Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

# Diagnosing diabetes mellitus using machine learning techniques

**Mazen Alzyoud[a*], Raed Alazaidah[b], Mohammad Aljaidi[b], Ghassan Samara[b], Mais Haj Qasem[b], Muhammad Khalid[c] and Najah Al-Shanableh[a]**

[a]*Faculty of Information Technology, Al al-Bayt University, Jordan*
[b]*Faculty of information technology, Zarqa university, Jordan*
[c]*School of computer science, university of Hull, Hull, United Kingdom*

**C H R O N I C L E**

*Article history*:
Received: July 18, 2023
Received in revised format: September 3, 2023
Accepted: October 7, 2023
Available online: October 7, 2023

*Keywords:*
*Classification*
*Diabetes*
*Feature selection*
*Medical diagnosis*
*Prediction*

**A B S T R A C T**

Diabetes Mellitus (DM) is a frequent condition in which the body's sugar levels are abnormally high for an extended length of time. It is a major cause of death with high mortality rates and the second leading cause of total years lived with disability worldwide. Its seriousness comes from its long-term complications, including nephropathy, retinopathy, and neuropathy leading to kidney failure, poor vision and blindness, and peripheral sensory loss, respectively. Such conditions are life-threatening and affect patients' quality of life. Therefore, this paper aims to identify the most relevant features in the diagnosis of DM and identify the best classifier that can efficiently diagnose DM based on a set of relevant features. To achieve this, four different feature selection methods have been utilized. Moreover, twelve different classifiers that belong to six learning strategies have been evaluated using two datasets and several evaluation metrics such as Accuracy, Precision, Recall, F1-measure, and ROC area. The obtained results revealed that the correlation attribute evaluation method would be the best choice to handle the task of feature selection and ranking for the considered datasets, especially when considering the Accuracy metric. Furthermore, MultiClassClassifier would be the best classifier to handle Diabetes datasets, especially when considering True Positive, precision, and Recall metrics.



## 1. Introduction

Diabetes mellitus, generally known as Diabetes, is a condition that affects the hormone insulin, which causes improper glucose metabolism and raises blood sugar levels (Cho et al., 2018). Numerous human body organs are impacted by high blood sugar levels, which in turn hamper many physiological processes, particularly those involving blood vessels and neurons. Although the exact origins of diabetes are still unknown, many experts think that both environmental conditions and inherited elements play a role. In any case, this disease is more common in adults, which is why it is classified as "adult-onset" diabetes. Currently, it is believed that DM accompanies people as they age. Diabetes affected 452 million patients worldwide in 2017, and it's predicted that number will increase to 694 million by 2045 (Cho et al., 2018). Unfortunately, 422 million of the world's 422 million Diabetes sufferers (80.6% of the total) reside in low-income nations. Statistics from the International Diabetes Federation (IDF) show that 141 million people in China (aged 20 to 79) had Diabetes in 2021, and that more than half of them went undiagnosed. The situation with DM is terrible right now (Mao et al., 2022).

Another scientific research has demonstrated the prevalence of DM. It found that half a billion people globally have the disease, and that in 2030 and 2045, the percentages are projected to increase to 25% and 51%, respectively. DM has no long-term cure, but if it is caught early enough, it can be controlled and its complications can be prevented (Khaleel & Al-Bakry, 2023). Current

* Corresponding author.
E-mail address: malzyoud@aabu.edu.jo (M. Alzyoud)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

doi: 10.5267/j.ijdns.2023.10.006

studies and medical professionals agree that early disease detection will improve recovery prospects. Nevertheless, Diabetes's patients suffer from several complication such as diabetic retinopathy, foot problems which might lead to amputation, high possibility for heart attacks and stroke, kidney problems, nerve damage and related conditions like gum disease, and special types of cancer.

With technological advancements, Machine Learning (ML) techniques can be used to predict diseases and diagnose illnesses early (Krishnamoorthi et al., 2022; Dritsas &Trigka, 2022). Asa part of Artificial Intelligence (AI), ML is much more than just a tool for data analysis. Data serves as the system's fuel (Huang et al., 2022). The diseases can be restrained, and human lives can be saved with early disease prediction (Forde et al., 2022). To accomplish this, this research study focuses on investigating how to diagnose DM before it manifests by taking into consideration a number of features. For this investigation, two datasets with different characteristics related to DM have been considered. By building predictive approaches using diagnostic medical datasets gathered from Diabetes patients, ML techniques efficiently extract knowledge. Knowing more about these facts can help predict Diabetes patients. It is possible to forecast DM using a variety of ML techniques. Choosing the optimal method to predict based on these variables is really challenging.

To predict DM using adult population data, we use four well-known feature selection and ranking methods. These methods are: InformationGainAttributeEvaluator (InfoGainAttributeEval.), GainRatioAttributeEvaluator (GainRatioAttributeEval.), ClassifierAttributeEvaluator (ClassifierAttributeEval.), and CorrelationAttributeEvaluator (CorrelationAttributeEval.). Moreover, various evaluation metrics are used here. Examples of these metrics are Accuracy, F1-measure, and ROC Area. Furthermore, 12 different classifiers that belong to six learning strategies have also been utilized to evaluate the considered feature selection and ranking methods. The main contributions of the present paper are summarized as follows:

- To identify the best feature selection and ranking method among four popular and well-known methods.
- To evaluate the predictive performance of twelve different classifiers that belong to six learning strategies in order to utilize them to handle Diabetes datasets.

