Informatics in Medicine Unlocked 2 (2016) 92–104

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

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

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

Yoichi Hayashi *, Shonosuke Yukita

Department of Computer Science, Meiji University, Tama-ku, Kawasaki, Kanagawa 214-8571, Japan

ARTICLE INFO

Article history:
Received 19 December 2015
Received in revised form
29 January 2016
Accepted 24 February 2016
Available online 23 April 2016

Keywords:
Rule extraction
Type 2 diabetes mellitus
Re-RX algorithm
Sampling selection
Pima Indian diabetes
Data mining

ABSTRACT

Diabetes is a complex disease that is increasing in prevalence around the world. Type 2 diabetes mellitus (T2DM) accounts for about 90–95% of all diagnosed adult cases of diabetes. Most present diagnostic methods for T2DM are black-box models, which are unable to provide the reasons underlying diagnosis to physicians; therefore, algorithms that can provide further insight are needed. Rule extraction can provide such explanations; however, in the medical setting, extracted rules must be not only highly accurate, but also simple and easy to understand. The Recursive-Rule eXtraction (Re-RX) algorithm is a "white-box" model that provides highly accurate classification. However, due to its recursive nature, it tends to generate more rules than other algorithms. Therefore, in this study, we propose the use of a rule extraction algorithm, Re-RX with J48graft, combined with sampling selection techniques (sampling Re-RX with J48graft) to achieve highly accurate, concise, and interpretable classification rules for the Pima Indian Diabetes (PID) dataset, which comprises 768 samples with two classes (diabetes or non-diabetes) and eight continuous attributes. The use of this algorithm resulted in an average accuracy of 83.83% after 10 runs of 10-fold cross validation. Sampling Re-RX with J48 graft achieved substantially better accuracy and provided a considerably fewer average number of rules and antecedents than the original Re-RX algorithm. These results suggest that sampling Re-RX with J48graft provides more accurate, concise, and interpretable extracted rules than previous algorithms, and is therefore more suitable for medical decision making, including the diagnosis of T2DM.

1. Introduction

Diabetes is a complex disease characterized by a lack of or resistance to insulin, a hormone critical for the regulation of blood sugar. In healthy individuals, the pancreas produces insulin to help metabolize sugar in the blood and keep blood glucose (sugar) levels within a normal range. Diabetics cannot produce or are resistant to insulin, and as a result, are unable to remove glucose from their bloodstreams. Consequently, glucose levels in the blood increase, leading to serious health problems [1].

In 2011, there were 347 million diabetics worldwide, and by 2030, this number is expected to increase to 552 million. About 4.6 million deaths were caused by diabetes in 2011, and by 2030, it is projected to be the seventh leading cause of death [2].

According to the Centers for Disease Control and Prevention, an estimated 29.1 million people, or 9.3% of the US population, have diabetes [3], 8.1 million of whom remain undiagnosed. In 2010, diabetes was listed as the underlying cause of death on 69,071 death certificates and a cause of death another 234,051, making it the seventh leading cause of death in the US.

Diabetes can affect the entire body and is associated with severe complications such as heart disease, stroke, vision loss, kidney failure, and lower-limb amputations. Good glucose control can help avoid some complications, particularly microvascular eye, kidney, and nerve disease, and early detection and treatment can help prevent disease progression; therefore, monitoring that includes dilated eye exams, urine tests, and foot exams is essential. Because diabetics and prediabetics are at an increased risk of cardiovascular disease, blood pressure and lipid management, and especially smoking cessation, are particularly important.

There are two main clinical classifications of diabetes: type 1 and type 2. Onset of type 1 diabetes, which was previously known as insulin-dependent diabetes mellitus or juvenile-onset

* Corresponding author.
E-mail addresses: hayashiy@cs.meiji.ac.jp (Y. Hayashi), redgtvo9606@gmail.com (S. Yukita).

http://dx.doi.org/10.1016/j.imu.2016.02.001

diabetes, accounts for about 5% of all diagnosed adult cases of diabetes. Although it can occur at any age, the peak age for diagnosis of type 1 diabetes is in the mid-teens.

The peak age of onset of type 2 diabetes mellitus (T2DM), which was previously known as non–insulin-dependent diabetes mellitus or adult-onset diabetes, is typically later than that of type 1 diabetes and accounts for about 90–95% of all diagnosed adult cases of diabetes. T2DM usually starts with insulin resistance, a disorder in which cells primarily within the muscles, liver, and fat tissue do not utilize insulin properly. The beta cells in the pancreas begin to gradually lose the ability to produce sufficient quantities of insulin as the need for the hormone increases. In contrast to beta cell dysfunction, the role of insulin resistance differs among individuals; some primarily have insulin resistance and only a minor defect in insulin secretion, while others primarily have a lack of insulin secretion and only slight insulin resistance.

Although the exact causes of complex diseases such as T2DM have yet to be identified [4], a combination of genetic, environmental, and lifestyle factors is suspected [5]. An ever-increasing amount of data is being collected in medical databases, and historical data on complex diseases, such as patients' blood glucose levels, is becoming more widely available; therefore, traditional methods of manual analysis have become inadequate. As a result, a variety of data mining techniques are being applied in order to discover new patterns of disease and promote the early detection and diagnosis of complex diseases such as diabetes [6].

The diagnosis of T2DM is a two-class classification problem, and numerous methods for diagnosing T2DM have been successfully applied to the classification of different tissues. However, most present diagnostic methods [1,7–47] for T2DM are black-box models. A drawback of black-box models is that they cannot adequately reveal information that may be hidden in the data.

For example, even in cases for which high-performance classifiers [2,4,8,24,25,32,33] allow the accurate assignment of instances to groups, black-box models are unable to provide the reasons underlying that assignment to physicians; therefore, algorithms that can provide insight into these underlying reasons are needed. Rule extraction can provide such explanations, and is it therefore becoming increasingly popular. However, in the medical setting, extracted rules must be not only highly accurate, but also simple and easy to understand. Rules are one of the most popular symbolic representations of knowledge discovered from data, and are more comprehensible, particularly "black boxes" like unseen medical datasets, than other representations [48].

The Recursive-Rule eXtraction (Re-RX) algorithm, originally intended to be a rule extraction tool, was recently developed by Setiono et al. [49]. Re-RX provides a hierarchical, recursive consideration of discrete variables prior to analysis of continuous data, and can generate classification rules from neural networks (NNs) that have been trained on the basis of both discrete and continuous attributes.

In contrast to black-box models, the Re-RX algorithm [49] is a "white-box" model that provides highly accurate classification. It is easy to explain and interpret in accordance with the concise extracted rules associated with IF-THEN forms. Due to its ease of understanding, the Re-RX algorithm is typically preferred by both physicians and clinicians alike.

However, due to its recursive nature, the Re-RX algorithm tends to generate more rules than other rule extraction algorithms. Therefore, one of the major drawbacks of the Re-RX algorithm is that it typically generates expansive extraction rules for middle-sized or larger datasets.

It is important to consider both accuracy and interpretability for extracted classification rules. The number of correctly classified test samples typically determines the accuracy of each extracted classification rule, while the number of extracted rules and the average number of antecedents in the extracted rules determines their interpretability.

To achieve both concise and highly accurate extracted rules while maintaining the good framework of the Re-RX algorithm, we recently proposed supplementing the Re-RX algorithm with J48graft [51], a class for generating a grafted C4.5 decision tree [50]. J48graft [52] is the result of the C4.5A [53] algorithm being implemented in open source data mining software referred to as the "all-tests-but-one partition (ATBOP)" [53]. In Re-RX with J48graft, J48graft [52] is employed to form decision trees in a recursive manner, while multi-layer perceptrons (MLPs) are trained using backpropagation (BP), which allows pruning [54], thereby generating more efficient MLPs for highly accurate rule extraction. Re-RX with J48graft provides rules that are not only highly accurate, but also easily explained and interpreted in terms of the concise extracted rules; that is, Re-RX with J48graft provides IF-THEN rules. This white-box model is easier to understand and is therefore often preferred in the medical setting.

In this study, we first proposed the use of a rule extraction algorithm, Re-RX with J48graft [51], combined with sampling selection techniques (sampling Re-RX with J48graft) [55,56] for preprocessing. We then investigated the accuracy, conciseness, and interpretability of diagnostic rules extracted for the Pima Indian Diabetes (PID) dataset using sampling Re-RX with J48graft based on a comparison with both crisp rule extraction techniques [21,27,28] and previous fuzzy rule extraction techniques [1,12–16,29–31,43]. As a typical example of T2DM, we used the PID dataset from the repository of machine learning at the University of California Irvine (UCI) [57]. The PID dataset comprises 768 samples with two classes (diabetes or non-diabetes) and eight continuous attributes. Important values missing from the PID dataset are discussed in Section 3.7.

In Section 5, we review the performance of rule extraction algorithms for the PID dataset since 2003, and compare the previous extracted fuzzy and crisp rules with the performance of the present extracted rules. In Sections 5.1–5.6, we compare the concrete rules for the PID dataset extracted by the proposed algorithm with those obtained using the four kinds of previous rule extraction algorithms recommended by the American Diabetes Association (ADA) for the diagnosis of diabetes. In Section 5.7, we also compare the classification accuracy obtained by the proposed algorithm with that obtained by other classifier systems for the PID dataset.

We explain the role of the oral glucose tolerance test (OGTT) and body mass index (BMI) for the diagnosis of the PID dataset in Section 6.1, and discuss the interpretation of rules extracted by the proposed algorithm from the perspective of medical informatics in Section 6.2. In Section 6.3, we discuss the trade-offs between accuracy and the number of extracted rules using trade-off curves, and in Section 6.4, we elucidate the trade-offs between accuracy and the average number of antecedents. Finally, we provide a summary and conclusion in Section 7.

2. Related works

In 1996, Shanker [10] evaluated the effectiveness of artificial NN (ANN) classifiers in predicting the onset of non–insulin-dependent diabetes mellitus among the Pima Indian female population. According to Knowler et al., the Pima Indians have the highest reported incidence of diabetes in the world [58]. Smith et al. [59] used the same dataset to test a model for predicting the onset of diabetes mellitus. In this study, ANNs were used to model the relationship between the onset of diabetes mellitus and various risk factors for diabetes among Pima Indian women.

