ding, Hua, & Yu, 2014). By tuning the parameters C and gamma with 10-fold cross validation, SVs can be obtained from the optimized SVM model with acceptable accuracy.

3.3. Rule generation and evaluation process

Extreme Gradient Boosting (XGBoost) developed by Chen and Guestrin (2016) is a powerful ensemble gradient boosting method for prediction. It is an improved implementation of the gradient boosted trees algorithm originally proposed by Friedman (2001, 2002). Gradient boosting is a learning technique that constructs a strong learner by integrating various weak learners to provide a more regularized model formulation that controls overfitting. This approach builds a multitude of decision trees one after the other which correctly predicts those instances that were wrongly predicted in the previous tree. Unlike XGBoost, RF constructs a group of independent and non-identical decision trees based on the concept of randomization. Another distinction is that XGBoost constructs numerous shallow trees whereas RF builds fewer but deeper trees. As compared to RF, the scalability and regularization component of XGBoost enables the creation of more generalized ensembles. Overall, due to its significant advantages over RF, XGBoost is utilized for generation of rule sets.

The process of rule generation advances in two stages. During the first stage, the SVM model is built using training data of the 10-fold CV. During each fold, this model is utilized for predicting the class labels of SVs, where a modified or artificial dataset is generated by discarding the original labels of SVs. In the second stage, the artificial dataset is used for training the XGBoost model and generating the rule sets. The rules are evaluated on the remaining 10% of test data for determining the accuracy, precision, recall and F-measure. In addition, ruleset size and mean rule length are also calculated for each fold of cross-validation.

4. Experimental procedure

The present study proposes a hybrid (SVM + XGBoost) approach for extracting rules from SVMs. This technique utilizes the knowledge given by the support vectors (SVs) of the SVM model and integrates it to boosted tree ensemble methods for extracting rules from the black box SVM model. Initially, the accuracy of SVM was compared with XGBoost, RF, C4.5, NB, BP NN, RIPPER and PART approaches, which proved to be the reason for extracting rules from SVMs. Then, SVM + C4.5, an eclectic rule-extraction approach (Barakat & Diederich, 2005) and SVM + RF, rule-extraction using Random Forest (Wang, Deng, & Choi, 2015) were employed for comparing the learning potential of the rules with the proposed SVM + XGBoost technique. The difference among these approaches exists in the rule induction methodology applied to SVMs after extraction of the SVs. The framework of the overall experimental procedure is depicted in Fig. 2.

4.1. JNMMC dataset

The secondary data was collected from the Department of Medicine, Dr. Bhim Rao Ambedkar Memorial Hospital, Pt. J.N.M. Medical College (JNMMC), Raipur, Chhattisgarh, India. In this cross-sectional study conducted between March 2015 and October 2016, a total of 300 diabetic subjects (129 males and 171 females) between 26 and 90 years of age participated. The Institutional Ethics Committee granted ethical approval for the study project including primary data collection. Informed consent was taken from all the study subjects and the data were collected on structured questionnaire.

The JNMMC dataset consists of 34 features in all containing anthropometric, non-invasive and metabolic laboratory measurements with class label hypertension categorized as positive and negative. The fasting blood sugar (FBS) levels of all subjects were measured for the diagnosis of type-I and type-II diabetes. Diabetes was defined according to criteria of the American Association of Clinical Endocrinologists (Feld, 2002) and the 1990 World Health Organization report (Alberti & Zimmet, 1998) as subjects having FBS > 110 mg/dl. Based on their blood pressure levels, the study subjects were divided in two groups as hypertensive and non-hypertensive. A person having blood pressure (systolic BP ≥ 140 mmHg and/or diastolic BP ≥ 90 mmHg) was considered as hypertensive otherwise non-hypertensive. Overall, 108 individuals were identified as diabetics having hypertension. The description of features for hypertension dataset is given in Table 2.

4.2. Feature selection

High – dimensionality and high noisiness in the dataset leads to degradation of the prediction accuracy. Therefore, it becomes increasingly necessary to employ robust computational techniques such as feature selection and extraction to select significant features for better disease prediction. Feature Selection (FS) aims at removal of redundant and irrelevant features to improve the quality and efficiency of machine learning algorithms. Both feature subset selection and feature extraction methods reduce the computational cost and avoid overfitting of the data. Our work focuses at accurately predicting the presence of hypertension among diabetics with reduced set of attributes attained by applying various FS techniques. Therefore, we used four filter methods namely univariate Logistic Regression (LR), Chi-square, Information Gain (IG) and mRMR and one embedded method Random Forest (RF) for choosing robust informative features.

Univariate LR is a statistical method that formulates a mathematical model which shows relationship among dependent or outcome variable and each independent variable. This method is used for testing the hypotheses. Univariate logistic regression equation is given by the relationship

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m \tag{13}$$

where π indicates the probability of occurrence of hypertension among diabetics, β_i ($i = 0, 1, 2, \ldots, m$) represent the regression coefficients associated with the reference group and x_i ($i = 1, 2, \ldots, m$) are independent variables. The univariate LR analysis indicated that 8 attributes depicted statistically significant association with class diabetic hypertensives (p-value < 0.05).

In statistical theory, chi-square test (Jin, Xu, Bie, & Guo, 2006) was used to measure the independence between feature f_k and class label c_i. Chi-square evaluates the independence of feature f_k and class label c_i. The higher the value of $\chi^2(f_k, c_i)$ is, the more class label c_i is related to that feature f_k. The formula for Chi-square is defined as

$$\chi^2(t_k, c_i) = \frac{N(a_{ki}d_{ki} - b_{ki}c_{ki})^2}{(a_{ki}+b_{ki})(a_{ki}+c_{ki})(b_{ki}+d_{ki})(c_{ki}+d_{ki})} \tag{14}$$

where N is the total number of samples in the dataset. Here, a_{ki} represents the frequency with which feature f_k belongs to category c_i, b_{ki} is the frequency with which feature f_k belongs to category other than c_i, c_{ki} is the frequency with which category c_i does not contain feature f_k and d_{ki} represents the number of times neither c_i nor t_k exists. Each feature is ranked with respect to the class label based on the value of χ^2.

Information Gain (Quinlan, 1986) measures the amount of information that can be obtained for prediction of class labels c_i by

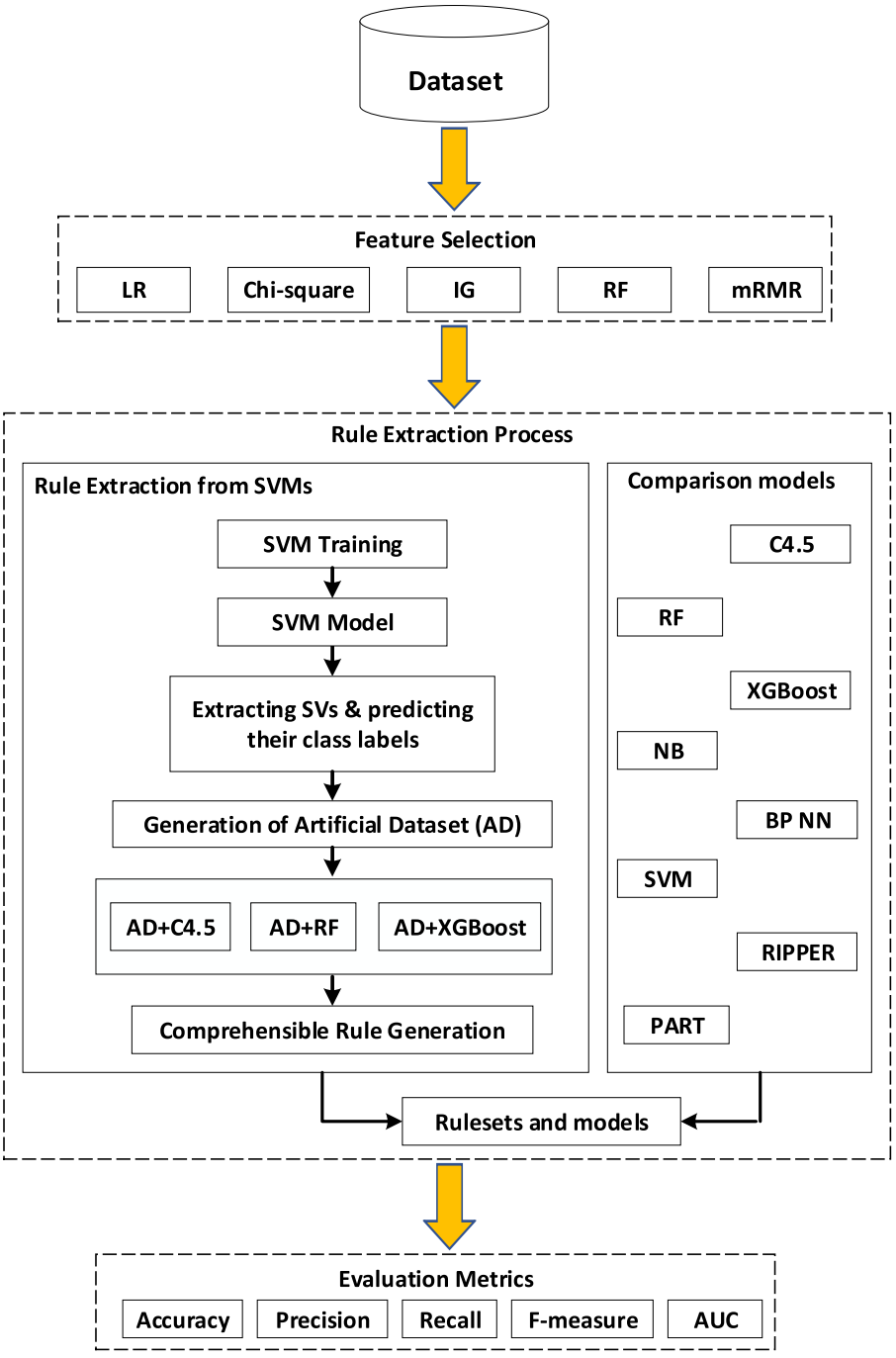

Fig. 2. Framework of the overall experimental procedure.

knowing the value of an attribute f_k. It is based on the decrease in entropy after the data is split on an attribute (Cover & Thomas, 2006; Brown, 2009). IG of an attribute f_k over the class c_i can be computed as

$$IG(f_k, c_i) = \sum_{c \in \{c_i, \bar{c}_i\}} \sum_{f \in \{f_k, \bar{f}_k\}} P(t, c) \log \frac{P(f, c)}{P(f)P(c)} \tag{15}$$

where $P(c)$ is the fraction of the samples associated with class label c_i over the total number of samples, $P(t,c)$ is the fraction of samples associated with class label c_i that contains the feature f over the total number of samples and $P(t)$ is the fraction of samples containing the feature f over the total number of samples. Eq. (15) represents the mutual information that measures the mutual dependence of the feature f and class label c_i. High IG value represents that f and c_i are closely related and 0 value depicts that both quantities are independent of each other.