Finally, the main task of ML considered in this research is the classification task. Classification is defined as the task of predicting the class label for a new sample or case as accurately as possible (Alazaidah et al. 2018, 2020, 2023; Alluwaici, Junoh, & Alazaidah, 2020; Alluwaici, Junoh, Ahmad, Mohsen, & Alazaidah, 2018). It is divided into two main types. The first type is called Single Label Classification (SLC), while the second one is called Multi Label Classification (MLC) (Alluwaici, Junoh, Ahmad, Mohsen, & Alazaidah, 2018). For SLC, instances are linked to one class label, while in the case of MLC, the opposite is correct, since it allows instances to be linked to one or several class labels at the same time. Class labels are mutually exclusive in SLC, while in MLC they are not (Alazaidah et al., 2017; Bose & Ramesh, 2023). For this research, MLC is not considered.

The rest of the paper is organized as follows: The literature is provided in Section II. Then, the proposed methodology, results, and discussion are discussed in Section iii. Finally, Section IV introduces the conclusion and future directions.

## 2. Related Work

To find a solution to the Diabetes prediction issue, it is necessary to examine all the prior work that has been done so far, and then, propose improved ideas. Thus, this section provides a comprehensive overview of the research. Authors in (Zhang, Wei, Ren, Cheng, & Zheng, 2018) recommended using a method that uses Classification and Regression Tree (CART) models. For a CART to be initialized, a decision tree must be generated from the available datasets. In CARS pruning and optimization, the regression tree is cut down to size based on a set of metrics, such as the maximum allowable depth of the tree, the minimum allowable number of leaf samples, and the minimum allowable impurity of any given node. A system proposing the use of K Nearest Neighbors (KNN), data preprocessing, the use of K-means, and the application of a classifying algorithm were proposed by authors in (Kumar & Umatejaswi, 2017). When analyzing medical data that is not evenly distributed, the authors in (Zeng et al., 2016) proposed using the Synthesis Minority Oversampling Method (SMOTE) as an efficient data sampling approach. The notion of using fuzzy logic to predict tasks is credited to the authors of (Liu, Zhang, Xiang & Zhou, 2017).

Fuzzy-based Information Decomposition (FID) rebalancing the train data and making more examples for the classes that are not in the majority. Both weighing and recuperation are involved in these procedures. It was recommended to employ several dimensions of data (Arsyadani & Purwinarko, 2023). Data overfitting in the predictive model is a new approach provided by (Ashiquzzaman et al., 2018). A system that compares machine learning and deep learning-based algorithms for Diabetes prediction has been provided in (Yahyaoui, Jamil, Rasheed &Yesiltepe, 2019) where Random Forest outperforms Support Victor Machine (SVM) and deep learning which are also employed. After looking at how well different combinations of machine learning algorithms work, Hasan et al. (2020) devised an ensemble approach for predicting Diabetes. The combination of AdaBoost and XgBoost gave the best results. Here, they introduce a framework for Diabetes prognosis that uses outliers, missing values, data normalization, feature selection, cross-validation, and several classifiers in addition to a Multi-Layer perceptron (MLP). The method presented in (Aliberti et al., 2019) involves first using the prediction models to infer future glucose level values on a new patient, and then, checking the accuracy of the models' using data of glucose signals from a large and diverse group of patients. Lee et al. (2014) use Logistic Regression (LR)and a Naive-biased classifier to predict fasting plasma glucose levels. SVM has an accuracy of 90% for LDA-acquired acute myocardial infarction, including hospital admissions for preventable causes—LDA and SVM both have 92%; RNN has 94.6%. The most significant cause of Diabetes is high blood sugar. Li et al. (2020) demonstrated a deep learning model that can accurately predict glucose levels. Their model gives an effective prediction horizon (PHeff) with a small-time lag for both simulated and real-world patient

datasets. The prediction method from Sneha and Gangil (2019) uses machine learning and the search for the best classifier to find the next best outcome. Predictive analysis is a technique for identifying variables and detecting DM early. The number that was found shows that both the Decision Tree (DT) method and the Random Forest (RF)are 98.20% and 98.00% accurate at making predictions. Diabetes is a problem that can be alleviated, according to the authors in (Srivastava & Dwivedi, 2022), by providing a solution in the form of a prototype using a smart trigger and several machine learning methods. The scikit-learn approach includes a train-test split and K-fold cross-validation, both of which may be used with a positive scheme that is provided. Lee and Kim (2016) used binary LR to compare HW and individual anthropometric parameters between healthy people and those with type 2 Diabetes. In figuring out how much energy you use and how much you move, (Georga et al., 2013) talked about three variables: the glucose profile, the plasma insulin concentration, and the arrival of glucose from meals. Six examples have been looked at using the above variables, and the SVR model is being used to test and validate the work suggested by cross-validation ten times. Table 1 summarizes the most related and recent research works in the domain of utilizing ML in the diagnosis of Diabetes.