Diagnosing T2DM is a two-class classification problem, and numerous methods for diagnosing T2DM have been successfully applied to the classification of different tissues. These methods include the following: bee colony algorithm [1]; extreme learning machines [7,46]; support vector machines (SVMs) [8,9,35]; NNs [10,11,36,37,41]; fuzzy classification [12,47]; fuzzy modeling [13]; fuzzy decision tree [14]; fuzzy rule extraction from SVMs [15]; evolving fuzzy rule-based classification [16]; mixture of expert models [17]; immune recognition systems [18,19]; neuro-fuzzy inference systems [20,38]; swarm optimization [21,27]; multiple classifier system [4]; hybrid intelligent system [2,22]; genetic programming [23]; hybrid prediction model [24,25]; granular computing [26]; genetic algorithm [28,39]; neuro-fuzzy system [29]; fuzzy classifier [30,31]; hybrid classifier [32]; classifier ensemble [33]; similarity classification [34,40]; radial basis function classifier [42]; evolutionary algorithm [43]; electromagnetism-like mechanism [44]; and ARTMAP-CART [45].

We provide a brief description for four rule extraction algorithms [1,12,16,30] used for comparisons in Section 5. The Artificial Bee Colony (ABC) algorithm proposed by Beloufa and Chikh [1] adds a mutation operator to an Artificial Bee Colony to improve its performance. This modified ABC can be used to automatically create and optimize membership functions and rules directly from the data.

The classification methodology proposed by Gadaras and Mikhailov [12] identifies fuzzy boundaries of classes by processing a set of labeled data. Fuzzy rules are obtained by exploring the characteristics of the identified boundaries and automatically producing membership functions for each class. When new patterns require classification, their numerical attributes are tested against generated knowledge to match a patient's symptoms with an antecedent.

A study on semi-supervised evolving fuzzy classification was conducted by Lekkas and Mikhailov [16] for the diagnosis of two medical problems. In their system, two domains contain records of actual patients with a known diagnosis. Their aim was to review the existing methodology for evolving fuzzy classification in order to improve upon it and evaluate its performance compared with other systems.

Finally, the design of fuzzy systems in relation to the data was investigated by Chang and Lilly [30]. They proposed the use of a new evolutionary approach to derive a compact fuzzy classification system directly from the data without any a priori knowledge or assumptions regarding the distribution of the data. The fuzzy classifier is initially empty with no rules in the rule base and no membership functions assigned to fuzzy variables. Rules and membership functions are then automatically created and optimized in an evolutionary process.

3. Methods

3.1. Re-RX algorithm

The Re-RX algorithm generates classification rules from both continuous and discrete datasets. It produces hierarchical rules, applying different rule conditions for discrete and continuous attributes, such that only the rules lowest in the hierarchy contain continuous attributes. Here, although the proposed algorithm can readily handle multiple groups, two-group classification problems are considered exclusively. The algorithm structure and functioning are described as follows.

Algorithm Re-RX (S, D, C).
Input: A set of data samples, S, having discrete attributes, D, and continuous attributes, C.
Output: A set of classification rules.

1. Train and prune [54] an NN using dataset S, including all of its D and C attributes.
2. Let D' and C' be the sets of discrete and continuous attributes, respectively, still present in the network, and let S' be the set of data samples correctly classified by the pruned network.
3. If $D'=\phi$, generate an axis hyperplane to split the samples in S' according to the values of the continuous attributes, C', then stop.

Otherwise, use only the discrete attributes, D', to generate the set of classification rules, R, for dataset S'.

4. For each rule, R_i, that is generated:

If support $(R_i) > \delta_1$ and error$(R_i) > \delta_2$, then

- Let S_i be the set of data samples that satisfy the condition of rule R_i, and let D_i be the set of discrete attributes that do not appear in rule condition R_i.
- If $D_i=\phi$, then generate an axis hyperplane to split the samples in S_i according to the values of their continuous attributes, C_i, then stop.
- Otherwise, call Re-RX (S_i, D_i, C_i).

Assuming a suitable pruning rate, Step 1 can employ a variety of NN training and pruning methods. Although the Re-RX algorithm makes no assumptions regarding the NN architecture, we have focused on BPNNs with a single hidden layer, allowing universal approximation.

The percentage of samples covered by a rule defines its support, and Step 4 assesses both the rule support and the corresponding error rate. The rule subspace is further partitioned if the error rate is above a threshold value, δ_2, and the support equals the approximate maximum threshold value, δ_1. If discrete attributes are absent from the rule conditions, *subdivision* is achieved by recursively calling Re-RX or by producing a separate axis hyperplane incorporating only the continuous data attributes.

The subdivision of the Re-RX algorithm is a unique function and inherent in its nature. This subdivision allows the use of other unused attributes, which increases both the number and accuracy of extracted rules by each subdivision process.

Needless to say, accuracy, comprehensibility, and conciseness in extracted rules have important trade-offs. Extracted rules before subdivision are more concise and interpretable, yet have lower accuracy, whereas extracted rules after subdivision are less concise, but have better accuracy.

A major advantage of the Re-RX algorithm developed by Setiono et al. [49] is that it was intended as a rule extraction tool. It provides a hierarchical, recursive consideration of discrete variables prior to analysis of continuous data, and is able to generate classification rules from NNs that have been trained on the basis of discrete and continuous attributes.

In other words, the Re-RX algorithm achieves a very high accuracy rule extraction method that also offered comprehensibility by generating perfect or strict separation between discrete attributes and continuous attributes in the antecedent of each extracted rule.

3.2. J4.8

J4.8 [60] is a Java-implemented version of C4.5 [50], an advanced version of the ID3 algorithm developed by Quinlan [61]. The decision trees generated by C4.5 are used for classification; therefore, this algorithm is typically described as a statistical classifier. C4.5 performs very similarly to ID3, except that it

determines the best target attribute using the gain ratio. Also, in contrast to ID3, C4.5 has the improved ability to handle numerical attributes by creating a threshold, and then splitting the data into those whose attribute value is either greater, or less than or equal to, that threshold. This algorithm also has the ability to handle attributes with variable cost. Finally, C4.5 can prune the decision tree after its creation, which reduces its size and thereby saves both time and memory.

3.3. J48graft

The concept of tree grafting is based on the desire to discard the "simplest is best" method for selecting a good tree. In contrast, in tree grafting, the focus is on the fact that similar objects tend to have the highest probability of belonging to the same class. In other words, if the final result is a better classification model, the need to yield more complex trees is eliminated.

Grafting is a post-process that can be readily applied to decision trees. Its main objective is to reclassify regions of an instance space where no training data exists or where there is only misclassified data, and as a result, to decrease prediction error. Grafting identifies the best-suited cuts of existing leaf regions and then branches out to create new leaves with classifications that differ from the original. In this process, the tree becomes more complex naturally. However, only branching that does not introduce classification errors in data that has already been correctly classified is considered, ensuring that the new tree reduces errors.

Webb introduced the C4.5A algorithm referred to as ATBOP, which is a more efficient method for evaluating potentially supporting evidence [53]. The ATBOP region of a leaf is formed by removing all the enclosing decision surfaces. Using ATBOP allows a reduction in computational requirements because the only set of training data considered for each leaf is that from the ATBOP region. The J48graft is the result of the C4.5A algorithm being implemented in open source data mining software known as the Waikato Environment for Knowledge Analysis (Weka) [60].

Pruning is a process that can be thought of as the opposite of grafting because it aims to reduce rather than increase the complexity of a decision tree while retaining good predictive accuracy. Surprisingly, Webb [62] concluded that, either despite or possibly because of the fact they are opposites, pruning and grafting work well in parallel. Grafting takes instances outside the analyzed leaf (global information) into account, while pruning only looks at instances within the analyzed leaf (local information). In this way, they seem to complement each other. In most cases, using both grafting and pruning on a decision tree yields a lower prediction error that using them separately [62].

3.4. Re-RX algorithm with J48graft

To enhance the accuracy and conciseness of classification rules, we proposed replacing the conventional Re-RX algorithm, which uses C4.5 as a decision tree [50], with Re-RX with J48graft. Concepts in the conventional pruning used in J4.8 and grafting used in J48graft [52] both contrast and complement each other. We believe that the performance of the Re-RX algorithm [49] is greatly affected by the decision tree. In consideration of the grafting properties in J48graft, our idea is to use the grafting concepts in the Re-RX algorithm to enhance the accuracy and conciseness of the extracted rules. Therefore, we replace J4.8 with J48graft in the Re-RX algorithm. We also expect that Re-RX with J48graft will generate much more accurate and concise classification rules.

One of the difficulties associated with using feedforward NNs is the need to determine the optimal number of hidden units before the training process can begin. Too many hidden units may lead to overfitting of the data and poor generalization, while too few may not result in an NN that learns the data. Setiono [54] proposed two different approaches to overcome the problem of determining the optimal number of hidden units required by an ANN to solve a given problem. The first begins with a minimal network and adds more hidden units only when they are needed to improve its learning capability. The second begins with an oversized network and then prunes redundant hidden units.

In the present paper, we first trained MLP using BP, then we started pruning from a trained MLP to a pruned MLP with a smaller number of connections; this allowed us to extract accurate and concise rules using Re-RX with J48graft. The amount of pruning carried out was about 70% for the PID dataset; the amount of pruning depends on the characteristics of the training dataset. This amount for pruning is considerably bigger than that is carried out within J48graft.

In summary, we frequently employ J48graft in Re-RX with J48graft [52] to form decision trees in a recursive manner, while we train MLPs using BP, which allows pruning [54] and therefore generates more efficient MLPs for rule extraction. The schematic overview of the Re-RX with J48graft is shown in Fig. 1.

3.5. Sampling selection

Instead of building more sophisticated models for two-class classification problems such as PID, Setiono [55,56] proposed a method that focuses on how the accuracy of the models can be improved by selecting relevant training data samples.

In a supervised learning scheme, classification models for PID, such as NNs, are trained using a historical dataset in which each sample has been labeled as either diabetes or non-diabetes. However, some of these class labels may be incorrectly assigned, and irregular data samples may be present. Although these samples have similar attributes, like the majority of samples in one class, they actually belong to a different class. The presence of irregular and/or mislabeled data samples in the training dataset is therefore likely to affect the performance of the NNs.

Therefore, the sampling selection technique proposed by Setiono et al. [55,56] removes these data samples before building a model that distinguishes between diabetes and non-diabetes. An NN ensemble is then trained to identify potentially irregular and/or mislabeled data samples, and data samples that are consistently misclassified by the majority of NNs in the ensemble are removed.