The mRMR (minimum Redundancy Maximum Relevance) method selects features f_k that possess highest relevance with class label c_i and which are maximally different from each other, i.e. minimally redundant. Both optimization criteria (maximum-relevance and minimum-redundancy) use mutual information as the evaluation measure. It is calculated as

$$\max D(S, c) = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c) \tag{16}$$

$$\min R(S) = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j) \tag{17}$$

Table 2
Summary of features for JNMMC hypertension dataset.

Feature No.	Feature Description	Range
1	Age (AGE)	26–90
2	Sex (SEX)	Male, female
3	Diabetes Duration (DIABDUR)	0.1–25
4	Diabetes Duration Category (DIABCAT)	1, 2, 3, 4
5	Medication Type (MEDTYPE)	OHA, insulin, both, none
6	History of Alcohol (ALCOH)	Weekly, occasionally, none
7	Ultrasound Sonography Abdomen (USGABD)	Normal, FLG-I, FLG-II, FLG-III
8	Haemoglobin (HB)	6.7–16
9	Total Leukocyte Count (TLC)	4–29.6
10	Serum Urea (UREA)	10–148
11	Creatinine (CREAT)	0.2–7
12	Total Bilirubin (TBIL)	0.2–3.7
13	Fasting Blood Sugar (FBS)	64–354
14	Postprandial Blood Sugar (PPBS)	110–554
15	HBA1C	5.5–15.3
16	SGOT	13–434
17	SGPT	6–561
18	Alkaline Phosphatase (ALP)	46–880
19	Total protein (TPROT)	4.5–8.4
20	Albumin (ALB)	2.1–4.6
21	Total cholesterol (TC)	89–277
22	Triglyceride Level (TGL)	35–500
23	High-Density Lipoprotein (HDL)	10–70
24	Prothrombin Time (PT)	14–19
25	International Normalized Ratio (INR)	1.01–1.65
26	Weight (WEIGHT)	50–98
27	Height (HEIGHT)	149–179
28	Body Mass Index (BMI)	21.2–40
29	Body mass index category (BMICAT)	1, 2, 3
30	Diabetes Mellitus Type (DMTYPE)	Type 1, Type 2
31	Waist circumference (WC)	80–132
32	Hip circumference (HC)	96–150
33	Waist-to-hip circumference ratio (WHCR)	0.76–0.98
34	Hypertension duration (HTNDUR)	0–20
35	Hypertension (Class)	Positive, Negative

Note: OHA = Oral Hypoglycemic Agents; FLG = Fatty Liver Grade.

Random Forest as an ensemble learner is based on randomized decision trees and provides different feature importance measures. One such measure is derived from the training of the RF classifier. RF performs an implicit feature selection, the outcome of which can be visualised as Gini importance (Menze et al., 2009). It is utilized as an embedded FS method. At each node τ among the binary trees T of the random forest, the optimal split is obtained using the Gini impurity $i(\tau)$ which provides a computationally efficient approximation to the entropy and measures how well a potential split is separating the instances of both the classes in this individual node. RF ranks the importance of variables by utilising Mean Decrease Impurity (MDI) as the measure of significance, which is depicted in Fig. 3. MDI is based on the total decrease in node impurity from splitting on the attribute, averaged over all trees (Louppe, Wehenkel, Sutera, & Geurts, 2013). With $p_k = n_k/n$ representing the fraction of n_k samples from positive and negative classes over the total number of n samples at node τ, the Gini impurity $i(\tau)$ is computed as

$$i(\tau) = 1 - p_1^2 - p_0^2 \tag{18}$$

where p_1 represents the positive samples and p_0 are the negative samples in the data.

The performance of a classification algorithm highly depends on an appropriate selection of the most informative and diagnostic features. To obtain greater predictive accuracy, we selected 8 attributes out of total 34 attributes on the basis of the scores or ranks returned by the 5 feature selection algorithms. The statistically significant (p-value < 0.05) features were selected using LR and Chi-square methods. The 8 features selected for diagnosing diabetic hypertensives were haemoglobin A1c (HBA1C), triglycerides (TGL), high-density lipoproteins (HDL), ultrasound sonography abdomen (USGABD), waist circumference (WC), waist-to-hip circumference ratio (WHCR), international normalized ratio (INR) and postprandial blood sugar (PPBS). The FS algorithms were implemented using FSelector package in R, the results of which are depicted in Table 3.

4.3. Extraction and generation of rules from SVMs

The eight features or risk factors selected from the FS methods were used for training the SVMs and extracting rules from the tree-based learners such as C4.5, RF, XGBoost. Subsequently, the generated rules were used for predicting hypertensive (positive) and non-hypertensive (negative) among diabetic patients. The prevalence of negative samples in the dataset was about 64% and positive samples was about 36%. The proportion of hypertensive versus non-hypertensive was about 1:1.8. For obtaining an unbiased estimate of the overall performance of the learners, the JNMMC hypertension dataset was divided into training and testing sets with 9:1 ratio using 10-fold cross validation.

During the first run, 90% of the data was utilized for training the SVMs using the default kernel type radial basis function (RBF) by 10-fold CV. The Gaussian RBF kernel is defined as $k_\gamma(x, y) = e^{-\gamma^2\|x-y\|_2^2}$ where $\gamma > 0$ and the two control parameters to be optimized are C and γ. Here, C is the regularization parameter which defines the trade-off between the training error and the margin d and γ defines the width (Steinwart, Hush, & Scovel, 2006). Grid search was used for finding the optimal parameters, which utilized a 10-fold CV for evaluation of each parameter combination (C, γ). The optimal parameter combination which achieved the highest mean accuracy in the CV was selected for training the SVM model.

Table 3
Feature selection results for the diagnostic model.

Feature Name	LR (*p*-value)	Chi-square (*p*-value)	IG	mRMR	RF
HBA1C	0.000	17.582 (0.000)	0.101	0.060	4.705
TGL	0.040	4.690 (0.030)	0.004	0.014	1.557
HDL	0.017	5.221 (0.022)	0.030	0.034	2.250
USGABD	0.005	4.340 (0.037)	0.043	0.026	1.087
WC	0.023	3.881 (0.049)	0.056	0.029	0.875
WHCR	0.000	7.959 (0.005)	0.080	0.065	4.198
INR	0.046	3.908 (0.049)	0.011	0.002	1.279
PPBS	0.036	3.943 (0.047)	0.007	0.001	1.632

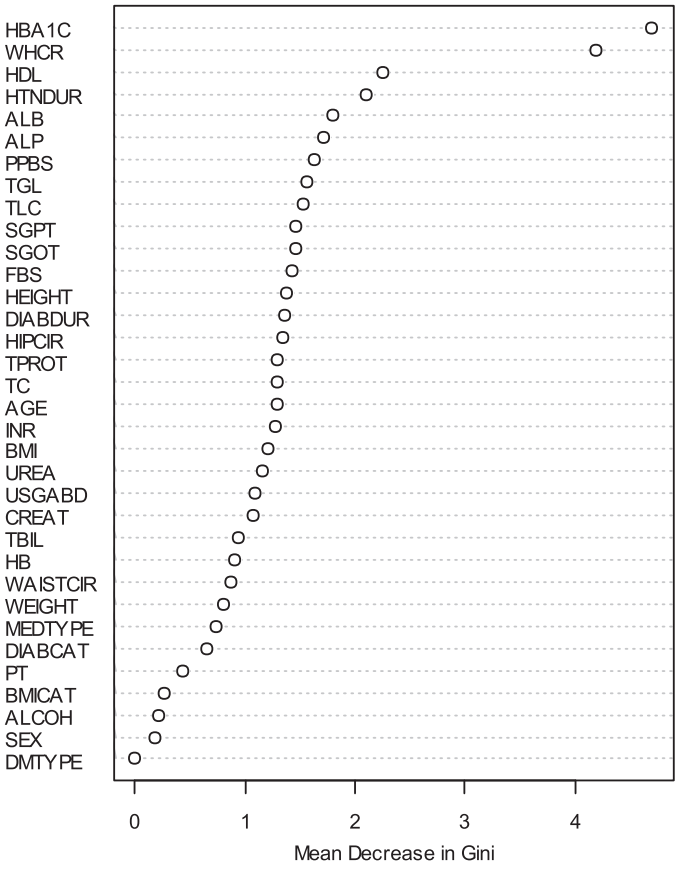

Fig. 3. Random Forest variable importance measure for prediction of hypertension among diabetic individuals in terms of mean decrease in Gini impurity.

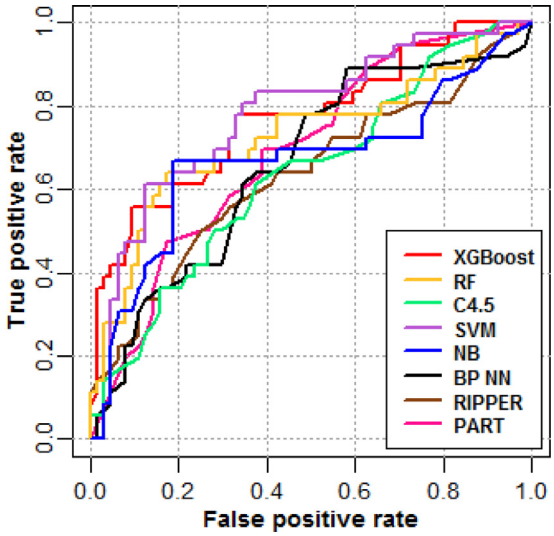

Fig. 4. Cross-validated ROC curves for positive class.

Subsequently, while performing 10-fold CV, SVM learner was built from the training data of each fold and tested on the remaining 10% dataset. To obtain a fair estimation of the performance of the trained model, nine runs were additionally conducted using the same parameters as selected in the first run. Although CV provides good results, the randomness in dividing the dataset produces different estimates between multiple runs. Furthermore, if the folds are unrepresentative, then the estimator may be biased giving pessimistic results.