**Table 1**
Summarization Table for the Related Work

| No. | Authors | Algorithms implemented | Contributions |
|---|---|---|---|
| 1 | (Zhang et al., 2018) | Classification And Regression Tree (CART) models | pruning and optimization, the regression tree is cut down to size based on a set of metrics, such as the maximum allowable depth of the tree, the minimum allowable number of leaf samples, and the minimum allowable impurity of any given node. |
| 2 | Kumar et al., 2017) | K Nearest Neighbors (KNN) | data preprocessing, the use of K-means, and the application of a classifying algorithm |
| 3 | (Zeng et al., 2016) | Synthesis Minority Oversampling Method (SMOTE) as an efficient data sampling approach. | analyze medical data that is not evenly distributed |
| 4 | (Liu et al., 2017) | fuzzy logic | tasks prediction |
| 5 | (Arsyadani et al., 2023) | Fuzzy-based Information Decomposition (FID) | rebalancing the train data and making more examples for the classes that are not in the majority. |
| 6 | (Yahyaoui et al., 2019) | Random Forest | comparing machine learning and deep learning-based algorithms for Diabetes prediction |
| 7 | (Hasan et al., 2019) | combination of AdaBoost and XgBoost | ensemble approach for predicting Diabetes. |
| 8 | (Lee et al., 2015) | Logistic Regression (LR)and a Naive-biased classifier | predict fasting plasma glucose levels. |
| 9 | (Li et al., 2019) | effective prediction horizon (PHeff) | predict glucose levels. |
| 10 | (Sneha et al., 2019) | Decision Tree (DT) and Random Forest (RF) | uses machine learning and the search for the best classifier to find the next best outcome. |
| 11 | (Lee et al., 2015) | Logistic Regression (LR) | Used binary LR to compare HW and individual anthropometric parameters between healthy people and those with type 2 Diabetes. |

## 3. Methodology, Results, and Analysis

In this section, methodology, results, and analysis are presented. At first, the methodology is described in Section A. Then, in Section B, the datasets that have been considered in this research are presented and described. After that, the feature selection and ranking steps are described in Section C. For Section D, several classification models have been evaluated and compared using several evaluation metrics. Finally, Section E provides a discussion of the obtained results.

### *3.1 Research Methodology*

The methodology that has been followed in this research is depicted in Figure 1. Five main steps have been considered in the methodology. The first step is data collection. For this step, two datasets have been considered. The first dataset consists of 1151 instances extracted from the MESSIDOR image set. MESSIDOR is short for Methods to Evaluate Segmentations and Indexing Techniques in the Field of Retinal Ophthalmology. MESSIDOR is a research project funded by the French Ministry of Research and Defense. More details regarding this dataset could be found at: https://www.adcis.net/en/third-party/messidor/. The second dataset has been collected by the Healthcare Cost and Utilization Project (HCUP), which considers a large group of healthcare databases as well as related products and software-specific tools. HCUP manages to collect the largest databases related to health care in the United States. More information regarding the HCUP and its healthcare databases can be found at https://www.hcup-us.ahrq.gov/overview.jsp.

The pre-processing step of this research considers only the task of handling the missing data. The first dataset is well formatted and does not need any pre-processing. The second dataset consists of 500,000 instances with a large number of instances, with a missing data. Therefore, it has been decided to remove all instances with missing data and keep only instances with no missing data. The new dataset consists of 13067 instances. The second dataset was well formatted, and hence, no other pre-processing step is needed.

The third step in research methodology is the feature selection step. This step aims to identify the most significant features in determining and identifying the object feature (Class) (Junoh et al., 2020; Alazaidah et al., 2017). For this step, four different

feature selection and ranking methods have been considered and evaluated. More information regarding this step can be found in Section C. The fourth step of the research methodology represents the core of this research. That is, to identify the best classifier to handle the Diabetes datasets. More information regarding this step is provided in Section D. The last step in Fig. 1 aims to analyze the results from the previous steps to identify the best classifier.

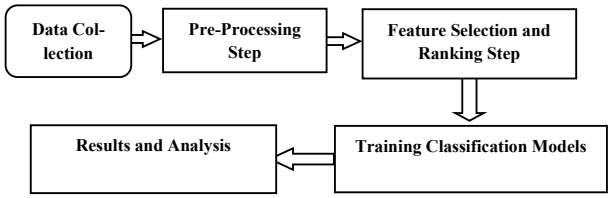


**Fig. 1.** Main phases in research methodology

### *3.2 Dataset Description*

The first dataset is known as Diabetic Retinopathy Debrecen (DRD)and consists of 20 features extracted from eye fundus color numerical images. This dataset contains 1151 instances, all associated with one of two class labels. Around 0.47% of the instances are associated with the "Negative" class, and 0.53% are associated with the "positive" class. Hence, this dataset is nearly balanced. The second dataset considered in this research consists of 13067 instances and 29 features, including the objective feature (class). Each instance could be associated with only one class label from a set consisting of two classes (true and false). Around 59% of instances are associated with the class "True", and 41% of the instances are associated with the class "False". Hence, the dataset is almost balanced. The dataset considered patients with ages ranging from 18 to 90, divided into eight groups as depicted in Table 2. More information regarding the dataset, features, and types could be found at https://www.hcup-us.ahrq.gov/databases.jsp.