The sample selection technique can be summarized as follows: 1) Ensemble creation: train an ensemble of M feedforward NNs using the available training data samples; 2) Sample selection: select training data samples based on the predictions of the NN ensemble; 3) Model generation: use the selected samples to train an NN; and 4) Rule extraction: apply an NN rule extraction algorithm to obtain concise and interpretable classification rules capable of distinguishing between diabetes and non-diabetes.

Sampling selection in Step 2 is a core component of the sampling selection technique. In this study, we employed an NN ensemble to identify outliers in the training dataset. Removing outliers and noise prior to learning has been shown to improve the predictive accuracy of numerous learning methods. A data sample was labeled as an outlier, and subsequently discarded, if it was incorrectly classified by a proportion of NNs exceeding the threshold ρ; otherwise, the sample is retained in the training dataset.

For example, the predictive output of each data sample from 30 NN ensembles is tabulated. If a value of $\rho=0.9$ is set, samples misclassified by 27 or more NNs in the ensemble are discarded.

Therefore, if we set a lower value for ρ, e.g., 0.6, overfitting of the data can be avoided. However, the primary purpose of sampling selection is not to remove all the outliners and noise, but to improve predictive accuracy. Considering the characteristics of the

training dataset, we can obtain values of ρ that will allow us to minimize the chance of overfitting.

3.6. Re-RX algorithm with J48graft and sampling selection technique (sampling Re-RX with J48graft)

We propose a new highly accurate, concise, and interpretable rule extraction algorithm using Re-RX with J48graft combined with sampling selection techniques (sampling Re-RX with J48graft) for preprocessing.

The objective of the present study was to achieve highly accurate, concise, and interpretable classification rules for the PID dataset. However, the PID dataset for rule extraction was a medical dataset, so the focus was on decreasing the number of extracted rules and the average number of antecedents. To extract concise rules, we employed sampling Re-RX with J48graft, which is better suited for achieving concise and interpretable, as opposed to accurate, medical rules.

We preprocessed the PID dataset using the sample selection technique [55,56] to extract a fewer number of rules and lower average number of antecedents. We then employed Re-RX with J48graft to extract a set of concise and interpretable diagnostic rules for the PID dataset. A schematic overview of sampling Re-RX with J48graft is shown in Fig. 2. As shown in the figure, a supplementary cross-validation loop is carried out with sampling selection by an NN ensemble.

The most important aim of sampling Re-RX with J48graft is to improve the conciseness and interpretability of extracted rules for

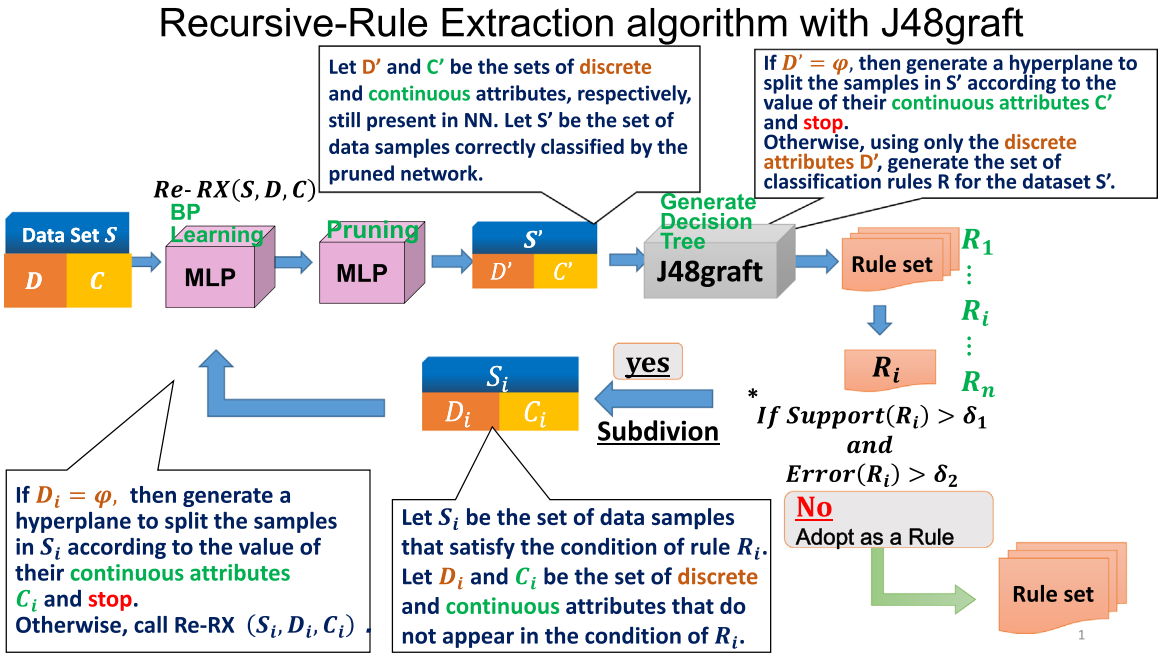

Fig. 1. Schematic overview of the Recursive-Rule eXtraction (Re-RX) algorithm with J48graft (Re-RX with J48graft).

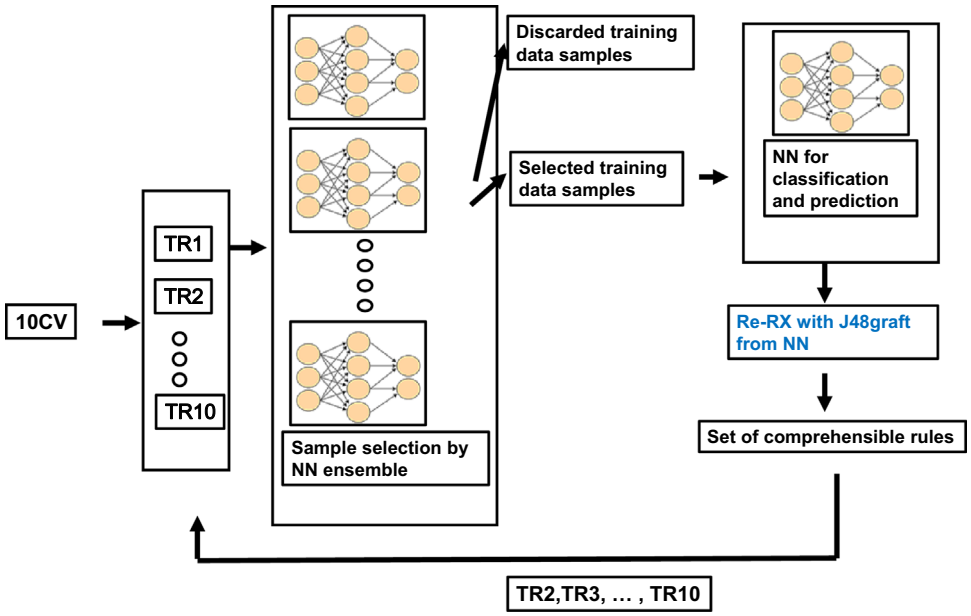

Fig. 2. Schematic overview of sampling Re-RX with J48graft **NN**: Neural network.

physicians, because the competition for achieving only better classification accuracy for the PID dataset has appeared to plateau [8,63], and unless diagnostic accuracy can be substantially improved, no significant contributions will be made to medical informatics.

3.7. Pima Indian Diabetes (PID) dataset and experimental setup

The PID dataset consists of 768 samples of eight numerical attributes [57] and is part of a larger dataset held by the National Institutes of Diabetes and Digestive and Kidney Diseases in the US. The values of these medical attributes come from Pima Indian women at least 21 years of age residing in or near Phoenix, AZ. The class variable takes the values "0" or "1" , indicating a negative and positive test for diabetes, respectively. In addition, theT2DM predominant in this population is said to have slow and gradual commencement. As a consequence, traditional diagnostic methods that are partially based on the plasma glucose test may be delayed by up to 10 years [64]. The eight clinical features for this population are as follows:

1. Number of times pregnant (NP).
2. Plasma glucose concentration after 2 h in an OGTT.
3. Diastolic blood pressure (mmHg) (DBP).
4. Triceps skinfold thickness (mm) (TSFT).
5. Two-hour serum insulin (μU/mL) (2HSI).
6. BMI.
7. Diabetes pedigree function (DPF).
8. Age (years) (AGE).

There are 500 samples from non-diabetic patients and 268 from diabetic patients. This dataset was selected because it is commonly used by the other classification systems evaluated in this study.

In 2001, 376 of 786 observations in the PID dataset were shown to lack experimental validity [65] because for some attributes, the value of *zero* was recorded in place of missing experimental observations [66]. It was also shown that if the instances with *zero* values were removed, performance could be dramatically improved [65]. In 2002, data preparation was shown to be a critically important step for the analysis of large diabetic datasets from the practical medical informatics view because the value of data mining or analysis depends on it [68]; that report was invaluable in regards to data mining for a moderate size diabetes dataset.

In 2011, Gagliardi [66] reported that an instance-based classifier (k-nearest neighbor classifier (k-NNC)) achieved 76.8% accuracy for Breault's [65] modified version of the PID dataset using the leave-one-out procedure as a cross validation (CV) technique.

In 2012, Chikh et al. [68] reported achieving a high classification accuracy (89.10%) after applying an Artificial Immune Recognition System2 (AIRS2) with k-NNC using 10-fold CV. After removing cases with unreasonable physical data, there were a total of 392 samples (262 normal and 130 diabetes samples).

Therefore, nearly all of the studies in the literature make use of the same version of the data; however, in these studies, these data are mistaken for correct data because the fact that *zeros* were actually missing values was not understood.

Therefore, in this study, the UCI machine learning database was used as the benchmark [57]. This dataset is very commonly used to test other classification systems, and thereby easier to compare their results with those of the proposed model for the diagnosis of the PID database.

4. Results

4.1. Performance

To guarantee the validity of the results, we used k-fold CV [69] to evaluate the classification rule accuracy of test datasets. The k-fold CV method is widely applied by researchers to minimize the bias associated with random sampling.

We trained the PID dataset using sampling Re-RX with J48graft and obtained 10 runs of 10-fold CV accuracies for the training dataset (TR ACC), 10 runs of 10-fold CV accuracies for the test dataset (TS ACC), the number of extracted rules (# rules), the average number of antecedents (Ave. # ante.), and the area under the receiver operating characteristics curve (AUC) [70] (Table 1). In this paper, the AUC was used as an appropriate evaluator because it does not include class distribution or misclassification costs [70].