One of the major motivations for extraction of rules from SVMs is to obtain rulesets which can best explain the SVM black box model. Tree-based models are integrated with SVMs for generation of 10-fold rules. In addition, XGBoost, RF, C4.5, SVM, NB, BP NN, RIPPER and PART are also implemented using 10-fold CV. The parameters of each of these classifiers are optimized in the same way using grid search in first run as previously performed in SVM. The hyper-parameters used for tuning each model with their detailed explanations are shown in Table 4. Table 5 depicts the fine-tuned hyper-parameter values for each classifier used in the experiments on JNMMC dataset. Accuracy, precision, recall, f-measure and AUC are used as performance parameters for these classifiers. Table 6 depicts the mean results of 10-fold CV and corresponding ROC curves are illustrated in Fig. 4.

In each of the 10 runs, support vectors were extracted from the trained SVM models, which were used to predict the class label of SVs. Thus, a modified artificial dataset was created consisting of the SVs and predicted class labels. This dataset was used as an input to train the tree-based rule learners for generation of symbolic rules in a comprehensible form which represented the concepts learned by the SVM model. The learners utilized were single decision tree learner C4.5 and ensemble tree learners RF and XGBoost. Each of these learners' constructs rule sets whose accuracy, precision, recall and F-measure values are shown in Table 7. The generated rule sets provide an explanation capability and represent a generalized behaviour of the SVMs. The process of construction of rulesets from C4.5, RF and XGBoost were achieved by R package "rpart", "randomForest" and "xgboost" respectively. Grid search with 10-fold CV was applied which selected best tuned values for model's parameters. The optimal parameter values obtained for SVM+C4.5, SVM+RF and SVM+XGBoost approaches were given in Table 5. The results of 10-fold CV in 10 runs were calculated for the proposed method SVM+XGBoost and compared with two approaches viz. SVM+C4.5 (eclectic method) and SVM+RF, the results of which are shown in Table 8. The performance of the proposed model, five rule induction techniques namely XGBoost, RF, C4.5, RIPPER, PART and other compared approaches such as SVM, NB and BP NN were evaluated on test set over ten runs. The rule induction method RIPPER (Cohen, 1995) implements a propositional rule learner whereas PART (Frank & Witten, 1998) utilizes the separate

Table 4
Parameters and their detailed explanations.

Parameter Name	Parameter Explanation
kernel type	radial basis function
cost (C)	regularization parameter
gamma (γ)	kernel width
ntree	number of trees to grow
mtry	number of attributes randomly sampled as candidates at each split
nodesize	minimum size of terminal nodes
minsplit	minimum number of observations that must exist in a node for a split to occur
cp	complexity parameter
maxdepth	maximum depth of any node of the final tree
eta	learning rate or shrinkage
gamma	minimum loss reduction required for further partitioning a leaf node of the tree
max_depth	maximum tree depth
min_child_weight	minimum sum of instance weight needed in a child
subsample	subsample ratio of the training observations
colsample_bytree	subsample ratio of columns when constructing each tree
nrounds	maximum number of iterations
threshold	stopping criteria
learningrate	learning rate
act.fct	activation function
NumOpt	number of optimization runs
NumFolds	amount of data used for reduced-error pruning
MinWeights	minimum total weight of the instances in a rule
threshold	confidence threshold used for pruning
pruned	whether pruning is performed

Table 5
Fine-tuned hyper-parameter values used in the experiments on JNMMC data.

S. No.	Classifiers	Parameter Name	Value
1.	SVM	kernel type	radial
		cost (C)	0.7
		gamma (γ)	0.04
		ntree	106
2.	RF	mtry	4
		nodesize	9
		minsplit	20
3.	C4.5	cp	0.001
		maxdepth	4
		eta	0.9
		gamma	1
		max_depth	4
4.	XGBoost	min_child_weight	1
		subsample	1
		colsample_bytree	1
		nrounds	200
		threshold	0.001
5.	BP NN	learningrate	0.3
		act.fct	logistic
		NumOpt	1
6.	RIPPER	NumFolds	4
		MinWeights	6
7.	PART	threshold	0.119
		pruned	yes
		minsplit	40
8.	SVM+C4.5 (C4.5 on SVs)	cp	0.001
		maxdepth	1
		ntree	26
9.	SVM+RF (RF on SVs)	mtry	4
		nodesize	1
		eta	0.6
		gamma	1
		max_depth	1
10.	SVM+XGBoost (XGBoost on SVs)	min_child_weight	1
		subsample	0.5
		colsample_bytree	1
		nrounds	200

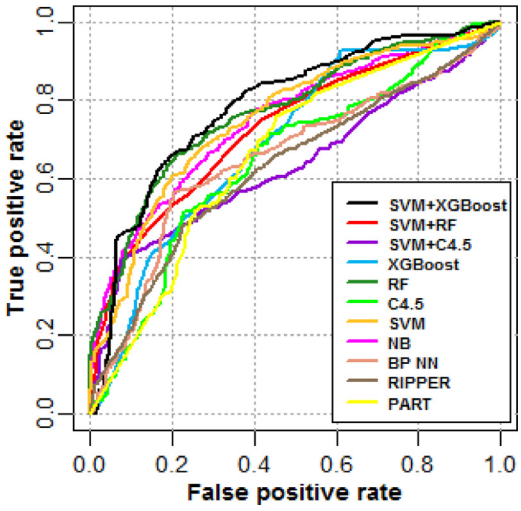

Fig. 5. Cross-validated ROC curves in 10 runs for positive class.

and conquer strategy to build the rule list. The cross-validated ROC curves in 10 runs for positive class are depicted in Fig. 5.

In addition, rules were also generated to evaluate the performance of the proposed SVM+XGBoost rule extraction algorithm. The training dataset partitioned in the 10-fold CV is utilized for generation of classification rules and these generated rules are evaluated on the test dataset to obtain performance results as shown in Table 7. The ruleset obtained from the proposed algorithm for each fold are applied to the testing data for obtaining the performance parameters. The rulesets obtained for each fold are depicted in Table 9, where the first column shows the fold number, the second column shows the rules generated for each fold, and the third and last column shows the ruleset size and average rule length (ARL) of the generated ruleset respectively. ARL is computed by dividing the total number of antecedents present in the ruleset by the total number of rules.

Table 6
Average and dispersion of tenfold cross validation results for positive class.

Classifiers	Accuracy (mean% ± std)	Precision (mean% ± std)	Recall (mean% ± std)	F-measure (mean ± std)	AUC (mean ± std)
XGBoost	70 ± 0.20	69.7 ± 0.35	62.7 ± 0.24	0.606 ± 0.26	0.809 ± 0.18
RF	72 ± 0.16	73.3 ± 0.28	**63.2 ± 0.29**	0.608 ± 0.22	0.830 ± 0.17
C4.5	64 ± 0.16	59.5 ± 0.31	44.3 ± 0.20	0.455 ± 0.20	0.683 ± 0.12
SVM	**75 ± 0.17**	**74.2 ± 0.30**	57.7 ± 0.26	**0.612 ± 0.25**	**0.841 ± 0.15**
NB	69 ± 0.13	73.3 ± 0.27	50.7 ± 0.24	0.546 ± 0.19	0.778 ± 0.18
BP NN	65 ± 0.07	51.2 ± 0.19	56.7 ± 0.22	0.507 ± 0.16	0.691 ± 0.18
RIPPER	64 ± 0.14	57.5 ± 0.28	52.5 ± 0.22	0.489 ± 0.13	0.63 ± 0.10
PART	70 ± 0.08	63.3 ± 0.29	61.9 ± 0.24	0.576 ± 0.2	0.656 ± 0.12

Table 7
Average and dispersion of tenfold cross validation results for extracted rules.

Classifiers	Accuracy (mean% ± std)	Precision (mean% ± std)	Recall (mean% ± std)	F-measure (mean ± std)
SVM+XGBoost	**69 ± 0.09**	**75.0 ± 0.33**	38 ± 0.18	0.488 ± 0.20
SVM+RF	68 ± 0.10	71.7 ± 0.25	**45 ± 0.23**	**0.493 ± 0.12**
SVM+C4.5	66 ± 0.19	53.3 ± 0.41	31.7 ± 0.26	0.391 ± 0.31

Table 8
Average and dispersion of test data in 10 runs for positive class with statistical significance values obtained using *t*-test.

Classifiers	Accuracy (mean% ± std) [*p*-value]	Precision (mean% ± std)	Recall (mean% ± std)	F-measure (mean ± std)	AUC (mean ± std) [*p*-value]
SVM+XGBoost	**74.7 ± 0.12** –	**75.3 ± 0.35**	42.1 ± 0.28	0.498 ± 0.26	**0.810 ± 0.15** –
SVM+RF	72.9 ± 0.15 0.0001[a]	72.6 ± 0.34	38.5 ± 0.30	0.452 ± 0.30	0.735 ± 0.19 0.0022[a]
SVM+C4.5	73.2 ± 0.14 0.0001[a]	68.2 ± 0.39	38.3 ± 0.27	0.454 ± 0.27	0.655 ± 0.14 0.0001[a]
XGBoost	63.4 ± 0.14 0.0001[a]	49.0 ± 0.29	54.3 ± 0.32	0.466 ± 0.24	0.709 ± 0.18 0.0001[a]
RF	73.2 ± 0.14 0.0001[a]	64.4 ± 0.32	**57.7 ± 0.30**	**0.573 ± 0.26**	0.761 ± 0.19 0.0443[a]
C4.5	67.6 ± 0.15 0.0001[a]	56.4 ± 0.33	50.8 ± 0.28	0.493 ± 0.25	0.677 ± 0.16 0.0001[a]
SVM	72.8 ± 0.13 0.0001[a]	68.6 ± 0.30	51.4 ± 0.26	0.552 ± 0.23	0.784 ± 0.15 0.2218[b]
NB	70.7 ± 0.14 0.0001[a]	67.8 ± 0.36	49.8 ± 0.31	0.533 ± 0.29	0.786 ± 0.17 0.2911[b]
BP NN	69.2 ± 0.14 0.0001[a]	60.8 ± 0.28	56.6 ± 0.28	0.550 ± 0.23	0.754 ± 0.18 0.0178[a]
RIPPER	67.9 ± 0.14	55.3 ± 0.29	53.5 ± 0.31	0.495 ± 0.23	0.649 ± 0.16
PART	65.8 ± 0.14	52.9 ± 0.30	50.2 ± 0.29	0.478 ± 0.24	0.681 ± 0.17

[a] denotes statistically significant upgradation.
[b] denotes statistically significant degradation.

4.4. Interestingness measures

Interestingness measures play an important role in data mining, regardless of the kind of patterns being mined (Geng & Hamilton, 2006). It is a metric to determine the strength of a rule where an appropriate interestingness measure for a dataset can be chosen depending on the domain application. Probability based objective measures such as support and confidence evaluate the generality and reliability of the obtained classification rules. To evaluate the interestingness of the rules generated from the proposed SVM + XGBoost algorithm, quantitative measures such as support and confidence are adopted.