**Table 2**
Patient's Ages in the Dataset

| No. | Age | No. of Patients | Percentage |
|---|---|---|---|
| 1 | 18-30 | 571 | 04.3% |
| 2 | 30-40 | 690 | 05.2% |
| 3 | 40-50 | 1634 | 12.5% |
| 4 | 50-60 | 2615 | 20.0% |
| 5 | 60-70 | 2830 | 21.6% |
| 6 | 70-80 | 2621 | 20.0% |
| 7 | 80-90 | 1746 | 13.3% |
| 8 | 90-100 | 360 | 02.7% |
| **Total** | | **13067** | **100%** |

### *3.3 Feature Selection and Ranking*

As mentioned earlier, identifying the best feature selection and ranking method is the main objective of this research. Therefore, four well-known feature selection and ranking methods have been chosen to be compared. These methods are: InfoGainAttributeEval (Hall et al., 2009), GainRatioAttributeEval (Hall et al., 2009), ClassifierAttributeEval, and CorrelationAttributeEval (Hall et al., 2009). Also, three evaluation metrics are used in order to identify the best feature selection and ranking method. These metrics are Accuracy, F-measure, and ROC Area. Moreover, 12 different classifiers that belong to six learning strategies have been used in the evaluation phase of the feature selection and ranking methods. Figs. 2–7 depict the evaluation results of the considered feature selection and ranking methods using only the best 50% of the features in the considered datasets.

It is worth mentioning that all experiments have been conducted using WEKA (Waikato Environment for Knowledge Analysis), which contains many classifiers and feature selection methods (Hall et al., 2009). All classifiers and feature selection methods have been used with their default settings as they have been implemented in WEKA. These classifiers are: BayesNet (Zhang & Wang, 2020), NaiveBayes (Hall et al., 2009), KStar (Cleary & Trigg, 1995), LWL (Frank, Hall, & Al- Pfahringer, 2012), MultiClassClassifier (Hall et al., 2009), FilteredClassifier (Hall et al., 2009), DecisionTable (Friedman et al., 2000), OneR (Holte, 1993), RandomForest (Breiman, 2001), RandomTree (Hall et al., 2009), SimpleLogistic (Landwehr et al., 2005), and VotedPerceptron (Al-Braihy, Dan, Ullah Khan, & Ullah Khan, 2022). Finally, 10-fold cross-validation has been used to validate the performance of the considered classifiers. Several evaluation metrics are used to evaluate the considered feature selection and ranking methods as well as the considered classifiers. These metrics are Accuracy, True Positive (TP) rate, False Positive (FP) rate, Precision, Recall, F1-measure, and Matthew's Correlation Coefficient (MCC). The previously mentioned metrics are computed using the following equations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$TP\ rate\ (\text{Recall}) = \frac{TP}{TP+FN} \tag{2}$$

$$FP\ rate\ = \frac{FP}{FP+TN} \tag{3}$$

$$\text{Precision}\ = \frac{TP}{TP+FP} \tag{4}$$

$$\text{Recall}\ = \frac{TP}{TP+FN} \tag{5}$$

$$\text{F1-Measure}\ = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \tag{6}$$

$$\text{MCC}\ = \frac{(TP\times TN - FP\times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{7}$$

According to Fig. 2 and Fig. 5, the highest Accuracy results have been achieved using the CorrelationAttributeEval. method on both datasets. ClassifierAttributeEval. achieved the highest F1-measure on the DRD dataset, as shown in Fig. 3, while the highest F1-measure result was achieved using the CorrelationAttributeEval. method on the HCUP dataset, as depicted in Fig. 6. For the ROC Area metric, and from Fig. 4, the highest result has been achieved using the CorrelationAttributeEval. method on the DRD dataset and the InfoGainAttributeEval. and GainRatioAttributeEval. methods on the HCUP dataset, as shown in Fig. 7. It is worth mentioning that one classifier (MultiClassClassifier) achieved the best results in the DRD dataset considering the three metrics. For the HCUP dataset, several classifiers managed to achieve identical performance considering the three-evaluation metrics.