Numerous types of rules have been suggested in the literature from the perspective of the expressive power of extracted rules, including propositional rules, which take the form of IF-THEN expressions and clauses defined using propositional logic, and M-of-N rules. Breaking from traditional logic, fuzzy rules allow partial truths instead of Boolean true/false outcomes.

Even if all types of rules are considered, the consensus is that no matter how they are defined, an ideal measure has yet to be developed; therefore, "what is a concise and/or interpretable rule? " remains a difficult question to answer.

To answer this question, we attempted to develop a "rough indicator" of conciseness by comparing the average number of antecedents from extracted rules generated using a variety of techniques.

Regarding the complexity of sampling Re-RX with J48graft, it took about 6.4 s to train the PID dataset using a standard workstation computer (3.1 GHz Intel Xeon E5-2687W, 3.5 GHz Turbo, 25 MB Cache; 64 GB RAM; 512 GB DDR3 System memory) and about 64 s for 10-fold CV. The testing time was negligible.

We achieved an average accuracy of 83.83% after 10 runs of 10-fold CV for the PID dataset. The performance of the original Re-RX algorithm [49], i.e., Re-RX with C4.5, is shown in Table 2.

Comparing Table 1 with Table 2, we confirmed that the sampling Re-RX with J48graft achieved more accurate and much more concise and interpretable extracted rules for the PID dataset. That is, sampling Re-RX with J48 graft achieved substantially better accuracy (83.83% for the PID dataset) than the original Re-RX algorithm (80.00%). In addition, sampling Re-RX with J48graft provided a considerably fewer average number of rules and antecedents compared with the original Re-RX algorithm.

Table 1
Performance of sampling Re-RX with J48graft for the Pima Indian Diabetes (PID) dataset (average of 10 runs of 10-fold cross validation [CV]).

PID dataset	TR ACC (%)	TS ACC (%)	# Rules	Ave. # ante.	AUC	TR ACC (SD)	TS ACC (SD)
Sampling Re-RX with J48graft	84.97	**83.83**	**8.21**	**2.01**	0.816	1.49	1.63

PID: Pima Indian Diabetes; CV: cross validation; Re-RX: Recursive-Rule eXtraction; Re-RX with J48graft: Recursive-Rule eXtraction algorithm with J48graft; TR: training dataset; TS: testing dataset; ACC: accuracy; Ave. # ante.: average number of antecedents; AUC: area under the receiver operating characteristic curve; SD: standard deviation.

5. Comparisons

We reviewed the rule extraction algorithms used for the PID dataset since 2003 and tabulated their performances in Table 3. The concrete rules extracted for the PID dataset by sampling Re-RX with J48graft are shown in Section 5.1. The four kinds of rules for the PID dataset reported in previous studies are described in Sections 5.2–5.5. In Section 5.6, we compare sampling Re-RX with J48graft with previous algorithms. In Section 5.7, we compare the classification accuracy obtained using sampling Re-RX with J48graft with other classifier systems for the PID dataset.

5.1. Rules extracted for the PID dataset by sampling Re-RX with J48graft

R1: If OGTT $\leq$ 125 Then Non-Diabetes
R2: If OGTT $\in$ (125, 139] AND BMI $\leq$ 36.1 Then Non-Diabetes
R3: If OGTT $\in$ (125, 129] AND BMI $\in$ (36.1, 39.6] Then Non-Diabetes
R4: If OGTT $\in$ (129, 139] AND BMI $>$ 36.1 Then Diabetes
R5: If OGTT $\in$ (139, 151] AND BMI $\leq$ 28.6 Then Non-Diabetes
R6: If OGTT $>$ 151 AND BMI $\leq$ 28.6 Then Diabetes
R7: If OGTT $>$ 139 AND BMI $>$ 28.6 Then Diabetes

5.2. Rules extracted for the PID dataset by artificial bee colony [1]

5.2.1. The rules with all features

R1: If (NP is H) AND (OGTT is L) AND (DBP is H) AND (TSFT is L) AND (2HSI is L) AND (BMI is H) AND (DPF is H) AND (Age is L) Then Non-Diabetes
R2: If (NP is L) AND (OGTT is L) AND (DBP is L) AND (TSFT is L) AND (2HSI is L) AND (BMI is L) AND (DPF is H) AND (Age is L) Then Non-Diabetes
R3: If (NP is H) AND (OGTT is L) AND (DBP is H) AND (TSFT is L) AND (2HSI is L) AND (BMI is H) AND (DPF is L) AND (Age is L) Then Non-Diabetes
R4: If (NP is L) AND (OGTT is L) AND (DBP is L) AND (TSFT is H) AND (2HSI is L) AND (BMI is L) AND (DPF is L) AND (Age is L) Then Non-Diabetes
R5: If (NP is H) AND (OGTT is L) AND (DBP is H) AND (TSFT is L) AND (2HSI is H) AND (BMI is L) AND (DPF is L) AND (Age is L) Then Non-Diabetes
R6: If (NP is L) AND (OGTT is L) AND (DBP is H) AND (TSFT is L) AND (2HSI is L) AND (BMI is L) AND (DPF is L) AND (Age is L) Then Non-Diabetes
R7: If (NP is L) AND (OGTT is L) AND (DBP is L) AND (TSFT is L) AND (2HSI is L) AND (BMI is L) and (DPF is L) AND (Age is L) Then Non-Diabetes
R8: If (NP is H) AND (OGTT is H) AND (DBP is H) AND (TSFT is L) AND (2HSI is H) AND (BMI is L) AND (DPF is H) AND (Age is L) Then Diabetes
R9: If (NP is L) AND (OGTT is H) AND (DBP is L) AND (TSFT is L) AND (2HSI is H) AND (BMI is H) AND (DPF is H) AND (Age is H) Then Diabetes
R10: If (NP is L) AND (OGTT is H) AND (DBP is L) AND (TSFT is H) AND (2HSI is L) and (BMI is L) AND (DPF is L) AND (Age is L) Then Diabetes
R11: If (NP is H) AND (OGTT is H) AND (DBP is H) AND (TSHT is H) AND (2HSI H) AND (BMI is H) AND (DPF is H) AND (Age is H) Then Diabetes
R12: If (NP is H) AND (OGTT is H) AND (DBP is H) AND (TSHT is H) AND (2HSI is L) AND (BMI is H) AND (DPF is L) AND (Age is H) Then Diabetes

5.2.2. Rules with feature selection

R1: If (OGTT is L) AND (BMI is L) AND (Age is L) Then Non-Diabetes
R2: If (OGTT is L) AND (BMI is H) AND (Age is L) Then Non-Diabetes
R3: If (OGTT is H) AND (BMI is L) AND (Age is L) Then Diabetes
R4: If (OGTT is H) AND (BMI is H) AND (Age is L) Then Diabetes
R5: If (OGTT is H) AND (BMI is H) AND (Age is H) Then Diabetes

Note: L=LOW, H=HIGH.

Table 2
Performance of the Re-RX algorithm for the PID dataset (average of 10 runs of 10-fold CV).

Pima Indian Diabetes	TR ACC (%)	TS ACC (%)	# Rules	Ave. # ante.	AUC	TR ACC (SD)	TS ACC (SD)
Re-RX with C4.5	82.80	80.00	12.5	3.07	0.769	1.00	1.10

PID: Pima Indian Diabetes; CV: cross validation; Re-RX: Recursive-Rule eXtraction; TR: training dataset; TS: testing dataset; ACC: accuracy; Ave. # ante.: average number of antecedents; AUC: area under the receiver operating characteristic curve; SD: standard deviation.

Table 3
Performance of previous rule extraction algorithms for the PID dataset.

Rule extraction method [validation method]	TR ACC (%)	TS ACC (%)	# Rules	Rule set	Total Ave. # ante.	# ante.	Year Refs.
Fuzzy rule-based classifier [5CV]	73.31	73.05	11.2 (FR)	No	40 FS	3.57	2003 [31]
Fuzzy classifier [Max. ACC]	78.2	77.0	3 (FR)	Yes	6 FS	2.0	2004 [30]
Neuro-fuzzy system [Max. ACC]	80.08	78.26	55 (FR)	No	–	–	2006 [29]
Granular computing [4CV]	91.97	78.78	5 (FR)	No	–	–	2006 [26]
Evolutionary algorithm [averaged over 10 runs]	77.9 $\pm$ 1.1	77.3 $\pm$ 0.7	Possible (FR)	Possible	–	–	2007 [43]
Fuzzy modeling [5CV]	–	77.65	125 (FR)	No	–	–	2008 [13]
Fuzzy rule extraction [averaged over 10 runs]	–	92.26	8 (FR)	Yes	64FS	8.0	2009 [12]
Evolving fuzzy rule-based classification [Max. ACC]	–	79.37	7 (FR)	Yes	56FS	8.0	2010 [16]
Particle swarm optimization [10CV]	–	88.70	19.3	No	–	–	2011 [27]
Distributed genetic algorithm [Max. ACC]	–	94.0	125	No	–	–	2011 [28]
Fuzzy rules extraction [Max. ACC]	–	73.47	30 (FR)	No	–	–	2013 [15]
Fuzzy decision tree [Max. ACC]	–	71.23	5.8 (FR)	No	–	–	2013 [14]
Artificial bee colony [5CV]	84.20	84.21	7.1 (FR)	**Yes**	–	3.7	2013 [1]
Swarm intelligence [5 $\times$ 2CV]	–	82.03	56	No	–	–	2015 [21]
Sampling Re-RX with J48graft [10 $\times$ 10CV]	84.97	**83.83**	**8.21**	**Yes**	13	**2.01**	**Present study**

PID: Pima Indian Diabetes; Re-RX: Recursive-Rule eXtraction; TR: training dataset; TS: testing dataset; ACC: accuracy; Ave. # ante.: average number of antecedents; Total # ante.: total number of antecedents: 10CV: 10-fold cross validation: 4CV: 4-fold cross validation: 10 $\times$ 10CV: 10 runs of 10-fold cross validation; 5 $\times$ 2CV: 5 runs of 2-fold cross validation; FR: fuzzy rule.