A classification rule is an implication of the form $X \rightarrow Y$, where X is the antecedent, Y is the consequent and $X \cap Y = \varphi$. X contains the predictor variables or risk factors that are related to outcome variable Y, which represents the occurrence of hypertension. Support is a significant measure because a rule having very low support may occur by chance. Therefore, support is utilized to eliminate uninteresting rules, while confidence is used to measure the precision of extracted rules. A given classification rule $X \rightarrow Y$ is deemed significant if it has high support and high confidence. Support value of 0.01 indicates that the given rule describes 1% of the dataset. Support determines the percentage of records in the training dataset that contain both X and Y, while confidence determines the percentage of records containing X that also contain Y. The formal definitions of these metrics are

$$\text{Support, } s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \tag{19}$$

$$\text{Confidence, } c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \tag{20}$$

Confidence ranges from 0 to 1. It is an estimate of $Pr(Y|X)$, the probability of observing Y given X. After obtaining the rulesets as shown in Table 9, support and confidence are used as the basis to understand the strong co-occurrence relationship between antecedent and consequent of the rule. For instance, in Table 9, the rule (HBA1C <= 9.45 AND TGL <= 205 AND INR <= 1.285 → Negative) is the most frequent rule for predicting negative class, because it has appeared in six of the tenfold rulesets. Each of the six rules possess high support and high confidence values. Glycated haemoglobin (HBA1C) is considered as a significant biomarker for the presence and severity of diabetes, which

Table 9
SVM + XGBoost rulesets generated using 10-fold CV.

Fold	Rule set	#Rules	ARL	Support	Confidence
1	If HBA1C <= 9.45 AND TGL <= 205 AND INR <= 1.285Then Class = Negative;	6	2	0.511	0.754
	If HBA1C > 9.45 Then Class = Positive;			0.144	0.813
	If HBA1C <= 9.45 AND WHCR <= 0.915Then Class = Negative;			0.422	0.844
	If HBA1C > 7.4 AND TGL > 204 AND WHCR > 0.905Then Class = Positive;			0.089	1
	If HBA1C <= 9.45 AND INR > 1.18 AND WHCR > 0.915Then Class = Positive;			0.033	1
	Else Class = Negative; (Default class)				
2	If HBA1C <= 9.45 AND TGL <= 205 AND INR <= 1.285Then Class = Negative;	7	1.7	0.533	0.762
	If HBA1C > 9.45 AND TGL > 77.5 Then Class = Positive;			0.111	0.91
	If TGL > 205 AND WHCR > 0.905 Then Class = Positive;			0.089	0.889
	If HDL <= 30 AND WHCR <= 0.905 Then Class = Negative;			0.211	0.864
	If INR > 1.285 Then Class = Positive;			0.033	0.75
	If PPBS <= 429 AND HBA1C > 9.45 Then Class = Positive;			0.089	0.8
	Else Class = Negative; (Default class)				
3	If HBA1C <= 10.2 AND TGL <= 202.5 AND INR <= 1.285 Then Class = Negative;	6	1.83	0.544	0.778
	If HBA1C > 10.2 Then Class = Positive;			0.122	0.786
	If HBA1C <= 10.2 AND WHCR <= 0.915Then Class = Negative;			0.478	0.843
	If TGL > 195.5 AND INR > 1.105 AND WHCR > 0.915Then Class = Positive;			0.111	1
	If HBA1C <= 7.95 AND INR <= 1.145 Then Class = Negative;			0.467	0.778
	Else Class = Positive; (Default class)				
4	If HBA1C <= 9.45 AND TGL <= 205 AND INR <= 1.285 Then Class = Negative;	5	1.6	0.567	0.797
	If HBA1C > 9.45 Then Class = Positive;			0.144	0.867
	If TGL > 205 AND WHCR > 0.915 Then Class = Positive;			0.1	1
	If TGL > 126.5 AND WHCR <= 0.915 Then Class = Negative;			0.3	0.871
	Else Class = Positive; (Default class)				
5	If HBA1C <= 9.45 AND TGL <= 205 AND INR <= 1.285 Then Class = Negative;	6	1.67	0.544	0.77
	If HBA1C > 9.45 Then Class = Positive;			0.122	0.846
	If HBA1C <= 9.45 AND WHCR <= 0.915Then Class = Negative;			0.455	0.837
	If HBA1C > 6.69 AND TGL > 204 AND WHCR > 0.895 Then Class = Positive;			0.1	1
	If HbA1C <= 6.69 Then Class = Negative;			0.33	0.91
	Else Class = Positive; (Default class)				
6	If HBA1C <= 9.3 AND TGL <= 205 AND INR <= 1.285 Then Class = Negative;	4	1.75	0.555	0.81
	If HBA1C > 9.3 Then Class = Positive;			0.122	0.786
	If HBA1C <= 8.95 AND HDL > 16.5 AND INR <= 1.225 Then Class = Negative;			0.533	0.814
	Else Class = Positive; (Default class)				
7	If HBA1C <= 9.45 AND TGL <= 205 AND INR <= 1.285 Then Class = Negative;	5	2	0.544	0.78
	If HBA1C > 9.45 Then Class = Positive;			0.144	0.867
	If PPBS <= 265.5 AND HBA1C <= 9.45 AND TGL > 205 AND INR > 1.1 Then Class = Positive;			0.044	0.667
	If HBA1C <= 7.95 AND INR <= 1.115Then Class = Negative;			0.355	0.865
	Else Class = Positive; (Default class)				
8	If HBA1C <= 8.4 AND TGL <= 217.5Then Class = Negative;	6	2	0.489	0.786
	If HBA1C > 8.95 AND HDL <= 35.5 Then Class = Positive;			0.155	0.824
	If HBA1C <= 8.95 AND WHCR <= 0.915Then Class = Negative;			0.478	0.878
	If HBA1C <= 8.95 AND TGL > 217.5 AND WHCR > 0.915 Then Class = Positive;			0.033	0.75
	If HBA1C <= 9.2 AND TGL <= 205 AND HDL > 23.5Then Class = Negative;			0.4	0.857
	Else Class = Positive; (Default class)				
9	If HBA1C <= 9.45 AND TGL <= 205 AND INR <= 1.285 Then Class = Negative;	4	1.5	0.533	0.75
	If HBA1C > 9.45 Then Class = Positive;			0.122	0.786
	If HBA1C <= 8.95 AND INR <= 1.225Then Class = Negative;			0.556	0.769
	Else Class = Positive; (Default class)				
10	If HBA1C <= 9.45 AND TGL <= 284.5 AND INR <= 1.285 AND WHCR <= 0.945 Then Class = Negative;	7	2	0.589	0.779
	If HBA1C > 9.45 AND TGL > 99.5 Then Class = Positive;			0.122	1
	If TGL > 215.5 AND INR > 1.105 AND WHCR > 0.905Then Class = Positive;			0.078	1
	If WHCR <= 0.905 Then Class = Negative;			0.456	0.911
	If HBA1C > 9.45 AND TGL <= 99.5 Then Class = Negative;			0.033	1
	If INR > 1.185 AND WHCR > 0.905 Then Class = Positive;			0.067	1
	Else Class = Negative; (Default class)				
Mean		5.6	1.805		

acts as a risk factor for various complications including hypertension. Elevated HBA1C is the consequence of an underlying disease state known as hyperglycaemia. The JNMMC dataset used in our study consists of all diabetic patients among which normotensives or hypertensives are diagnosed. Due to this, HBA1C variable appears more frequently in each fold in Table 9. Moreover, high correlation of HBA1C with TGL helps in prediction of diabetes status. Similarly, INR is one of the risk factors for identification of T2DM. Thus, all the three biomarkers viz. HBA1C, TGL and INR can be used for screening or diagnosis of diabetes.

Another most common rule (HBA1C > 9.45 → Positive) occurs individually as well as in association with TGL and PPBS in other folds for predicting positive class. Although HBA1C remains the gold standard for diabetes diagnosis, PPBS is also a better predictor of overall glycaemic control in absence of HBA1C. In addition, there exists a strong relation between PPBS and development of diabetes complications (Ketema & Kibret, 2015). Sometimes, HBA1C variable alone is inadequate for diagnosing hypertension and more accurate diagnosis requires confirmation from other variables. Thus, extracted rules combined with several biomarkers can more precisely

Table 10
Benchmark diabetic datasets.

No.	Datasets	Missing values	# of instances	# of attributes	# of classes	Reference
1.	PID	No	768	8	2	Dheeru and Karra Taniskidou (2017)
2.	Messidor	No	1151	19	2	Dheeru and Karra Taniskidou (2017)
3.	BDD	Yes	403	19	2	Biostat Diabetes Dataset (2018)
4.	T1DD	Yes	306	21	2	Asaduzzaman et al. (2018)
5.	PAFR	Yes	95	68	2	Kurano et al. (2018)

Table 11
Accuracy assessment on selected benchmark and JNMMC diabetes dataset.

Approaches	Mean (%) ± Standard Deviation					
	PID	Messidor	BDD	T1DD	PAFR	JNMMC
SVM+XGBoost	**79.2 ± 0.04**	**73.6 ± 0.05**	**93.5 ± 0.04**	96.6 ± 0.04	**94.9 ± 0.12**	**74.7 ± 0.12**
SVM + RF	76 ± 0.05	70.9 ± 0.04	91.7 ± 0.04	97.3 ± 0.03	84 ± 0.14	72.9 ± 0.15
SVM + C4.5	75 ± 0.04	66.8 ± 0.04	92.2 ± 0.04	93.2 ± 0.05	82.8 ± 0.18	73.2 ± 0.14
XGBoost	74.7 ± 0.05	63.1 ± 0.05	90.5 ± 0.03	95.1 ± 0.03	94.7 ± 0.07	63.4 ± 0.14
RF	76.8 ± 0.05	70 ± 0.04	91.7 ± 0.07	**98.1 ± 0.02**	93.4 ± 0.08	73.2 ± 0.14
C4.5	74.5 ± 0.05	65.8 ± 0.04	91.3 ± 0.04	95.3 ± 0.04	94.4 ± 0.06	67.6 ± 0.15
SVM	76.5 ± 0.04	70.4 ± 0.04	91.8 ± 0.05	96.9 ± 0.03	76.3 ± 0.13	72.8 ± 0.13
NB	75.4 ± 0.05	56.8 ± 0.03	90.7 ± 0.04	97.7 ± 0.03	85.4 ± 0.12	70.7 ± 0.14
BP NN	75.9 ± 0.04	72.2 ± 0.04	89 ± 0.05	97.3 ± 0.03	72.3 ± 0.17	69.2 ± 0.14
RIPPER	74 ± 0.06	62.6 ± 0.05	90.8 ± 0.04	96 ± 0.04	93.4 ± 0.08	67.9 ± 0.14
PART	73.7 ± 0.05	63.9 ± 0.05	88.4 ± 0.05	96.4 ± 0.04	92.4 ± 0.07	65.8 ± 0.14

Table 12
AUC assessment on selected benchmark and JNMMC diabetes datasets.