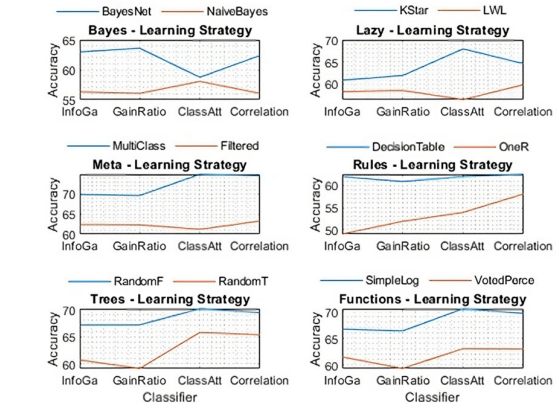


**Fig. 2.** Evaluation results for Accuracy – DRD dataset

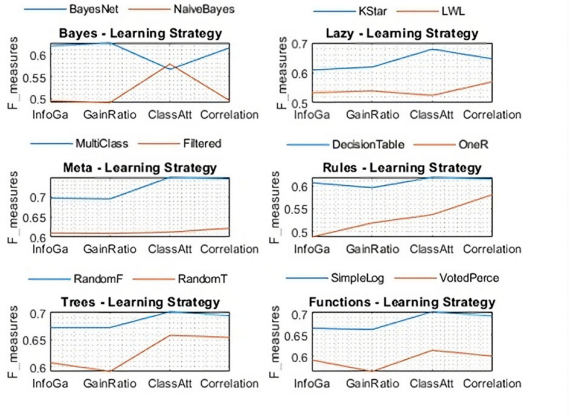


**Fig. 3.** Evaluation results for F_measures – DRD dataset

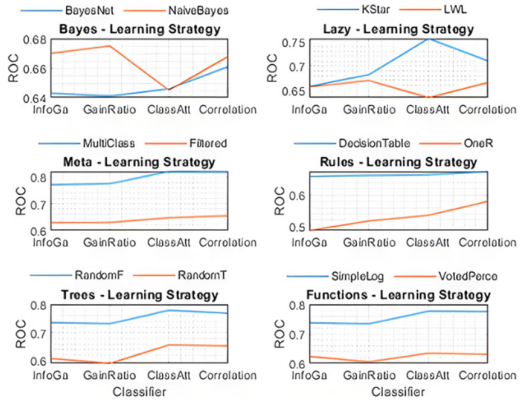


**Fig. 4.** Evaluation results for ROC = DRD dataset

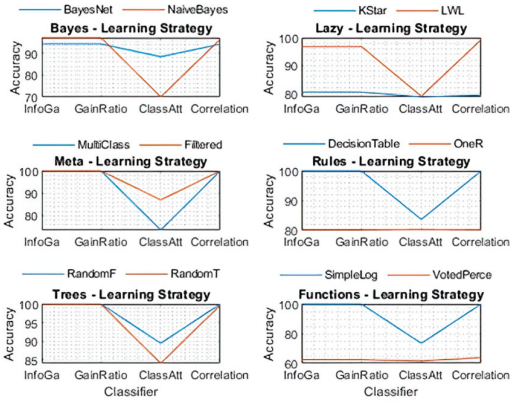


**Fig. 5.** Evaluation results for Accuracy -HCUP dataset

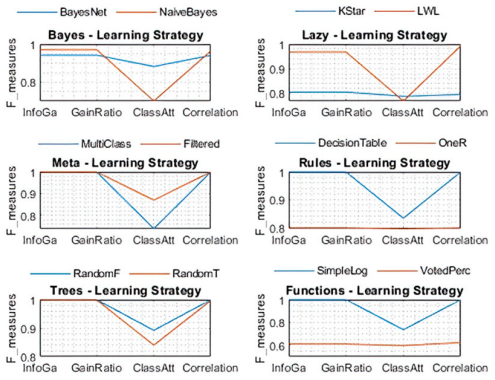


**Fig. 6.** Evaluation results for F_measures - HCUP

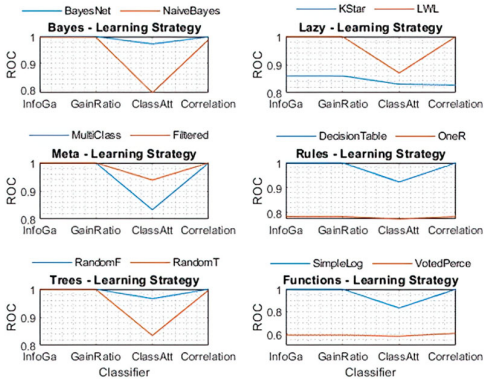


**Fig. 7.** Evaluation results for ROC – FCUP dataset

Table 3 summarizes the Average for the three-evaluation metrics in the two considered datasets. For Table 3 and the following tables, 'A' stands for Accuracy and 'F' stands for F1-measure.

**Table 3**
Summarization of the Results Obtained in the Evaluation phase of the Considered Feature Selection and Ranking Methods

| Method | DRD Dataset | | | HCUP Dataset | | |
|---|---|---|---|---|---|---|
| | A | F | ROC | A | F | ROC |
| **InfoGainAttributeEval.** | 61.468 | 0.600 | 0.657 | 93.619 | 0.935 | 0.945 |
| **GainRatioAttributeEval.** | 61.381 | 0.599 | 0.659 | 93.619 | 0.935 | 0.945 |
| **ClassifierAttributeEval.** | 63.539 | 0.629 | 0.683 | 78.706 | 0.783 | 0.824 |
| **CorrelationAttributeEval.** | 64.039 | 0.628 | 0.689 | 93.735 | 0.936 | 0.943 |

As shown in Table 2, it is obvious that the CorrelationAttributeEval. method is the best choice when attempting to optimize both Accuracy and F1-measure. Also, the InfoGainAttributeEval and GainRatioAttributeEval methods have identical performance considering the three-evaluation metrics on the HCUP dataset. Both InfoGainAttributeEval. and GainRatioAttributeEval. methods are the best choices when attempting to optimize the ROC Area metric on the HCUP dataset, while CorrelationAttributeEval. is the best choice to optimize the ROC Area on the DRD dataset. Table 4 summarizes the number of classifiers that achieved the highest possible value considering the three-evaluation metrics in the two considered datasets.

**Table 4**
The Number of Classifiers that achieved the Highest Possible Value for the Considered evaluation Metrics

| **Method** | **DRD Dataset** | | | **HCUP Dataset** | | |
|---|---|---|---|---|---|---|
| | **A** | **F** | **ROC** | **A** | **F** | **ROC** |
| **InfoGainAttributeEval.** | **1** | **1** | **1** | **4** | **4** | **7** |
| **GainRatioAttributeEval.** | **1** | **1** | **1** | **4** | **5** | **7** |
| **ClassifierAttributeEval.** | **1** | **1** | **1** | **0** | **0** | **0** |
| **CorrelationAttributeEval.** | **1** | **1** | **1** | **4** | **5** | **6** |

Based on Table 3, the best choices to optimize the Accuracy metric are InfoGainAttributeEval, GainRatioAttributeEval, and CorrelationAttributeEval. Also, the best choices to optimize the F1-measure metric are GainRatioAttributeEval. and CorrelationAttributeEval., while the best choices to optimize the ROC Area metric are the InfoGainAttributeEval. method and the GainRatioAttributeEval. method. Therefore, based on the last two tables, the CorrelationAttributeEval. method would be the best choice to handle the task of feature selection and ranking for the considered datasets in this research.