5.3. Rules extracted for the PID dataset using evolving fuzzy rule-based classification [16]

R1: If (NP is Around 5.000) AND (OGTT is around 88.000) AND (DBP is around 66.000) AND (TSFT is around 21.000) AND (2HSI is around 23.000) AND (BMI is 24.400) AND (DPF is around 0.342) AND (AGE is around 30.000) Then Non-Diabetes
R2: If (NP is around 13.000) AND (OGTT is around 145.000) AND (DBP is around 82.000) AND (TSFT is around 19.000) AND (2HSI is around 110.000) AND (BMI is 24.200) AND (DPF is around 0.245) AND (AGE is around 57.000) Then Non-Diabetes
R3: If (NP is around 6.000) AND (OGTT is around 183.000) AND (DBP is around 94.000) AND (TSFT is around 0.000) AND (2HSI is around 0.000) AND (BMI is 40.800) AND (DPF is around 1.461) AND (AGE is around 45.000) Then Non-Diabetes
R4: If (NP is around 8.000) AND (OGTT is around 167.000) AND (DBP is around 106.000) AND (TSFT is around 46.000) AND (2HSI is around 231.000) AND (BMI is 37.600) AND (DPF is around 0.165) AND (AGE is around 43.000) Then Diabetes
R5: If (NP is around 4.000) AND (OGTT is around 189.000) AND (DBP is around 110.000) AND (TSFT is around 31.000) AND (2HSI is around 0.000) AND (BMI is 28.500) AND (DPF is around 0.680) AND (AGE is around 37.000) Then Non-Diabetes
R6: If (NP is around 6.000) AND (OGTT is around 165.000) AND (DBP is around 68.000) AND (TSFT is around 26.000) AND (2HSI is around 168.000) AND (BMI is 33.600) AND (DPF is around 0.631) AND (AGE is around 49.000) Then Non-Diabetes
R7: If (NP is around 13.000) AND (OGTT is around 153.000) AND (DBP is around 88.000) AND (TSFT is around 37.000) AND (2HSI is around 140.000) AND (BMI is 40.600) AND (DPF is around 1.174) AND (AGE is around 39.000) Then Non-Diabetes

5.4. Rules extracted for the PID dataset using fuzzy rule extraction [12]

	NP	OGTT	DBP	TSFT	2HSI	BMI	DPF	Age	Class
R1	High	High	Low	High	High	High	High	Low	Diabetes
R2	Low	Low	High	Low	Low	Low	Low	High	Non-Diabetes
	NP	OGTT	DBP	TSFT	2HSI	BMI	DPF	Age	Class
R3.1	Med-high	Med-high	Med-high	Med-low	Med-high	Med-high	Med-high	Med-low	Diabetes
R3.2	Med-low	Med-low	Med-low	Med-high	Med-low	Med-low	Med-low	Med-high	Non-Diabetes
	NP	OGTT	DBP	TSFT	2HSI	BMI	DPF	Age	Class
R3.3.1	Med-med-high	Med-med-low	Med-med-high	Med-med-low	Med-med-high	Med-med-high	Med-med-high	Med-me-low	Diabetes
	NP	OGTT	DBP	TSFT	2HSI	BMI	DPF	Age	Class
R3.3.2	Med-med-low	Med-med-high	Med-med-low	Med-med-high	Med-med-low	Med-med-low	Med-med-low	Med-med-high	Non-Diabetes
	NP	OGTT	DBP	TSFT	2HSI	BMI	DPF	Age	Class
R3.3.3.1	Med-med-med-high	Med-med-med-high	Med-med-med-high	Med-med-med-high	Med-med-med-high	Med-med-med-high	Med-med-med-high	Med-med-med-high	Diabetes
	NP	OGTT	DBP	TSFT	2HSI	BMI	DPF	Age	Class
R3.3.3.2	Med-med-med-low	Med-med-med-low	Med-med-med-low	Med-med-med-low	Med-med-med-low	Med-med-med-low	Med-med-med-low	Med-med-med-low	Non-diabetes

5.5. Rules extracted for the PID dataset using fuzzy classifier [30]

R1: If (OGTT is H) AND (BMI is H) Then Diabetes
R2: If (OGTT is H) AND (BMI is L) Then Non-Diabetes
R3: If (OGTT is L) AND (BMI is L) Then Non-Diabetes

5.6. Comparison of the extracted rules in the present study with those from four previous rule extraction algorithms

Regarding the accuracy, the number of rules, and the average number of antecedents, three rules with an average number of 2.0 antecedents extracted using fuzzy classifier [30] achieved an accuracy of 77.0%. Based on this performance, this rule set can be considered relatively good.

However, two membership functions (OGTT and BMI) in the antecedents were not accurately depicted, and no descriptions regarding how to define these two fuzzy sets could be found in the literature [30]. This suggests that engineers and/or scientists should construct the fuzzy classifier in cooperation with physicians as domain experts.

Furthermore, no descriptions could be found regarding attribute (feature) selection similar to rules with feature selection in Section 5.2. The default number of attributes in antecedents of rules extracted by fuzzy classifier is eight, which is identical to the number of attributes in the PID dataset; therefore, rules extracted in Section 5.5 may be intuitively interpretable, but also overly subjective.

Rules extracted in Sections 5.3 and 5.4 were seemingly too complicated. Although the rules with feature selection in Section 5.2 achieved approximately the same level of accuracy and number of rules as those from the present study shown in Section 5.1, the average number of antecedents was considerably larger. Therefore, these extracted rules are less interpretable than the rules obtained using the proposed algorithm.

Table 4
Accuracy obtained using sampling Re-RX with J48graft compared with other classifier systems.

Author (Year) [Refs.]	Method	Classification accuracy (%)
Luukka (2007) [34]	Similarity Classifier using PCA and Entropy Optimization	75.82
Polat and Gűneş (2007) [18]	Fuzzy-Artificial immune recognition system [10CV]	84.42
Polat and Gűneş (2008) [20]	PCA+ANFIS [10CV]	89.47
Ghazavi and Liao (2008) [13]	Fuzzy Modeling with Selected Features	77.65
Polat et al. (2008) [35]	Generalized discriminant analysis-Least square-SVM [10CV]	82.05
Kahramanli and Allahverdi (2008) [36]	ANN+FNN [10CV]	84.24
Temurtas et al. (2009) [37]	Multilayer NN with Levenberg-Marquardt algorithm [10CV]	79.62
Übeyli (2009) [17]	Modified Mixture of Experts	99.17
Patil et al. (2010) [24]	Hybrid Prediction Model with Simple K-means Clustering [10CV]	92.38
Übeyli (2010) [38]	Adaptive Neuro-Fuzzy Inference Systems [ANFIS]	98.14
Őrkcű and Bal (2011) [39]	Real-coded Genetic Algorithm [10CV]	77.60
Luukka (2011) [40]	Similarity Classifier+Feature Extraction	75.97
Isa et al. (2011) [41]	Clustered-Hybrid MLP [10 × 5CV]	80.59
Ozcift and Gulten (2011) [33]	Rotation Forest Ensemble Classifier [leave-one-out 10CV]	74.47
Aslam et al. (2013) [23]	Genetic Programming+K-Nearest Neighbor [10CV]	80.50
Seera and Lim (2014) [22]	Fuzzy-Max-Min NN-CART-Random Forest [10CV]	78.39
Yilmaz et al. (2014) [8]	Modified K-Means Clustering+SVM [10CV]	96.71
Gürbüz et al. (2014) [9]	Adaptive Support Vector Machine	97.39
Belle et al. (2014) [42]	Radial Basis Function Classifier	76.70
Wang et al. (2015) [44]	Improved Electromagnetism-like Mechanism [10CV]	77.21
Seera et al. (2015) [45]	Hybrid Fuzzy ARTMAP-CART model [10CV]	87.64
Zhu et al. (2015) [4]	Multiple Factors Weighted Combination [5CV]	≒93
Purwar and Singh (2015) [25]	Hybrid prediction model with Missing Value Imputation [10CV]	99.82
Ding et al. (2015) [7]	Extreme Learning Machine	77.63
Mohapatra et al. (2015) [46]	Improved Cuckoo Search based Extreme Learning Machine	78.50
Feng et al. (2015) [47]	Variable Coded Hierarchical Fuzzy Classification	79.17
Sampling Re-RX with J48graft [10 × 10CV]		
Present study		**83.83**

PID: Pima Indian Diabetes; ReRX: Recursive-Rule eXtraction; MLP: Multilayer Perceptron; 10CV: 10-fold cross validation; 10 × 10CV: 10 runs of 10-fold cross validation; 5 × 2CV: 5 runs of 2-fold cross validation; PCA: Principal Component Analysis; FNN: Fuzzy Neural Network

Consequently, we believe that the present rules extracted by sampling Re-RX with J48graft achieved excellent performance (83.83% accuracy, an average of 8.21 concise rules and 2.01 antecedents).

5.7. Comparison of the classification accuracy in the present study with other classifier systems for the PID dataset

We reviewed the performance of previous rule extraction algorithms in terms of classification accuracy and number of extracted rules for the PID dataset (Table 3).

Medical datasets typically include incomplete data due to missing values in the attributes. Missing values can result from various reasons, such as human error during manual data entry, equipment errors, or incorrect measurements. Missing values in data mining can lead to several problems in the knowledge extraction process, including inefficiency, complications in managing and analyzing the data, and bias due to differences between the missing and complete data.

Therefore, we compared the classification accuracy obtained using sampling Re-RX with J48graft in the present study with that obtained by other classifier systems, some of which conduct preprocessing for filtering and/or imputing missing data. We reviewed classifier systems reported since 2007 and tabulated their performances in Table 4.

Table 4 shows a comparison of studies that carried out k-fold CV to measure classification accuracy. The performance of our proposed sampling Re-RX with J48graft as a classifier achieved substantially better classification accuracy on average than the previous classifiers.

Generally, rule extraction algorithms attempt to achieve both highly accurate and highly concise extracted rules with a well-balanced trade-off. Strictly in terms of classification accuracy, sampling Re-RX with J48graft may not be superior to recent high performance classifiers.

In 2007, Luukka [34] examined the appropriateness of a similar classifier for diagnosis of the PID dataset. Principal Component Analysis (PCA) was used for data preprocessing, and the entropy minimization method was used as a dimension reduction method; these were tested with the classifier.

In addition, we reviewed all recent high performance classifiers that achieved at least 90% classification accuracy for the PID dataset using 10-fold CV.