Approaches	Mean ± Standard Deviation					
	Pima	Messidor	BDD	T1DD	PAFR	JNMMC
SVM + XGBoost	**0.866 ± 0.05**	**0.808 ± 0.05**	**0.948 ± 0.11**	0.998 ± 0.01	**0.981 ± 0.10**	**0.810 ± 0.15**
SVM + RF	0.775 ± 0.05	0.783 ± 0.05	0.829 ± 0.12	0.998 ± 0.01	0.907 ± 0.12	0.735 ± 0.19
SVM + C4.5	0.719 ± 0.06	0.695 ± 0.05	0.804 ± 0.11	0.94 ± 0.04	0.856 ± 0.17	0.655 ± 0.14
XGBoost	0.793 ± 0.05	0.681 ± 0.05	0.904 ± 0.07	0.978 ± 0.03	0.973 ± 0.05	0.709 ± 0.18
RF	0.829 ± 0.04	0.771 ± 0.04	0.922 ± 0.07	**0.999 ± 0**	0.970 ± 0.04	0.761 ± 0.19
C4.5	0.778 ± 0.06	0.669 ± 0.05	0.886 ± 0.10	0.97 ± 0.04	0.948 ± 0.06	0.677 ± 0.16
SVM	0.831 ± 0.04	0.774 ± 0.04	0.901 ± 0.09	0.998 ± 0	0.833 ± 0.13	0.784 ± 0.15
NB	0.814 ± 0.05	0.636 ± 0.05	0.908 ± 0.07	0.997 ± 0.01	0.915 ± 0.12	0.786 ± 0.17
BP NN	0.826 ± 0.05	0.789 ± 0.04	0.882 ± 0.09	0.998 ± 0	0.818 ± 0.16	0.754 ± 0.18
RIPPER	0.708 ± 0.06	0.63 ± 0.05	0.841 ± 0.10	0.963 ± 0.04	0.933 ± 0.09	0.649 ± 0.16
PART	0.785 ± 0.05	0.675 ± 0.05	0.803 ± 0.15	0.969 ± 0.03	0.926 ± 0.09	0.681 ± 0.17

identify those at the high risk of developing diabetes and its progression to hypertension.

Among the various anthropometric indices utilized, WHCR is a strong risk factor superior in predicting hypertension. The determination of appropriate WHCR cut-off level is important in prevention and treatment of hypertension. The inference rule (WHCR > 0.895 → Positive) obtained from Table 9 is in line with the threshold recommended by WHO (2011). Analogously, the inference rule (TGL > 195.5 → Positive) along with the presence of other concomitant risk factors potentiates the overall risk of hypertension. Overall, through the obtained rulesets meaningful insights and inferences could be derived which can aid the medical experts in diagnosing the disease.

4.5. Comparison on other diabetes datasets

For extended evaluation of the proposed approach, the classification methodology has been applied on other different and independent datasets. The comparison of the classification accuracy and AUC of the proposed model (i.e., SVM + XGBoost) with the other ten approaches are shown in Tables 11 and 12, respectively. Table 10 summarizes the characteristics of the datasets including the number of instances, attributes and classes. For each dataset, the missing values present in the numeric and nominal attributes have been replaced by mean and mode respectively. The description of the selected datasets is as follows:

Pima Indian diabetes (PID). This dataset is used in most of the data mining techniques for analysing the algorithmic results. It concerns with the presence or absence of diabetes among Pima-Indian women (at least 21 years old) living near Phoenix, Arizona. The set contains cases belonging to two classes (labelled as "test positive" and test negative") and described by 8 attributes: (1) number of times pregnant, (2) plasma glucose concentration, a 2 h oral glucose tolerance test, (3) diastolic blood pressure, (4) triceps skin fold thickness, (5) 2-h serum insulin, (6) BMI, (7) diabetes pedigree function, and (8) age.

Diabetic retinopathy (Messidor). This dataset contains features extracted from the Messidor image database for diabetic retinopathy (DR) screening. The features extracted from the output of various retinal image processing algorithms contains image-level, lesion specific and anatomical components. The goal is to discriminate between samples that contain signs versus no signs of DR. Based on the scores or ranks returned by the 5 different feature selection algorithms as given in Section 4.2, we selected 8 attributes out of total 19. The selected subset of features consists of the binary result of pre-screening, results of MA detection and binary

result of AM/FM-based classification. The eight attributes in each sample were found to differ significantly between the DR and no-DR samples.

Biostat diabetes dataset (BDD). The BDD dataset contains subjects screened for diabetes. A positive diagnosis of diabetes is taken if glycosylated haemoglobin > 7.0. Therefore, variable glyhb is considered as the class label. The main focus of the data is to predict the presence or absence of diabetes. Among the 19 attributes, a total of 13 are selected for use in problem space. These attributes are determined as total cholesterol (chol), stabilized glucose (stab.glu), high density lipoprotein (hdl), cholesterol/hdl ratio (ratio), age, gender, weight, body_frame, first systolic blood pressure (bp.1 s), first diastolic blood pressure (bp.1d), waist, hip and postprandial time (time.ppn).

Type I diabetes dataset (T1DD). The dataset comprises of significant risk factors for Type I diabetes among Bangladeshi residents. The entire dataset consisting of 152 cases and 154 controls was collected based on a specific questionnaire, which included various factors. In order to determine the essential risk factors of Type I diabetes, the aforementioned five feature selection approaches selected 14 out of total 19 attributes. These are detailed as age, sex, area of residence, HbA1c, height, weight, other disease, adequate nutrition, education of mother, standardized growth-rate in infancy, hypoglycaemia, pancreatic disease affected in child, family history affected in Type 1 Diabetes and Type 2 Diabetes. The objective of this dataset is to discriminate healthy subjects from those with T1DM.

PAFR research dataset. The dataset consists of clinical and gene data of 95 patients with T2DM. The data characterizes about the association of mRNA expression of platelet activating factor receptor (PAFR) in peripheral blood mononuclear cells with albuminuria and vascular dysfunction among patients with T2DM. The feature selection techniques identified 19 features to be significantly associated with the outcome variable. These are summarized as: systolic blood pressure (SBP), albumin to creatinine ratio (ACR) first, low density lipoprotein cholesterol (LDLC), Creatinine (cr), estimated glomerular filtration rate (eGFR), plasma monocyte chemoattractant protein-1 level (p-MCP1), urinary monocyte chemoattractant protein-1 to creatinine ratio (u-MCP1/Cr), flow mediated dilation (FMD), metformin, calcium channel blocker (CCB), ARB/ACEI, TNF, CD36, TLR4, p67phox, PAFR, past smoking, cardiovascular disease and cerebrovascular disease. For ease of binary classification, the output class is divided into two groups: normoalbuminuria group (42 subjects), and elevated albuminuria group (53 subjects) created by combining microalbuminuria and macroalbuminuria groups. The task is to differentiate between the presence and absence of diabetic nephropathy (renal dysfunction).

4.6. Statistical comparison of classifiers

4.6.1. Parametric test

An unpaired *t*-test was conducted between SVM+XGBoost and other algorithms for accuracy and AUC, to evaluate the statistical significance of outperforming SVM+XGBoost algorithm. The *p*-values obtained from the *t*-test are demonstrated in the Table 8 with superscript ([a]) denotes that SVM+XGBoost approach is better than the compared method whereas superscript ([b]) denotes that the SVM+XGBoost approach is inferior than the compared method. The significance level is shown by a *p*-value of less than 0.05 i.e. a confidence level of 95%.

4.6.2. Non-parametric test

According to the work proposed by (Demšar, 2006), the rank obtained by Friedman test (Friedman, 1940) was utilized to compare the accuracy and AUCs of the various classifiers. This non-parametric test is conducted to check whether the differences in the obtained performance results are due to chance. The Friedman statistic depends on the average ranked (AR) performances of the eleven classification approaches on each of the six independent datasets and is computed as follows:

$$\chi_F^2 = \frac{12N}{k(k+1)}\left[\sum_{j=1}^{k} AR_j^2 - \frac{k(k+1)^2}{4}\right], \text{ where } AR_j = \frac{1}{N}\sum_{i=1}^{N} r_i^j \tag{21}$$

In (21), N denotes the number of datasets, k is the total number of classifiers and r_i^j is the rank of the classifier j on the dataset i. Under the null hypothesis where no significant difference exists, the Friedman statistic is distributed according to Chi-square distribution χ_F^2 with k-1 degrees of freedom. Table 13 illustrates the ranking results obtained for different algorithms. From the analysis of the table, it is revealed that using any of the evaluation measures (Accuracy and AUC), the proposed SVM + XGBoost approach outperforms in comparison to the other approaches. If the null hypothesis is rejected by the Friedman test, we proceed by performing the post hoc Bonferroni–Dunn test (Dunn, 1961) to compare all classifiers with the best performing classifier.

The null hypothesis that there is no difference between the algorithms can be rejected if the χ_F^2 value is very large. The χ_F^2 value 25.42 for accuracy ranks with 10 degrees of freedom demonstrates that the *p*-value computed by Friedman test is 0.0045. This indicates that the null hypothesis of the equivalence of performance of classifiers for accuracy is rejected. Similarly, the χ_F^2 value 33.67 for AUC ranks with 10 degrees of freedom demonstrates that the *p*-value computed by Friedman test is 0.0002. As consequence, the null hypothesis of equivalence of performance of classifiers for AUC is also rejected. In Tables 14 and 15, the z statistic is applied for comparing the rank of best classifier R_0 and the rank of j th classifier R_j where SE denotes the standard error of $R_0 - R_j$. Here, p denotes the unadjusted significance level and p_{Bonf} is the adjusted level of significance using Bonferroni-Dunn method.

Since, the null hypotheses for each of the case of accuracy and AUC is rejected, therefore the post hoc Bonferroni test is applied to compare the performance of best classifier SVM+XGBoost with

Table 13
Ranking accuracy and AUC of algorithms using Friedman rank test.