Moreover, based on the previous evaluation results, several classifiers showed identical performance. Hence, more evaluation should be conducted to determine the best classifier that suits the datasets considered in this research. Consequently, Tables 5 to 7 show the predictive performance of the twelve different classifiers that belong to the six learning strategies and using six evaluation metrics.

### *3.4 Training Classification Models Step*

This step aims to identify the best classification model (classifier) that can handle the Diabetes datasets effectively. Hence, twelve different classification models that belong to six different learning strategies have been evaluated using six evaluation

metrics. All these classifiers have been used with their default settings as implemented in WEKA. Table 5 depicts the evaluation results of the twelve classifiers using the TP Rate and the FP Rate metrics.

Based on Table V, the best results for TP Rate and FP Rate metrics have been achieved by five classifiers. These classifiers are: MultiClassClassifier and FilteredClassifier from the Meta learning strategy; DecisionTable from the Rules learning strategy; RandomForest from the Trees learning strategy; and SimpleLogistic from the Function learning strategy. MultiClassClassifier is the dominant classifier on the DRD dataset.

**Table 5**
Evaluation of the Considered Classifiers Using TP Rate and FP Rate

| Learning Strategy | Classifier | DRD Dataset | | HCUPDataset | |
|---|---|---|---|---|---|
| | | TP | FP | TP | FP |
| Bayes | BayesNet | 0.633 | 0.345 | 0.945 | 0.08 |
| | NaiveBayes | 0.568 | 0.387 | 0.963 | 0.054 |
| Lazy | KStar | 0.613 | 0.385 | 0.797 | 0.209 |
| | LWL | 0.597 | 0.370 | 0.839 | 0.11 |
| Meta | MultiClassClassifier | **0.749** | **0.245** | **1** | **0** |
| | FilteredClassifier | 0.632 | 0.356 | **1** | **0** |
| Rules | DecisionTable | 0.624 | 0.359 | **1** | **0** |
| | OneR | 0.533 | 0.476 | 0.801 | 0.233 |
| Trees | RandomForest | 0.692 | 0.304 | **1** | **0** |
| | RandomTree | 0.617 | 0.386 | 0.971 | 0.031 |
| Functions | SimpleLogistic | 0.712 | 0.279 | **1** | **0** |
| | VotedPerceptron | 0.636 | 0.385 | 0.616 | 0.467 |

Table 6 depicts the evaluation results of the twelve classifiers using Precision (P) and Recall (R) metrics.

**Table 6**
Evaluation of the Considered Classifiers Using Precision and Recall Metrics

| Learning Strategy | Classifier | DRD Dataset | | HCUP Dataset | |
|---|---|---|---|---|---|
| | | P | R | P | R |
| Bayes | BayesNet | 0.666 | 0.633 | 0.95 | 0.945 |
| | NaiveBayes | 0.695 | 0.568 | 0.965 | 0.963 |
| Lazy | KStar | 0.615 | 0.613 | 0.799 | 0.797 |
| | LWL | 0.661 | 0.597 | 0.885 | 0.839 |
| Meta | MultiClassClassifier | **0.756** | **0.749** | **1** | **1** |
| | FilteredClassifier | 0.646 | 0.632 | **1** | **1** |
| Rules | DecisionTable | 0.646 | 0.624 | **1** | **1** |
| | OneR | 0.530 | 0.533 | 0.8 | 0.801 |
| Trees | RandomForest | 0.696 | 0.692 | **1** | **1** |
| | RandomTree | 0.617 | 0.617 | 0.971 | 0.971 |
| Functions | SimpleLogistic | 0.723 | 0.712 | **1** | **1** |
| | VotedPerceptron | 0.642 | 0.636 | 0.602 | 0.616 |

From Table 6, the same classifiers that achieved the best values for TP rate and FP rate metrics also achieved the best results for Precision and Recall metrics. Moreover, MultiClassClassifier is the dominant classifier on the DRD dataset, while several classifiers show identical performance on the HCUP dataset.

Table 7 depicts the evaluation results of the twelve classifiers using MCC and PRC metrics.

According to Table 6, the best results achieved for the MCC metric on the HCUP dataset have been achieved by five classifiers: MultiClassClassifier and FilteredClassifier from the meta-learning strategy, DecisionTable from the rules learning strategy, RandomForest from the tree learning strategy, and SimpleLogistic from the function learning strategy. For the PRC Area metric, the same previously mentioned classifiers achieved the best results, in addition to BayesNet from the Bayes learning strategy and LWL from the lazy learning strategy. For the DRD dataset, MultiClassClassifier achieves the best results for MCC and PRC metrics.