In 2014, Yilmaz et al. [8] devised a new data preparation method for diagnosis of the PID based on clustering algorithms. In this study, we used a modified k-means algorithm to eliminate noise and inconsistent data, and SVMs for classification. This newly developed approach was tested in the diagnosis of the PID. A classification accuracy of 96.71% was obtained using 10-fold CV.

In 2015, Zhu et al. [4] proposed a dynamic weighted voting scheme referred to as multiple factors weighted combination (MFWC) for decision combination in a multiple classifier system. In contrast to other methods, their dynamic weighting method considered the local accuracy factor for each classifier and used a validation set to estimate classification accuracy at the global level. In addition, because the generalization error of a classifier is a key function for measuring its performance generalized to unseen samples, their method also considered the relationship between training and testing samples to involve a generalization error. They obtained a classification accuracy of about 93% for the PID dataset using 10-fold CV.

In 2015, Purwar and Singh [25] presented a novel hybrid prediction model with missing value imputation (HPM-MI). Their model used simple k-means clustering to analyze various imputation techniques, and applied the best one to a dataset. Their proposed hybrid model was the first to use a combination of k-means clustering and a multilayer perceptron. Before applying the classifier, they used k-means clustering to validate the class labels of given data (incorrectly classified instances were deleted, i.e., extracted from original data). As a result, the quality of the data

was significantly improved. The efficiency of their model as a predictive classification system was then investigated using the PID dataset. The results showed HPM-MI achieved an accuracy of 99.82%, making it the most accurate compared with the existing methods.

6. Discussion

In Section 6.1, we explain the role of OGTT and BMI in the diagnosis of the PID dataset. Next, in Section 6.2, we discuss the medical informatics interpretation of the rules extracted in the present study, and in Sections 6.3 and 6.4, we address two kinds of important trade-off issues. We also discuss the significance of the present rules extracted by sampling Re-RX with J48graft.

6.1. Role of OGTT and BMI for diagnosis of the PID dataset

Traditionally, routine screening for diabetes has been challenging, both in primary practice and the community. No global consensus on the optimal screening strategy for diabetes has been reached. Although fasting plasma glucose (FPG) is commonly used in screening for diabetes; however, this measurement varies widely; therefore, OGTT remains the most valid tool for diagnosing diabetes.

As shown in Section 3.7, the PID dataset does not include Hemoglobin A1c (HbA1c) as an attribute; therefore, we cannot extract the most important attribute to diagnose T2DM according to the ADA Diabetes Guidelines [71]. In the 1980s, the measurement of HbA1c became routine in patients known to have diabetes, and it has been suggested that this test could supplant the measurement of blood or FPG as the diagnostic tool.

In fact, nearly all research conducted in relation to the complications and/or treatment of diabetes heavily cites the results from the 1998 UK Prospective Diabetes Study (UKPDS) [72,73] and the 1993 Diabetes Control and Complications Trial (DCCT) [74]. This demonstrates the huge impact of the UKPDS in the field [75].

The proposed diagnostic cutoff of 6.5% for HbA1c [71] is a value that most studies have shown would lead to a diabetes prevalence equivalent to that using FPG, so fewer patients will be newly diagnosed if HbA1c at this level is used alone. OGTT is one option [71] as a criterion of diabetes diagnosis written as 2-hour post glucose (2-h FPG) $\geq$ 200 mg during OGTT (75 g).

Overall, only 25% of individuals with diabetic OGTT had an HbA1c $\geq$ 6.5%, while 45% of individuals who exceeded both the FPG and OGTT criteria (1% of the entire population) were not diagnosed with diabetes using HbA1c [76].

In fact, HbA1c $<$ 5.8 alone is sometimes diagnosed using OGTT, whose values vary between non-diabetes, borderline (prediabetes) and diabetes, i.e., they sometimes overlap. Prediabetes does not belong to diabetes or non-diabetes, i.e., impaired glucose tolerance (IGT) is defined as FPG $<$ 100 mg/dL and OGTT $\geq$ 140 and $<$ 199 mg/dL, while impaired fasting glycaemia (IFG) is defined as FPG $\geq$ 100 and $<$ 125 mg/dL and OGTT $<$ 140 mg/dL or a complication of IGT and IFG.

The transition from early metabolic abnormalities that precede diabetes such as IFG and IGT to diabetes may take years; however, current estimates indicate that most individuals in a prediabetic state eventually develop diabetes. The complications of diabetes, which are the major causes of morbidity and mortality, are related to its duration, chronic level of glycemia, and other risk factors. Clinical trials have demonstrated the effectiveness of intensive glycemic and blood pressure control to reduce the long-term complications of diabetes; however, the public health burden of the disease remains substantial [77].

As shown in Section 5.1, we extracted seven rules which include two important attributes, i.e., OGGT and BMI. Thus, we describe the importance of OGGT and BMI in the diagnosis of T2DM as follows.

Cavagnolli et al. [78] reported that lowering the HbA1c cutoff level and adding a glucose-based method improved HbA1c performance in the diagnosis of diabetes. This suggested that each method identifies different patient populations. In their results, HbA1c $\geq$ 6.5% showed high specificity, but limited sensitivity, to a diabetes diagnosis. This suggests that a cutoff point of $\geq$ 6.5% would not be sufficient to diagnose diabetes. The use of HbA1c as the sole diagnostic test for diabetes should be approached with caution to assure the correct classification of diabetics.

Hayashi et al. [79] reported that the insulin concentration pattern during an OGTT is a strong predictor of future T2DM among Japanese-Americans. Although many of these patterns were associated with insulin sensitivity and secretion, independent associations were seen with the incidence of diabetes. Therefore, the OGTT pattern of insulin concentration may represent a useful adjunct in the prediction of future T2DM.

While BMI is not included in the four options of the ADA guidelines [71], testing for T2DM and prediabetes in asymptomatic adults is recommended in all adults who are overweight or obese (BMI $\geq$ 25 or $\geq$ 23 in Asian-Americans) and who have more than one risk factor as defined by the ADA [71].

Araneta et al. [80] recently proposed that the BMI cutoff point for identifying Asian-Americans who should be screened for undiagnosed T2DM should be $<$ 25, and $\geq$ 23 may be the most practical. A similar result was reported by Hsia et al. [81].

Boffeta et al. [82] conducted a cross-sectional pooled analysis of 900,000 individuals in the Asia Cohort Consortium and estimated the shape and the strength of the association between BMI and the prevalence of diabetes in Asian populations. They also identified patterns of association by age, country, and other risk factors for diabetes.

Therefore, we believe that OGTT and BMI can be included as attributes in the antecedent of rules extracted for diagnosis of the PID dataset.

6.2. Interpretation of rules extracted by the proposed algorithm from medical informatics view

The ADA criteria for the diagnosis of diabetes has four options that include four important attributes, i.e., HbA1c, FPG, OGTT and random PG. Due to the constraints of the PID dataset, we could only two attributes, i.e., OOGT and BMI, so we attempted to explore diagnostic rules for primarily borderline type T2DM or prediabetes.

We explained the reason why the present extracted rules achieved very good performance in Section 5.6. In this section, we attempt to interpret how seven rules play a role in the diagnosis of borderline type T2DM or prediabetes as follows.

R1 is the safety-side rule for diagnosing non-diabetes. The main purpose of the ADA guidelines is to detect diabetes. The safety-side criterion may not be explicitly indicated. R1 is useful for providing clear relief to patients after laboratory tests.

R2 and R4 state that an OGTT of 139 and a BMI of 36.1 are critical cutoff points. In the same range of OGTT, if BMI is $>$ 36.1, then diabetes is diagnosed as shown in R4.

R3 states the upper limit of BMI for non-diabetes. If BMI is $>$ 39.6, then it is definitely diabetes, regardless of the OGTT, because the OGTT range is very limited (125–129).

R5 states the upper limit of OGTT for non-diabetes. Thus, in this case, BMI must be $\leq$ 28.1. This value is also very critical in R6 and R7.

R6 states that even if OGTT is >151, then it is diabetes with a relatively good BMI value of ≤ 28.6.
R7 states that if OGTT is >139, then it is non-diabetes only in the case where BMI ≤ 28.6; otherwise, it is diabetes.

We think that these rules can be applied to the diagnosis of the PID dataset. We hope that the proposed algorithm could also be adapted to similar T2DM datasets that include HbA1c, FPG, and random PG to extract diagnostic rules.

6.3. Trade-off between the accuracy and the number of extracted rules

In multi-objective optimization and economics, the so-called Pareto optimality (ideally balanced trade-off) is always an important issue. In the case of medical rule extraction, there is a trade-off between high diagnostic accuracy and the interpretability of extracted rules. Thus, if a physician wishes to extract rules with high diagnostic accuracy from medical datasets, they can choose the algorithm with high diagnostic accuracy but reduced interpretability. However, in other situations, a physician may want to obtain extracted diagnostic rules with reduced accuracy and more interpretability.

Needless to say, if the optimal solution (best trade-off) can be found, then the best extracted rules can be obtained. Ideally, we hope to extend the Pareto optimal curve to obtain a wider viable region that provides improvements in both diagnostic accuracy and interpretability.

Recently, Fortuny and Martens [83] expressed the same opinion: Rule extraction is a technique that attempts to find compromise between both requirements by building a simple rule set that mimics how the well-performing complex model (black-box) makes decisions.

As described in Section 4.1, even if all types of rules are considered, the consensus is that no matter how they are defined, an ideal measure has yet to be developed; therefore, “what is a concise rule? ” remains a difficult question to answer.

However, we believe that this perspective is very important in comparing the quality of rules extracted from the PID dataset. As shown in Table 3, 11 fuzzy rule extraction algorithms were proposed for the PID dataset. In contrast, only three concise rule extraction algorithms were proposed for the same dataset.

To date, these two kinds of rules have been incomparable in terms of quality of rules extracted for the PID dataset. In the design of both fuzzy classifier and rule extraction algorithms, the interpretability of the rule set has been considered an important factor. This interpretability is measured by calculating the number of rules; fuzzy classifiers containing fewer fuzzy rules are always more interpretable than those with more fuzzy rules [1].

However, many types of rules have been suggested in the literature. Propositional rules take the form of IF-THEN expressions, where clauses are defined in propositional or fuzzy logic. The trade-off between the accuracy and the number of rules also needs to be balanced.