Algorithm	Average Rank (Accuracy)	Average Rank (AUC)
SVM + XGBoost	**1.83**	**1.42**
SVM + RF	4.5	6.42
SVM + C4.5	6.25	9.33
XGBoost	8.17	5.5
RF	3.25	3
C4.5	7.17	7.33
SVM	5	4.58
NB	6.67	5.5
BP NN	6.42	5.42
RIPPER	7.92	9.33
PART	8.83	8.17

Table 14
Friedman and post hoc statistical analysis for accuracy.

j	Algorithm	$z = (R_0 - R_j)/SE$	p	p_{Bonf}
10	PART	3.655631	0.000257	0.002566
9	XGBoost	3.307475	0.000941	0.009414
8	RIPPER	3.176917	0.001488	0.014885
7	C4.5	2.785242	0.005349	0.053488
6	NB	2.524126	0.011599	0.115986
5	BP NN	2.393568	0.016685	0.166854
4	SVM+C4.5	2.306529	0.021081	0.210811
3	SVM	1.653738	0.098181	0.981808
2	SVM+RF	1.392621	0.163734	1.637344
1	RF	0.73983	0.459403	4.594031

Table 15
Friedman and post hoc statistical analysis for AUC.

j	Algorithm	$z=(R_0-R_j)/SE$	p	p_{Bonf}
10	SVM+C4.5	4.134344	0.000036	0.000356
9	RIPPER	4.134344	0.000036	0.000356
8	PART	3.525073	0.000423	0.004234
7	C4.5	3.089878	0.002002	0.020024
6	SVM+RF	2.611165	0.009023	0.090234
5	XGBoost	2.132451	0.03297	0.329698
4	NB	2.132451	0.03297	0.329698
3	BP	2.088932	0.036714	0.036714
2	SVM	1.653738	0.098181	0.981808
1	RF	0.826869	0.408311	4.083114

the other classifiers. The post hoc statistical analysis has been performed in Table 14 for accuracy and in Table 15 for AUC.

In view of the above analysis, it is evident that the proposed algorithm significantly outperformed the other ten benchmark algorithms. In other words, it could be statistically inferred that the SVM + XGBoost methodology is able to significantly improve the classification performance as compared to the other techniques. Thus, the proposed approach provides a new insight into better diagnosis and prediction of hypertension among diabetics.

5. Results and discussions

This study primarily aims at detecting probable diabetic individuals with undiagnosed hypertension. For identifying hypertension, the quality of proposed model is assessed using five evaluation parameters namely accuracy, precision, recall, F-measure and AUC. Among all, high priority is given to accuracy that determines the proportion of true results over the total observed cases and AUC which compares the diagnostic performance of the proposed model with the other models. The average results of proposed SVM + XGBoost model on test data in 10 runs are depicted in Table 8. In the JNMMC dataset, the proposed model produced 74.7% accuracy, 75.3% precision and AUC value 0.810 for positive class. Further, it can be derived from Table 6 that the proposed model outperforms the other approaches.

The average results of tenfold cross validation for positive i.e. hypertensive class are shown in Table 6. It depicts that SVM yields highest accuracy (75%), precision (74.2%), F-measure (0.612) and AUC (0.841) among the six methods, which demonstrates superior performance of SVM over the rule induction methods. This evidence provides the motivation for extraction of rules from SVMs. The proposed rule-extraction technique SVM + XGBoost generates understandable rule sets, as shown in Table 8 with better accuracy, precision and AUC in comparison with seven other rule induction approaches such as SVM+RF, SVM+C4.5, XGBoost, RF, C4.5, RIPPER and PART. Although, the recall values of the proposed method are only higher than two approaches namely SVM + C4.5 and SVM + RF and F-measure values are lower than RF, SVM, NB and BP NN, it yields maximum accuracy, precision and AUC values among total eleven models. When compared with SVM, SVM + XGBoost exhibits better accuracy, precision and AUC values, which enhances the generalization performance of SVMs. Overall, the proposed SVM + XGBoost method exploits the learning capabilities of SVM black-box model and interpretability of XGBoost tree ensemble model.

The rule sets generated by SVM + XGBoost approach in each fold of tenfold CV are represented in Table 9. In addition, Table 9 shows the total number of rules and average rule length (ARL) produced by each fold with the proposed method obtaining a mean rule set size of 5.6 and a mean ARL of 1.981. Each fold consists of certain set of rules with the default rule specified at the end of each set. The list of rules together with the *default-class* is represented as: $\langle \mathcal{R}_1, \ \mathcal{R}_2, \ldots, \ \mathcal{R}_k, \ default-class \rangle$, where $\mathcal{R}_i$ is the rule mined from the training data. If the test sample to be predicted does not belong to any of the rules specified in the rule-list $(\mathcal{R}_1, \ \mathcal{R}_2, \ldots, \ \mathcal{R}_k)$ of each fold, then it is classified by the *default-class*. The *default-class* is usually the majority class among all the classes in the training data, which is utilized only when no rule is learned from the training data.

The tenfold cross validated rules extracted from SVM + XGBoost produce better accuracy (69%) and precision (75%) when compared with SVM + RF and SVM + C4.5 approaches, as shown in Table 7. This proves that SVM+XGBoost model produces more accurate and good quality rule sets which enhances diagnostic capability and reduces the complexity of proposed approach. The rules extracted from RF show highest recall and F-measure values of 45% and 0.493 respectively when compared with XGBoost and C4.5 rule induction techniques. Through Table 8, SVM+XGBoost achieves much higher accuracy, precision and AUC than the individual XGBoost model. Thus, as XGBoost implements extreme gradient boosting ensemble technique, it proves to perform better than single tree model such as C4.5. Since RF is also based on bagging ensemble technique, it provides better results than C4.5, XGBoost and SVM+RF. The rules learned by XGBoost are generated from the SVs, which not only reduce the size of rules in terms of antecedents but also time and complexity. The generated comprehensible rulesets can aid clinicians in gaining more interpretability to the diagnostic results. Overall, the generated rule sets are more transparent and understandable than larger rule sets that are complicated and ambiguous. In this way, SVM+XGBoost has emerged as a technique superior to other algorithms in providing lucid and legitimate rules.

In order to validate the generality of the proposed method, we have compared it with five more diabetes-related databases obtained from different sources. The simulation results obtained from Tables 11 and 12 prove the effectiveness of the proposed approach. The results indicate that SVM + XGBoost outperforms in five among six datasets for both accuracy and AUC evaluation measures. One of the important results recognized by this paper is that factors namely age, gender, HbA1c, PPBS, DBP, SBP, waist circumference, hip circumference, HDL and weight are more effective in identifying the risk of diabetes and other associated microvascular and macrovascular complications.

6. Conclusions and future work

In this paper, we developed a hybrid system for diagnosing hypertension among diabetic individuals. Specifically, SVMs have been used for identifying hypertension in patients with diabetes for which a rule-based explanation module was added to produce intelligible and understandable rules. Such SVM extracted rules can be considered as a second opinion for diagnosing hypertension and other complications related to diabetes. These rules can also be implemented as a tool for predicting hypertension in individuals suffering from high risk of diabetes by an amateur user. The medical experts recognize the rules generated by the system to be very beneficial for outpatient screening where basic measurements can help to identify the severity of diabetes and its complications. The incidence of diabetes and related risks can be alleviated if timely, proper and appropriate interventions take place. Else it may lead to severe consequences which can be potentially life-threatening.

For evaluation, the proposed model is applied to the new diabetic hypertensive dataset (JNMMC) and five more diabetes-related datasets. The results depict that on JNMMC dataset, our proposed model i.e. SVM + XGBoost produces high quality accurate and precise rules that determines the diagnostic ability of the proposed model. Also, the proposed technique outperforms other such hybrid approaches which utilize SVM with rule-based component for

generation of comprehensible rules. It performs better than XGBoost and RF ensemble tree learners and C4.5 single tree model. Moreover, the proposed model also shows superiority over SVM and BPNN black box algorithms, RIPPER and PART rule-based approaches and NB probabilistic classifier. The proposed method outperforms the ten other methods on five out of six datasets. This underlines that the proposed approach gives statistically better results than the two hybrid and eight benchmark methods. The significant improvement is validated by both parametric and non-parametric statistical test.

One of the possible extensions of this work is pruning of the rule sets obtained from the proposed work, so that less number of small size rules are generated without compromising the diagnostic accuracy. Further, the small size of JNMMC data is a serious limitation of this study that may adversely affect the performance of the proposed model and reliability of the diagnostic rules generated from it. An in-depth analysis conducted with large amount of data can boost the accuracy of the proposed algorithm.

Apart from this, ensemble approaches which combine multiple learners can be utilized for classification. Another extension to the work would be to apply these methods on diverse disease datasets which possess a wider variety of class distributions. Moreover, the real-world medical datasets usually suffer from class-imbalance problems since they are composed of a large number of normal samples with only a small proportion of abnormal ones. As the proposed approach did not deal with the class imbalance problem of the datasets, the obtained results can be improved by handling this issue. Additionally, based on the outcomes of this study, a more advanced risk score can be developed for early identification of complications related to diabetes which can greatly reduce the healthcare costs.

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Credit author statement

1. Namrata Singh (1st and corresponding author): Substantially contributed to the conception, design and development of the algorithm and experimental work. **2. Pradeep Singh (2nd author):** Supervised the research, contributed to the manuscript and revised it critically for important results and conclusions. **3. Deepika Bhagat (3rd author):** Acquired the diabetic-hypertensive medical dataset (JNMMC) from Pt. J N M Medical College and associated Dr. Bhimrao Ambedkar Hospital, Raipur, Chhattisgarh, India. She performed the analysis and interpretation of medical data. **All the three authors:** Finally approved the revised version of the manuscript to be published.

References

Alberti, K. G. M. M., & Zimmet, P. Z. (1998). Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus. Provisional report of a WHO consultation. *Diabetic Medicine, 15*(7), 539–553 Retrieved from. doi:10.1002/(SICI)1096-9136(199807)15:7\%3C539::AID-DIA668\%3E3.0.CO;2-S.

Andrews, R., Diederich, J., & Tickle, A. B. (1995). Survey and critique of techniques for extracting rules from trained artificial neural networks. *Knowledge-Based Systems, 8*(6), 373–389 Retrieved from. doi:10.1016/0950-7051(96)81920-4.

Asaduzzaman, S., Masud, F. Al, Bhuiyan, T., Ahmed, K., Paul, B. K., & Rahman, S. A. M. M. (2018). Dataset on significant risk factors for Type 1 diabetes: A Bangladeshi perspective. *Data in Brief, 21*, 700–708. doi:10.1016/j.dib.2018.10.018.