**Table 7**

Evaluation of the Considered Classifiers Using MCC and PRC Area Metrics

| Learning Strategy | Classifier | DRD Dataset | | HCUP Dataset | |
|---|---|---|---|---|---|
| | | MCC | PRC | MCC | PRC |
| **Bayes** | BayesNet | 0.305 | 0.669 | 0.890 | 1 |
| | NaiveBayes | 0.259 | 0.664 | 0.924 | 0.99 |
| **Lazy** | KStar | 0.227 | 0.669 | 0.583 | 0.821 |
| | LWL | 0.265 | 0.649 | 0.723 | 1 |
| **Meta** | MultiClassClassifier | **0.505** | **0.828** | **1** | **1** |
| | Filtered | 0.281 | 0.641 | **1** | **1** |
| **Rules** | Decision Tree | 0.275 | 0.669 | **1** | **1** |
| | OneR | 0.057 | 0.517 | 0.582 | 0.736 |
| **Trees** | RandomForest | 0.387 | 0.753 | **1** | **1** |
| | RandomTree | 0.231 | 0.573 | 0.941 | 0.958 |
| **Functions** | SimpleLogistic | 0.436 | 0.780 | **1** | **1** |
| | VotedPerceptron | 0.268 | 0.594 | 0.166 | 0.563 |

## 4. Discussion

According to all previous results, several classifiers managed to achieve the best possible results on the HCUP dataset with respect to the considered metrics. The case was the opposite in the DRD dataset, where only one classifier showed a good and stable performance. The reason for that is the nature of the datasets, where more than 50% of features in the HCUP dataset are discrete nominal features. Hence, several features showed identical performance. For the DRD dataset, the majority of features are continuous. Therefore, the prediction task was more complicated, and hence, only the MultiClassClassifier managed to achieve a good predictive performance on this dataset. The conclusion that could be drawn here is that it will be a better choice to use MultiClassClassifier with a dataset where most of the features are of continuous type, while it would be better to use RandomForest with a dataset where the majority of features are discrete.

Moreover, the predictive performance of the classifiers on the DRD dataset was less than that on the HCUP dataset. The main reason for that could be the low number of instances compared to the number of features in the DRD dataset. Therefore, it is highly recommended to balance and tune the total number of instances according to the total number of features in the dataset. Finally, regarding the most appropriate feature selection method, the best choice is to use correlation attribute evaluation method, especially with datasets that consist of many numeric features like the DRD dataset.

Regarding the limitations of this research, it is highly recommended for future research to consider more feature selection and ranking methods. Also, it will be a good idea to consider other classification models from other learning strategies or to utilize the capabilities of deep learning in a similar research project. Moreover, other datasets with more and different features might be useful in this domain.

## 5. Conclusion and Future Work

In this paper, two datasets related to DM with different characteristics have been considered and pre-processed. Also, four different feature selection and ranking methods have been applied to the considered datasets to determine the best method that suits them. Moreover, twelve different classifiers that belong to six learning strategies have been evaluated on the considered datasets using several evaluation metrics. The results showed that the correlation attribute evaluation method is the best feature selection method to use with the Diabetes datasets. Moreover, MultiClassClassifier achieved the best results compared with the other eleven classifiers. In future work, more evaluation should be considered with respect to other feature selection methods as well as other classifiers. Moreover, an investigation of deep learning models' performance on the same datasets is highly recommended.

## References

Alazaidah, R., Ahmad, F. K., & Mohsen, M. F. M. (2017). A comparative analysis between the three main approaches that are being used to. *International Journal of Soft Computing, 12*(4), 218-223.

Alazaidah, R., Ahmad, F. K., & Mohsin, M. (2020). Multi label ranking based on positive pairwise correlations among labels. *The International Arab Journal of Information Technology, 17*(4), 440-449.

Alazaidah, R., Ahmad, F. K., Mohsen, M. F. M., &Junoh, A. K. (2018). Evaluating conditional and unconditional correlations capturing strategies in multi label classification. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 10*(2-4), 47-51.

Alazaidah, R., Samara, G., Almatarneh, S., Hassan, M., Aljaidi, M., & Mansur, H. (2023). Multi-Label Classification Based on Associations. *Applied Sciences, 13*(8), 5081.

Aliberti, A., Pupillo, I., Terna, S., Macii, E., Di Cataldo, S., Patti, E., & Acquaviva, A. (2019). A multi-patient data-driven approach to blood glucose prediction. *IEEE Access, 7*, 69311-69325.

Alluwaici, M. A., Junoh, A. K., &Alazaidah, R. (2020). New problem transformation method based on the local positive pairwise dependencies among labels. *Journal of Information & Knowledge Management, 19*(01), 2040017.

Arsyadani, F., &Purwinarko, A. (2023). Implementation of Synthetic Minority Oversampling Technique and Two-phase Mutation Grey Wolf Optimization on Early Diagnosis of Diabetes using K-Nearest Neighbors. *Recursive Journal of Informatics, 1*(1), 9-17.

Ashiquzzaman, A., Tushar, A. K., Islam, M. R., Shon, D., Im, K., Park, J. H., ... & Kim, J. (2018). Reduction of overfitting in diabetes prediction using deep learning neural network. *In IT Convergence and Security 2017: Volume 1 (pp. 35-43).* Springer Singapore.

Bose, A. S. C., & Ramesh, V. (2023). Highly accurate grey neural network classifier for an abdominal aortic aneurysm classification based on image processing approach. Int. Arab J. Inf. Technol., 20(2), 215-223.

Breiman, L. (2001). Random forests machine learning. *Journal of Clinical Microbiology, 2*, 199-228.

Cho, N. H., Shaw, J. E., Karuranga, S., Huang, Y., da Rocha Fernandes, J. D., Ohlrogge, A. W., & Malanda, B. I. D. F. (2018). IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes research and clinical practice, 138*, 271-281.

Cleary, J. G., & Trigg, L. E. (1995). K*: An instance-based learner using an entropic distance measure. *In Machine Learning Proceedings 1995* (pp. 108-114). Morgan Kaufmann.