To allow a better understanding of our claim, a Pareto optimal (the best trade-off) curve between the accuracy and the number of rules extracted is shown in Fig. 3. The reciprocal of the number of rules extracted is shown on the x-axis. The red dot, which is located at the trade-off curve, shows the performance of the proposed algorithm. This shows that the present algorithm provided extracted rules for the PID dataset that were both accurate and concise.

The four green dots obtained by artificial bee colony [1], fuzzy rule extraction [12], granular computing [26] and fuzzy classifier [30] may provide better accuracy and/or the number of rules than that of the present rules.

However, in general, fuzzy rules involve strong expressive power by linguistic and intuitive expressions. Thus, the number of fuzzy rules is not equivalent to the same number of concise rules in terms of expressive power. On the contrary, the number of fuzzy rules should be considered much more important than the number of concise rules.

Considering the potential for the more expressive power of fuzzy rules, all of the green dots for fuzzy rules should be shifted horizontally to the left, which result in being beyond the trade-off curve.

Consequently, the red dot obtained by the proposed algorithm is the closest to the trade-off curve, and provides well balanced performance between accuracy and the number of rules.

6.4. Trade-off between the accuracy and the average number of antecedents in one rule

In this manner, propositional rules take the form of IF-THEN expressions, in which clauses are defined in propositional or fuzzy

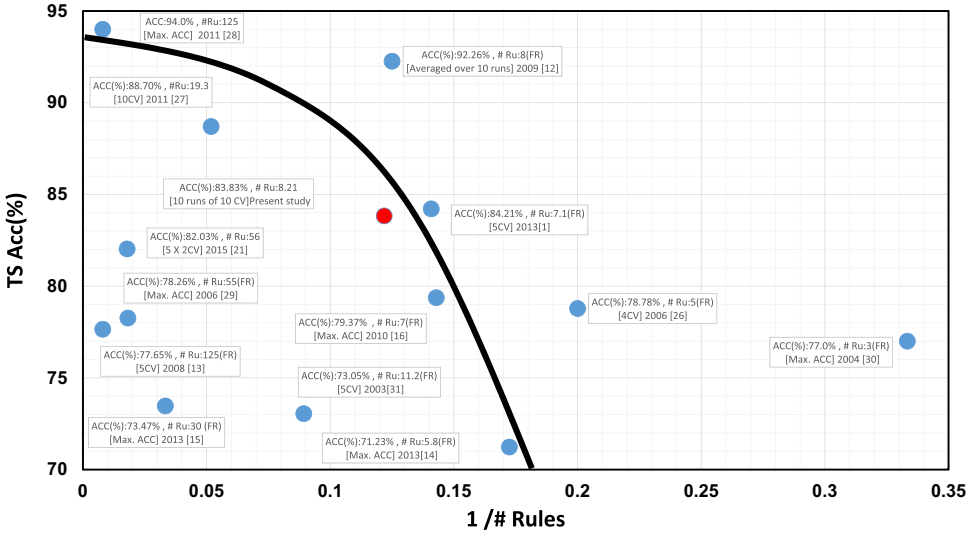

Fig. 3. Trade-off curve between the accuracy and the number of rules extracted. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

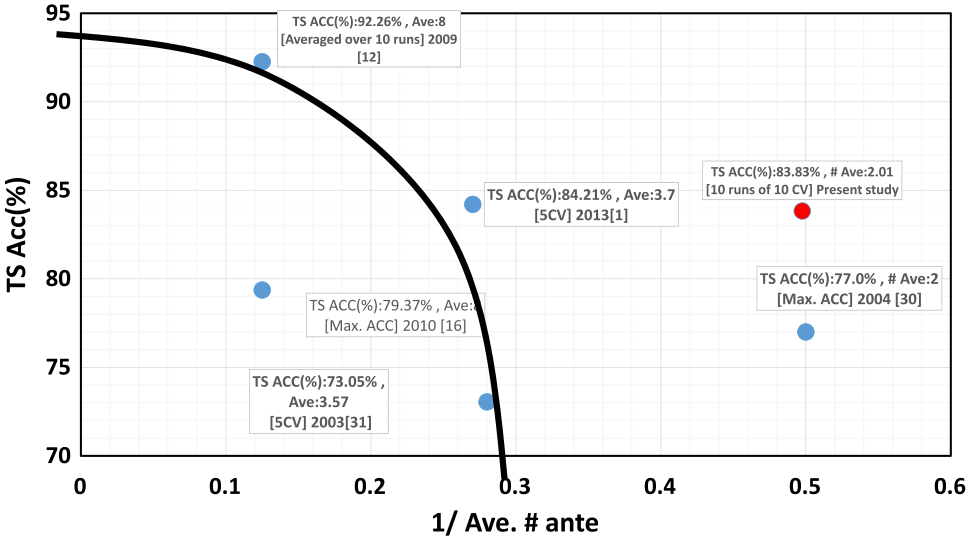

Fig. 4. Trade-off curve between the accuracy and the number of antecedents.

logic. The trade-off between the accuracy and the average number of antecedents also needs to be balanced.

To allow a better understanding of our claim, a Pareto optimal (the best trade-off) curve between the accuracy and the number of antecedents is shown in Fig. 4. The reciprocal of the average number of antecedents is shown on the x-axis.

Fuzzy rules consist of many antecedents that use fuzzy sets defined by membership functions. Considering the potential for the more expressive power of fuzzy rules, all of the green dots for fuzzy rules should be shifted horizontally to the left, which result in being beyond the trade-off curve.

The red dot obtained by the proposed algorithm is located in a wider viable region that provides substantially more improvement in both accuracy and the average number of antecedents than the other previous algorithms.

7. Conclusions

In this paper, we proposed sampling Re-RX with J48graft as a new algorithm for extracting highly accurate, concise, and interpretable rules for the PID dataset. We also demonstrated that the extracted rules based on the two kinds of trade-offs, were more accurate, concise, and interpretable, and therefore more suitable for medical decision making. Actually, high accuracy, conciseness, and interpretability are achieved simultaneously by the proposed sampling Re-RX with J48graft algorithm for the PID dataset.

Although the attributes of the PID dataset are substantially different from the attributes of the current T2DM dataset [67], which includes HbA1c, FPG, and random PG, we think that sampling Re-RX with J48graft provides better clinical information regarding T2DM. Specifically, we investigated how OGGT and BMI values may interact with extracted rules to predict T2DM. The use of sampling Re-RX with J48graft is expected to be particularly useful in patients with T2DM whose fracture risk is relatively high.

Needless to say, the diagnosis of T2DM remains a complex problem; therefore, sampling Re-RX with J48graft should be tested on more recent and complete diabetes datasets in future studies in order to ensure that the most highly accurate rules can be extracted for diagnosis.

References

[1] Beloufa F, Chikh MA. Design of fuzzy classifier for diabetes disease using modified artificial bee colony algorithm. Comput Methods Prog Biomed 2013;112:92–103.
[2] Marateb HR, Mansourian M, Faghihimani E, Amini M, Farina D. A hybrid intelligent system for diagnosing microalbuminuria in type 2 diabetes patients without having to measure urinary albumin. Comput Biol Med 2014;45:34–42.
[3] Centers for Disease Control and Prevention. National Diabetes Statistics Report: Estimate of Diabetes and its Burden in the United States, 2014. Atlanta, GA: Department of Health and Human Services,; 2014.
[4] Zhu J, Xie Q, Zheng K. An improved early detection method of type-2 diabetes mellitus using multiple classifier system. Inf Sci 2015;292:1–14.
[5] Hunter DJ. Gene-environment interactions in human diseases. Nat Rev Genet 2005;6:287–98.
[6] Homme MB, Reynolds KK, Valdes R, Linder MW. Dynamic pharmacogenetic models in anticoagulation therapy. Clin Lab Med 2008;28:539–52.
[7] Ding S, Zhao H, Zhang X, Xu X, Nie R. Extreme learning machine: algorithm, theory and applications. Artif Intell Rev 2015;44:103–15.
[8] Yilmaz N, Inan O, Uzer MS. A new data preparation method based on clustering algorithms for diagnosis systems of heart and diabetes diseases. J Med Syst 2014;38:48–59.
[9] Gürbüz E, Kılıç E. A new adaptive support vector machine for diagnosis of diseases. Expert Syst 2014;31:389–97.
[10] Shanker MS. Using neural networks to predict the onset of diabetes mellitus. J Chem Inf Comput Sci 1996;36:35–41.
[11] Park J, Edington DW. A sequential neural network model for diabetes prediction. Artif Intell Med 2001;23:277–93.
[12] Gadaras I, Mikhailov L. An interpretable fuzzy rule-based classification methodology for medical diagnosis. Artif Intell Med 2009;47:25–41.
[13] Ghazavi SN, Liao TW. Medical data mining by fuzzy modeling with selected features. Artif Intell Med 2008;43:195–206.
[14] Liu X, Feng X, Pedrycz W. Extraction of fuzzy rules from fuzzy decision trees: an axiomatic fuzzy sets (AFS) approach. Data Knowl Eng 2013;84:1–25.
[15] Chavas ADF, Vallasco MMBR, Tanscheit R. Fuzzy rules extraction from support vector machines for multi-class classification. Neural Comput Appl 2013;22:1571–80.
[16] Lekkas S, Mikhailov L. Evolving fuzzy medical diagnosis of Pima Indians diabetes and of dermatological disease. Artif Intell Med 2010;50:117–26.
[17] Übeyli ED. Modified mixture of experts for diabetes diagnosis. J Med Syst 2009;33:299–305.
[18] Polat K, Gűnes S. An improved approach to medical data sets classification: artificial immune recognition system with fuzzy resource allocation mechanism. Expert Syst 2007;24:252–70.
[19] Chikh MA, Saidi M, Settouti N. Diagnosis of diabetes diseases using an Artificial Immune Recognition System2 (AIRS2) with fuzzy K-nearest neighbor. J Med Syst 2012;36:2721–9.
[20] Polat K, Gűnes S. An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes diseases. Digit Signal Process 2007;17:702–10.
[21] Christopher JJ, Nehemiah HK, Kannan A. A swarm optimization approach for clinical knowledge mining. Comput Methods Prog Biomed 2015;121:137–48.
[22] Seera M, Lim CP. A hybrid intelligent system for medical data classification. Expert Syst Appl 2014;41:2239–49.