Barakat, N., & Bradley, A. P. (2006). Rule extraction from support vector machines: Measuring the explanation capability using the area under the ROC curve. In *Proceedings of the eighteenth international conference on pattern recognition (ICPR): 2* (pp. 812–815). IEEE. Retrieved from. doi:10.1109/ICPR.2006.1021.

Barakat, N., & Diederich, J. (2005). Eclectic rule-extraction from support vector machines. *International Journal of Computational Intelligence, 2*(1), 59–62.

Barakat, N. H., & Bradley, A. P. (2007). Rule extraction from support vector machines: A sequential covering approach. *IEEE Transactions on Knowledge and Data Engineering, 19*(6), 729–741 Retrieved from. doi:10.1109/TKDE.2007.190610.

Barakat, N. H., Bradley, A. P., & Barakat, M. N. H. (2010). Intelligible support vector machines for diagnosis of diabetes mellitus. *IEEE Transactions on Information Technology in Biomedicine, 14*(4), 1114–1120 Retrieved from. doi:10.1109/TITB.2009.2039485.

Biostat Diabetes Dataset. (2018). Retrieved January 20, 2019, from http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

Brown, G. (2009). A new perspective for information theoretic feature selection. In *Proceedings of the twelfth international conference on artificial intelligence and statistics (AISTATS)* (pp. 49–56).

Chen, P., & Pan, C. (2018). Diabetes classification model based on boosting algorithms. *BMC Bioinformatics, 19*(1), 109 Retrieved from. doi:10.1186/s12859-018-2090-9.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the twenty-second ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794). Retrieved from. doi:10.1145/2939672.2939785.

Cho, N. H., Shaw, J. E., Karuranga, S., Huang, Y., da Rocha Fernandes, J. D., Ohlrogge, A. W., et al. (2018). IDF diabetes atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Research and Clinical Practice, 138*, 311–321 Retrieved from. doi:10.1016/j.diabres.2018.02.023.

Cohen, W. W. (1995). Fast effective rule induction. *Machine learning proceedings 1995* (pp. 115–123). Elsevier. doi:10.1016/B978-1-55860-377-6.50023-2.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297 Retrieved from. doi:10.1007/BF00994018.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc Retrieved from. doi:10.1002/047174882X.

Craven, M. W., & Shavlik, J. W. (1996). Extracting tree-structured representations of trained neural networks. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in neural information processing systems* (8, pp. 24–30). Cambridge, MA: MIT Press.

de Fortuny, E. J., & Martens, D. (2015). Active learning-based pedagogical rule extraction. *IEEE Transactions on Neural Networks and Learning Systems, 26*(11), 2664–2677 Retrieved from. doi:10.1109/TNNLS.2015.2389037.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research, 7*(Jan), 1–30.

Dheeru, D., & Karra Taniskidou, E. (2017). UCI Machine Learning Repository. Retrieved from http://archive.ics.uci.edu/ml

Diederich, J. (2008). Rule Extraction from Support Vector Machines: An Introduction, *Studies in computational intelligence* (80, pp. 3–31). BerlinHeidelberg: Springer. doi:10.1007/978-3-540-75390-2_1.

Diederich, J., & Barakat, N. (2004). Hybrid rule-extraction from support vector machines. In *Proceedings of the IEEE conference on cybernetics and intelligent systems* Retrieved from. doi:10.1109/ICCIS.2004.1460774.

Ding, S., Hua, X., & Yu, J. (2014). An overview on nonparallel hyperplane support vector machine algorithms. *Neural Computing and Applications, 25*(5), 975–982 Retrieved from. doi:10.1007/s00521-013-1524-6.

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association, 56*(293), 52. doi:10.2307/2282330.

Farquad, M. A. H., Ravi, V., & Bapi, R. S. (2008). Rule extraction using Support Vector Machine based hybrid classifier. In *Proceedings of the TENCON IEEE Region 10 Conference* (pp. 1–6). Retrieved from. doi:10.1109/TENCON.2008.4766534.

Farquad, M. A. H., Ravi, V., & Bapi, R. S. (2009). Support Vector Machine based Hybrid Classifiers and Rule Extraction thereof: Application to Bankruptcy Prediction in Banks. *Handbook of research on machine learning applications and trends: Algorithms, methods, and techniques* (pp. 404–426) Retrieved from. doi:10.4018/978-1-60566-766-9.ch019.

Farquad, M. A. H., Ravi, V., & Raju, S. B. (2010). Support vector regression based hybrid rule extraction methods for forecasting. *Expert Systems with Applications, 37*(8), 5577–5589 Retrieved from. doi:10.1016/j.eswa.2010.02.055.

Farran, B., Channanath, A. M., Behbehani, K., & Thanaraj, T. A. (2013). Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: Machine-learning algorithms and validation using national health data from Kuwait—a cohort study. *BMJ Open, 3*(5), e002457 Retrieved from. doi:10.1136/bmjopen-2012-002457.

Feld, S. (2002). The American association of clinical endocrinologists medical guidelines for the management of diabetes mellitus: The AACE system of intensive diabetes self-management - 2002 update. *Endocrine Practice, 8*(Supplement 1), 40–82. doi:10.4158/EP.8.S1.40.

Frank, E., & Witten, I. H. (1998). Generating accurate rule sets without global optimization. In *Proceedings of the fifteenth international conference on machine learning* (pp. 144–151).

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232 Retrieved from. doi:10.1214/aos/1013203451.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics and Data Analysis, 38*(4), 367–378 Retrieved from. doi:10.1016/S0167-9473(01)00065-2.

Friedman, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics, 11*(1), 86–92.

Fu, X., Ong, C., Keerthi, S., Hung, G. G., & Goh, L. (2004). Extracting the knowledge embedded in support vector machines. In *Proceedings of the IEEE international joint conference on neural networks* (pp. 291–296). Retrieved from. doi:10.1109/IJCNN.2004.1379916.

Fung, G., Sandilya, S., & Rao, R. B. (2005). Rule extraction from linear support vector machines. In *Proceeding of the eleventh ACM SIGKDD international conference on knowledge discovery in data mining - KDD '05* (p. 32). New York: ACM Press. Retrieved from. doi:10.1145/1081870.1081878.

Geng, L., & Hamilton, H. J. (2006). Interestingness measures for data mining. *ACM Computing Surveys (CSUR), 38*(3) 9–es. doi:10.1145/1132960.1132963.

Gorzałczany, M. B., & Rudziński, F. (2017). Interpretable and accurate medical data classification –A multi-objective genetic-fuzzy optimization approach. *Expert Systems with Applications, 71*, 26–39 Retrieved from. doi:10.1016/j.eswa.2016.11.017.

Guo, Q., Lu, X., Gao, Y., Zhang, J., Yan, B., Su, D., et al. (2017). Cluster analysis: A new approach for identification of underlying risk factors for coronary artery disease in essential hypertensive patients. *Scientific Reports, 7*, 43965 Retrieved from. doi:10.1038/srep43965.

Guzman, J. C., Melin, P., & Prado-Arechiga, G. (2017). Design of an optimized fuzzy classifier for the diagnosis of blood pressure with a new computational method for expert rule optimization. *Algorithms, 10*(3), 79. doi:10.3390/a10030079.

Guzmán, J. C., Melin, P., & Prado-Arechiga, G. (2018). Fuzzy optimized classifier for the diagnosis of blood pressure using genetic algorithm. In O. Castillo, P. Melin, & J. Kacprzyk (Eds.), *Fuzzy logic augmentation of neural and optimization algorithms: Theoretical aspects and real applications* (pp. 309–318). Cham: Springer. doi:10.1007/978-3-319-71008-2_23.

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 Retrieved from. doi:10.1109/JBHI.2014.2325615.

Hayashi, Y., & Yukita, S. (2016). Rule extraction using Recursive-Rule extraction algorithm with J48graft combined with sampling selection techniques for the diagnosis of type 2 diabetes mellitus in the Pima Indian dataset. *Informatics in Medicine Unlocked, 2*, 92–104 Retrieved from. doi:10.1016/j.imu.2016.02.001.

Hsu, C.-W., & Lin, C.-J. (2002). A simple decomposition method for support vector machines. *Machine Learning, 46*(1–3), 291–314 Retrieved from. doi:10.1023/A:1012427100071.

International Diabetes Federation IDF Diabetes Atlas-8th Edition. (2017). Retrieved June 7, 2018, from http://www.diabetesatlas.org/

Jayanthi, N., Babu, B. V., & Rao, N. S. (2017). Survey on clinical prediction models for diabetes prediction. *Journal of Big Data, 4*(1), 26 Retrieved from. doi:10.1186/s40537-017-0082-7.

Jin, X., Xu, A., Bie, R., & Guo, P. (2006). Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles. In *Proceedings of the international workshop on data mining for biomedical applications* (pp. 106–115). BerlinHeidelberg: Springer. Retrieved from. doi:10.1007/11691730_11.

Johansson, U., König, R., & Niklasson, L. (2003). Rule extraction from trained neural networks using genetic programming. In *Proceedings of the thirteenth international conference on artificial neural networks* (pp. 13–16).

Johansson, U., König, R., & Niklasson, L. (2004). The truth is in there-rule extraction from opaque models using genetic programming. In *Proceedings of the FLAIRS Conference* (pp. 658–663).

Kahramanli, H., & Allahverdi, N. (2009). Extracting rules for classification problems: AIS based approach. *Expert Systems with Applications, 36*(7), 10494–10502 Retrieved from. doi:10.1016/j.eswa.2009.01.029.

Katayama, S., Hatano, M., & Issiki, M. (2018). Clinical features and therapeutic perspectives on hypertension in diabetics. *Hypertension Research, 41*(4), 213–229 Retrieved from. doi:10.1038/s41440-017-0001-5.

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 Retrieved from. doi:10.1016/j.csbj.2016.12.005.

Ketema, E. B., & Kibret, K. T. (2015). Correlation of fasting and postprandial plasma glucose with HbA1c in assessing glycemic control; systematic review and meta-analysis. *Archives of Public Health, 73*(1), 43. doi:10.1186/s13690-015-0088-6.

Kurano, M., Darestani, S. G., Shinnakasu, A., Yamamoto, K., Dochi, Y., Uemura, K., et al. (2018). mRNA expression of platelet activating factor receptor (PAFR) in peripheral blood mononuclear cells is associated with albuminuria and vascular dysfunction in patients with type 2 diabetes. *Diabetes Research and Clinical Practice, 136*, 124–133. doi:10.1016/j.diabres.2017.11.028.