Dritsas, E., & Trigka, M. (2022). Data-driven machine-learning methods for diabetes risk prediction. *Sensors, 22*(14), 5304.

Forde, H., Davenport, C., Rochfort, K. D., Wallace, R. G., Durkan, E., Agha, A., ... & Smith, D. (2022). Serum OPG/TRAIL ratio predicts the presence of cardiovascular disease in people with type 2 diabetes mellitus. *Diabetes Research and Clinical Practice, 189*, 109936.

Frank, E., Hall, M., & Pfahringer, B. (2012). Locally weighted naive bayes. arXiv preprint arXiv:1212.2487.

Friedman, J., Hastie, T., &Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics, 28*(2), 337-407.

Georga, E. I., Protopappas, V. C., Ardigo, D., Marina, M., Zavaroni, I., Polyzos, D., & Fotiadis, D. I. (2013). Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression. *IEEE journal of biomedical and health informatics, 17*(1), 71-81.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter, 11*(1), 10-18.

Hasan, M. K., Alam, M. A., Das, D., Hossain, E., & Hasan, M. (2020). Diabetes prediction using ensembling of different machine learning classifiers. *IEEE Access, 8*, 76516-76531.

Holte, R. C. (1993). Very simple classification rules perform well on most commonly used datasets. Machine learning, 11, 63-90.

Huang, J., Yeung, A. M., Armstrong, D. G., Battarbee, A. N., Cuadros, J., Espinoza, J. C., ... &Klonoff, D. C. (2023). Artificial intelligence for predicting and diagnosing complications of diabetes. *Journal of Diabetes Science and Technology, 17*(1), 224-238.

Junoh, A. K., Ahmad, F. K., Mohsen, M. F. M., & Alazaidah, R. (2018, April). Open research directions for multi label learning. *In 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE) (pp. 125-128). IEEE*.

Junoh, A. K., AlZoubi, W. A., Alazaidah, R., & Al-luwaici, W. (2020). New features selection method for multi-label classification based on the positive dependencies among labels. Solid State Technology, 63(2s).

Khaleel, F. A., & Al-Bakry, A. M. (2022). Diagnosis of diabetes using machine learning algorithms. *Materials Today: Proceedings, 80*, 3200-3203.

Krishnamoorthi, R., Joshi, S., Almarzouki, H. Z., Shukla, P. K., Rizwan, A., Kalpana, C., & Tiwari, B. (2022). A novel diabetes healthcare disease prediction framework using machine learning techniques. *Journal of Healthcare Engineering, 2022.*

Kumar, P. S., & Umatejaswi, V. (2017). Diagnosing diabetes using data mining techniques. International Journal of Scientific and Research Publications, 7(6), 705-709.

Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine learning, 59*, 161-205.

Lee, B. J., & Kim, J. Y. (2015). Identification of type 2 diabetes risk factors using phenotypes consisting of anthropometry and triglycerides based on machine learning. *IEEE journal of biomedical and health informatics, 20*(1), 39-46.

Lee, B. J., Ku, B., Nam, J., Pham, D. D., & Kim, J. Y. (2014). Prediction of fasting plasma glucose status using anthropometric measures for diagnosing type 2 diabetes. *IEEE journal of biomedical and health informatics, 18*(2), 555-561.

Li, K., Daniels, J., Liu, C., Herrero, P., & Georgiou, P. (2019). Convolutional recurrent neural networks for glucose prediction. *IEEE journal of biomedical and health informatics, 24*(2), 603-613.

Liu, S., Zhang, J., Xiang, Y., & Zhou, W. (2017). Fuzzy-based information decomposition for incomplete and imbalanced data learning. *IEEE Transactions on Fuzzy Systems, 25*(6), 1476-1490.

Mao, Y., Zhu, Z., Pan, S., Lin, W., Liang, J., Huang, H., ... & Chen, G. (2023). Value of machine learning algorithms for predicting diabetes risk: A subset analysis from a real-world retrospective cohort study. *Journal of Diabetes Investigation, 14*(2), 309-320.

Sneha, N., &Gangil, T. (2019). Analysis of diabetes mellitus for early prediction using optimal features selection. *Journal of Big data, 6*(1), 1-19.

Srivastava, R., & Dwivedi, R. K. (2022). A survey on diabetes mellitus prediction using machine learning algorithms. *In ICT Systems and Sustainability: Proceedings of ICT4SD 2021, Volume 1 (pp. 473-480).* Springer Singapore.

Yahyaoui, A., Jamil, A., Rasheed, J., &Yesiltepe, M. (2019, November). A decision support system for diabetes prediction using machine learning and deep learning techniques. *In 2019 1st International informatics and software engineering conference (UBMYK) (pp. 1-4). IEEE*.

Zeng, M., Zou, B., Wei, F., Liu, X., & Wang, L. (2016, May). Effective prediction of three common diseases by combining SMOTE with Tomek links technique for imbalanced medical data. *In 2016 IEEE International Conference of Online Analysis and Computing Science (ICOACS) (pp. 225-228). IEEE*.

Zhang, B., Wei, Z., Ren, J., Cheng, Y., & Zheng, Z. (2018). An empirical study on predicting blood pressure using classification and regression trees. IEEE access, 6, 21758-21768.

Zhang, C., & Wang, P. (2000, September). A new method of color image segmentation based on intensity and hue clustering. *In Proceedings 15th International Conference on Pattern Recognition. ICPR-2000* (Vol. 3, pp. 613-616). IEEE.