[23] Aslam MW, Zhu Z, Nandi AK. Feature generation using genetic programming with comparative partner selection for diabetes classification. Expert Syst Appl 2013;40:5402–12.
[24] Patil BM, Joshi RC, Toshniwal D. Hybrid prediction model for type-2 diabetic patients. Expert Syst Appl 2010;37:8102–8.
[25] Purwar A, Singh SK. Hybrid prediction model with missing value imputation for medical data. Expert Syst Appl 2015;42:5621–31.
[26] Su CT, Chen LS, Yih Y. Knowledge acquisition through information granulation for imbalanced data. Expert Syst Appl 2006;31:531–41.
[27] Özbaki L, Delice Y. Exploring comprehensible classification rules from trained neural networks integrated with a time-varying binary particle swarm optimizer. Eng Appl Artif Intell 2011;24:491–500.
[28] Rodríguez M, Escalante DM, Peregrín A. Efficient distributed genetic algorithm for rule extraction. Appl Soft Comput 2011;11:733–43.
[29] Gonçalves LB, Vellasco MMBR, Pacheco MAC, Souza FJ. Invited hierarchical neuro-fuzzy model for pattern classification and rule extraction in diabetes. IEEE Trans Syst Man Cyber-Part C: Appl Rev 2006;36:236–48.
[30] Chang X, Lilly JH. Evolutionary design of a fuzzy classifier from data. IEEE Trans Syst Man Cyber-Part B: Cyber 2004;34:1894–906.
[31] Abonyi J, Roubos JA, Szeifert F. Data-driven generation of compact, accurate, and linguistically sound fuzzy classifiers based on a decision-tree initialization. Int J Approx Reason 2003;32:1–21.
[32] Salari N, Shohaimi S, Najafi F, Nallappan M, Karishnarajah I. A novel hybrid classification model of genetic algorithms, modified k-nearest neighbor and developed backpropagation neural network. PLoS One 2014;9:e112987.
[33] Ozcift A, Gulten A. Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. Comput Methods Prog Biomed 2011;104:443–51.
[34] Luukka P. Similarity classifier using similarity measure derived fromYu's norms in classification of medical data sets. Comput Biol Med 2007;37:1133–40.
[35] Polat K, Gűnes S, Arslan A. A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine. Expert Syst Appl 2008;34:482–7.
[36] Kahramanli H, Allahverdi N. Design of a hybrid system for the diabetes and heart diseases. Expert Syst Appl 2008;35:82–9.
[37] Temurtas H, Yumusak N, Temurtas F. A comparative study on diabetes disease diagnosis using neural networks. Expert Syst Appl 2009;36:8610–5.
[38] Übeyli ED. Automatic diagnosis of diabetes using adaptive neuro-fuzzy inference systems. Expert Syst 2010;27:259–66.
[39] Örkcü HH, Bal H. Comparing performances of backpropagation and genetic algorithms in the data classification. Expert Syst Appl 2011;38:3703–9.
[40] Luukka P. Feature selection using fuzzy entropy measures with similarity classifier. Expert Syst Appl 2011;38:4600–7.
[41] Isa NAM, Mamat WMFW. Clustered-Hybrid Multilayer Perceptron network for pattern recognition application. Appl Soft Comput 2011;11:1457–66.
[42] Belle VV, Lisboa P. White box radial basis function classifiers with component selection for clinical prediction models. Artif Intell Med 2014;60:53–64.
[43] Hoffmann F, Baesens B, Mues C, Gestel TV, Vanthienen J. Inferring descriptive and approximate fuzzy rules for credit scoring using evolutionary algorithms. Eur J Oper Res 2007;177:540–55.
[44] Wang K-J, Adrian AM, Chen K-H, Wang K-M. An improved electromagnetism-like mechanism algorithm and its application to the prediction of diabetes mellitus. J Biomed Inf 2015;54:220–9.
[45] Seera M, Lim CP, Tan SC, Loo CK. A hybrid FAM–CART model and its application to medical data classification. Neural Comput Appl 2015;26:1799–811.
[46] Mohapatra P, Chakravarty S, Dash PK. An improved cuckoo search based extreme learning machine for medical data classification. Swarm Evol Comput 2015;24:25–49.
[47] Feng T-C, Li T-HS, Kuo P-H. Variable coded hierarchical fuzzy classification model using DNA coding and evolutionary programming. Appl Math Model 2015;39:7401–19.
[48] Napierala K, Stefanowski J. BRACID: a comprehensive approach to learning rules from imbalanced data. J Intell Inf Syst 2012;39:335–73.
[49] Setiono R, Baesens B, Mues C. Recursive neural network rule extraction for data with mixed attributes. IEEE Trans Neural Netw 2008;19:299–307.
[50] Quinlan JR. C4.5: programs for machine learning. Morgan Kaufmann Series in Machine Learning. San Mateo, California: Morgan Kaufman, Inc.; 1993.
[51] Hayashi Y, Tanaka Y, Takagi T, Saito T, Iiduka H, Kikuchi H, Bologna G, Mitra S. Recursive-Rule Extraction algorithm with J48graft and applications to generating credit scores. J Artif Intell Soft Comput Res 2016;6:35–44.
[52] ⟨http://fiji.sc/javadoc/weka/classifiers/trees/J48graft.html⟩; [Last accessed 30.09.15].
[53] Webb GI. Decision tree grafting from the all-tests-but-one partition, In: Proc. 16th international joint conference on artificial intelligence (IJCAI), 2; 1999. p. 702–07.
[54] Setiono R. A penalty-function approach for pruning feedforward neural networks. Neural Comp 1997;9:185–204.
[55] Setiono R. Sampling selection and neural network rule extraction for credit scoring. In: Proceedings of the 43rd decision sciences institutes annual meeting; 2012. p. 1280–90.
[56] Setiono R, Azcarraga A, Hayashi Y. Using sample selection to improve accuracy and simplicity of rules extracted from neural networks for credit scoring applications. Int J Comput Intell Appl 2015;14:1550021-1–20.
[57] University of California, Irvine Learning Repository, ⟨http://archive/ics.uci/edu/m/⟩; [last accessed 01.10.15].
[58] Knowler WC, Bennett PH, Bottazzo GF, Doniach D. Islet cell antibodies and diabetes mellitus in Pima Indians. Diabetologia 1979;17:161–4.
[59] Smith JW, et al. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In: Proc. Twelfth Annual Symposium Comput. Applications Medical Care; 1988. p. 261–65.
[60] Witten IH, Frank E. Data Mining: Practical Machine Learning Tools with Java Implementations. San Mateo, CA: Morgan Kaufmann, Inc,; 1999.
[61] Quinlan JR. Induction of decision trees. Mach Learn 1986;1:81–106.
[62] Webb GI. Decision tree grafting, learning. In: Proc. IJCAI'97 15th international conference on artificial intelligence (IJCAI) 2; 1997. p. 846–85.
[63] Purwar A, Singh SK. Hybrid prediction model with missing value information for medical data. Expert Syst Appl 2015;42:5621–31.
[64] Holt R, Hanley N. Essential endocrinology and diabetes. Malden, MA: Blackwell Publishing,; 2006.
[65] Breault JL. Data mining diabetic databases: are rough sets a useful addition? In: Wegman E, Braverman A, Goodman A, Smyth P, editors. Computing Science and Statistics, 33. Fairfax Station, VA: Interface Foundation of North America; 2001. p. 51–60.
[66] Gagliardi F. Instance-based classifiers applied to medical databases: Diagnosis and knowledge extraction. Artif Intell Med 2011;52:123–39.
[67] Breault JL, Goodall CR, Fos PJ. Data mining a diabetic data warehouse. Artif Intell Med 2002;26:37–54.
[68] Amine Chikh M, Saidi M, Settouti N. Diagnosis of diabetes diseases using an Artificial Immune Recognition System2 (AIRS2) with fuzzy K-nearest neighbor. J Med Syst 2012;36:2721–9.
[69] Salzberg SL. On comparing classifiers: pitfalls to avoid and a recommended approach. Data Min Knowl Disco 1997;1:317–28.
[70] Marqués AI, García V, Sánchez JS. On the suitability of resampling techniques for the class imbalance problem in credit scoring. J Oper Res Soc 2013;64:1060–70.
[71] American Diabetes Association. Standards of medical care in diabetes-2015. Diabetes Care 2015;38:S1–93.
[72] UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet 1998;352:837–53.
[73] UK Prospective Diabetes Study (UKPDS) Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). Lancet 1998;352:854–65.
[74] The DCCT Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Eng J Med 1992;329:977–86.
[75] Geneuth S. The UKPDS and its global impact. Diabet Med 2008;25:57–62.
[76] Kilpatrick ES, Winocour PH. ABCD position statement on haemoglobin A_{1c} for the diagnosis of diabetes. Pr Diabetes 2010;27:1–5.
[77] Nathan DM, Davidson MB, Defronzo RA, Heine RJ, Henly RR, Prately R, Zinman B. Impaired fasting glucose and impaired glucose tolerance. Diabetes Care 2007;30:753–9.
[78] Cavagnolli G, Comerlatot J, Comerlatot C, Renzt PB, Gross JJ, Camargo JL. HbA1c measurement for the diagnosis of diabetes: is it enough? Diabet Med 2011;28:31–5.
[79] Hayashi T, Boyko EJ, Sato KK, McNeely MJ, Leonetti DL, Kahn SE, Fujimoto WY. Patterns of insulin concentration during the OGGT predict the risk of type 2 diabetes in Japanese Americans. Diabetes Care 2013;36:1229–35.
[80] Araneta MRG, Kanaya AM, Hsu WC, Chang HK, Grandinetti A, Boyko EJ, Hayashi T, Kahn SE, Leonetti DL, Mcneely MJ, Onishi Y, Sato KK, Fujimoto WY. Optimum BMI cut points to screen Asian Americans for type 2 diabetes. Diabetes Care 2015;38:814–20.
[81] Hsia DS, Larrivee S, Cefalu WT, Johnson WD. Impact of lowering BMI cut points as recommended in the revised American Diabetes Association's Standards of Medical Care in Diabetes-2015 on diabetes screening in Asian Americans. Diabetes Care 2015;38:2166–8.
[82] Boffeta P, Mclerran D, Chen Y, Inoue M, Sinha R, et al. Body mass index and diabetes in Asia: a cross-sectional pooled analysis of 900,000 individuals in the Asia cohort consortium. Plos One 2011;6:e19930.
[83] Fortuny EJD, Martens D. Active learning-based pedagogical rule extraction. IEEE Trans Neural Netw Learn Syst 2015;26:2664–77.