LaFreniere, D., Zulkernine, F., Barber, D., & Martin, K. (2016). Using machine learning to predict hypertension from a clinical dataset. In *Proceedings of the IEEE symposium series on computational intelligence (SSCI)* (pp. 1–7). Retrieved from. doi:10.1109/SSCI.2016.7849886.

Liu, J., Zhao, D., Liu, J., Qi, Y., Sun, J., & Wang, W. (2013). Prevalence of diabetes mellitus in outpatients with essential hypertension in China: A cross-sectional study. *BMJ Open, 3*(11), e003798 Retrieved from. doi:10.1136/bmjopen-2013-003798.

Lopez-jaramillo, P., Lopez-Lopez, J., Lopez-Lopez, C., & Rodriguez-Alvarez, M. I. (2014). The goal of blood pressure in the hypertensive patient with diabetes is defined: Now the challenge is go from recommendations to practice. *Diabetology & Metabolic Syndrome, 6*(1), 31 Retrieved from. doi:10.1186/1758-5996-6-31.

Louppe, G., Wehenkel, L., Sutera, A., & Geurts, P. (2013). *Understanding variable importances in forests of randomized trees*. Proceedings of the Advances in neural information processing systems. Retrieved from NIPS2013_4928

Luo, G. (2016). Automatically explaining machine learning prediction results: A demonstration on type 2 diabetes risk prediction. *Health Information Science and Systems, 4*(1) Retrieved from. doi:10.1186/s13755-016-0015-4.

Malmir, B., Amini, M., & Chang, S. I. (2017). A medical decision support system for disease diagnosis under uncertainty. *Expert Systems with Applications, 88*, 95–108 Retrieved from. doi:10.1016/j.eswa.2017.06.031.

Maniruzzaman, M., Kumar, N., Abedin, Menhazul, M. , Shaykhul Islam, M. , Suri, El-Baz, A. S., et al. (2017). Comparative approaches for classification of diabetes mellitus data: Machine learning paradigm. *Computer Methods and Programs in Biomedicine, 152*, 23–34 Retrieved from. doi:10.1016/j.cmpb.2017.09.004.

Marling, C., Wiley, M., Bunescu, R., Shubrook, J., & Schwartz, F. (2012). Emerging applications for intelligent diabetes management. *AI Magazine, 33*(2), 67–78 Retrieved from. doi:10.1609/aimag.v33i2.2410.

Martens, D., Baesens, B. B., & Van Gestel, T. (2009). Decompositional rule extraction from support vector machines by active learning. *IEEE Transactions on Knowledge and Data Engineering, 21*(2), 178–191 Retrieved from. doi:10.1109/TKDE.2008.131.

Melin, P., Miramontes, I., & Prado-Arechiga, G. (2018). A hybrid model based on modular neural networks and fuzzy systems for classification of blood pressure and hypertension risk diagnosis. *Expert Systems with Applications, 107*, 146–164 Retrieved from. doi:10.1016/j.eswa.2018.04.023.

Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., et al. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics, 10*(1), 213 Retrieved from. doi:10.1186/1471-2105-10-213.

Miramontes, I., Martínez, G., Melin, P., & Prado-Arechiga, G. (2017). A Hybrid Intelligent System Model for Hypertension Diagnosis. In P. Melin, O. Castillo, & J. Kacprzyk (Eds.), *Nature-inspired design of hybrid intelligent systems* (pp. 541–550). Springer. doi:10.1007/978-3-319-47054-2_35.

Miramontes, I., Martínez, G., Melin, P., & Prado-Arechiga, G. (2018). A hybrid intelligent system model for hypertension risk diagnosis. In P. Melin, O. Castillo, J. Kacprzyk, M. Reformat, & W. Melek (Eds.), *Fuzzy logic in intelligent system Design. NAFIPS 2017* (648, pp. 202–213). Cham: Springer International Publishing. doi:10.1007/978-3-319-67137-6_22.

Nanditha, A., Ma, R. C., Ramachandran, A., Snehalatha, C., Chan, J. C. N., Chia, K. S., et al. (2016). Diabetes in Asia and the Pacific: Implications for the global epidemic. *Diabetes Care, 39*(3), 472–485 Retrieved from. doi:10.2337/dc15-1536.

Núñez, H., Angulo, C., & Català, A. (2002). Rule extraction from support vector machines. In *Proceedings of the European symposium on artificial neural networks (ESANN)* (pp. 107–112). Retrieved from. doi:10.1016/j.neucom.2010.02.016.

Núñez, H., Angulo, C., & Català, A. (2004). Rule Based Learning Systems from SVM and RBFNN. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.144.2476

Ogurtsova, K., Rocha Fernandes, J. D., Huang, Y., Linnenkamp, U., Guariguata, L., Cho, N. H., et al. (2017). IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Research and Clinical Practice, 128*, 40–50 Retrieved from. doi:10.1016/j.diabres.2017.03.024.

Pourpanah, F., Lim, C. P., & Saleh, J. M. (2016). A hybrid model of fuzzy ARTMAP and genetic algorithm for data classification and rule extraction. *Expert Systems with Applications, 49*, 74–85 Retrieved from. doi:10.1016/j.eswa.2015.11.009.

Pulido, M., Melin, P., & Prado-Arechiga, G. (2018). A new model based on a fuzzy system for arterial hypertension classification. In O. Castillo, P. Melin, & J. Kacprzyk (Eds.), *Fuzzy logic augmentation of neural and optimization algorithms: Theoretical aspects and real applications* (pp. 319–327). Cham: Springer. doi:10.1007/978-3-319-71008-2_24.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81–106 Retrieved from. doi:10.1023/A:1022643204877.

Rodbard, D., & Vigersky, R. A. (2011). Design of a decision support system to help clinicians manage glycemia in patients with type 2 diabetes mellitus. *Journal of Diabetes Science and Technology, 5*(2), 402–411 Retrieved from. doi:10.1177/193229681100500230.

Sakr, S., Elshawi, R., Ahmed, A., Qureshi, W. T., Brawner, C., Keteyian, S., et al. (2018). Using machine learning on cardiorespiratory fitness data for predicting hypertension: The Henry Ford ExercIse Testing (FIT) Project. *PloS One, 13*(4), e0195344 Retrieved from. doi:10.1371/journal.pone.0195344.

Schmitz, G. P. J., Aldrich, C., & Gouws, F. S. (1999). ANN-DT: An algorithm for extraction of decision trees from artificial neural networks. *IEEE Transactions on Neural Networks, 10*(6), 1392–1401 Retrieved from. doi:10.1109/72.809084.

Seera, M., & Lim, C. P. (2014). A hybrid intelligent system for medical data classification. *Expert Systems with Applications, 41*(5), 2239–2249 Retrieved from. doi:10.1016/j.eswa.2013.09.022.

Sim, L. L. W., Ban, K. H. K., Tan, T. W., Sethi, S. K., & Loh, T. P. (2017). Development of a clinical decision support system for diabetes care: A pilot study. *PloS One, 12*(2), e0173021 Retrieved from. doi:10.1371/journal.pone.0173021.

Sowers, J. R., Epstein, M., & Frohlich, E. D. (2001). Diabetes, hypertension, and cardiovascular disease: An update. *Hypertension, 37*(4), 1053–1059 Retrieved from. doi:10.1161/01.HYP.37.4.1053.

Steinwart, I. (2003). Sparseness of support vector machines. *Journal of Machine Learning Research, 4*(11), 1071–1105 Retrieved from doi:http://www.jmlr.org/papers/v4/steinwart03a.html.

Steinwart, I., Hush, D., & Scovel, C. (2006). An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels. *IEEE Transactions on Information Theory, 52*(10), 4635–4643 Retrieved from. doi:10.1109/TIT.2006.881713.

Stoean, R., & Stoean, C. (2013). Modeling medical decision making by support vector machines, explaining by rules of evolutionary algorithms with feature selection.

Expert Systems with Applications, 40(7), 2677–2686 Retrieved from. doi:10.1016/j.eswa.2012.11.007.

Teramukai, S., Okuda, Y., Miyazaki, S., Kawamori, R., Shirayama, M., & Teramoto, T. (2016). Dynamic prediction model and risk assessment chart for cardiovascular disease based on on-treatment blood pressure and baseline risk factors. *Hypertension Research, 39*(2), 113–118 Retrieved from. doi:10.1038/hr.2015.120.

Tickle, A. B., Andrews, R., Golea, M., & Diederich, J. (1998). The truth will come to light: Directions and challenges in extracting the knowledge embedded within trained artificial neural networks. *IEEE Transactions on Neural Networks, 9*(6), 1057–1068 Retrieved from. doi:10.1109/72.728352.

Tickle, A. B., Orlowski, M., & Diederich, J. (1994). DEDEC: Decision detection by rule extraction from neural networks. *Neurocomputing Research Center QUT NRC*.

van Dieren, S., Beulens, J. W. J., Kengne, A. P., Peelen, L. M., Rutten, G. E. H. M., Woodward, M., et al. (2012). Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes: A systematic review. *Heart, 98*(5), 360–369 Retrieved from. doi:10.1136/heartjnl-2011-300734.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer Verlag Inc Retrieved from. doi:10.1007/978-1-4757-3264-1.

Wang, G., Deng, Z., & Choi, K. S. (2015). Detection of epileptic seizures in EEG signals with rule-based interpretation by random forest approach. In *Proceedings of the international conference on intelligent computing* (pp. 738–744). Cham: Springer. Retrieved from. doi:10.1007/978-3-319-22053-6_78.

Wang, X., Huang, F., & Cheng, Y. (2014). Super-parameter selection for Gaussian-Kernel SVM based on outlier-resisting. *Measurement, 58*, 147–153 Retrieved from. doi:10.1016/j.measurement.2014.08.019.

WHO. (2011). *Waist circumference and waist-hip ratio: Report of a WHO expert consultation*. World Health Organization Technical Report.

World Health Organization. (2013). *A global brief on hypertension: Silent killer, global public health crisis: World Health Day 2013*. Retrieved from http://apps.who.int/iris/handle/10665/79059

Zhang, Y., Li, Z., Tang, Y., & Cui, K. (2004). DRC-BK: Mining classification rules with help of SVM. In *Proceedings of the Pacific-Asia conference on knowledge discovery and data mining* (pp. 191–195). BerlinHeidelberg: Springer. Retrieved from. doi:10.1007/978-3-540-24775-3_24.

Zhang, Y., Su, H., Jia, T., & Chu, J. (2005). Rule extraction from trained support vector machines. In *Proceedings of the Pacific-Asia conference on knowledge discovery and data mining* (pp. 61–70). BerlinHeidelberg: Springer. Retrieved from. doi:10.1007/11430919_9.