Computational and Structural Biotechnology Journal 23 (2024) 212–233

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

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

Research Article

Secure and privacy-preserving automated machine learning operations into end-to-end integrated IoT-edge-artificial intelligence-blockchain monitoring system for diabetes mellitus prediction

Alain Hennebelle [a], Leila Ismail [a,b,c,*], Huned Materwala [b,c], Juma Al Kaabi [d,e], Priya Ranjan [f], Rajiv Janardhanan [g]

[a] School of Computing and Information Systems, The University of Melbourne, Australia
[b] Intelligent Distributed Computing and Systems Lab, Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University, United Arab Emirates
[c] National Water and Energy Center, United Arab Emirates University, United Arab Emirates
[d] College of Medicine and Health Sciences, Department of Internal Medicine, United Arab Emirates University, United Arab Emirates
[e] Tawam and Mediclinic Hospitals, Al Ain, Abu Dhabi, United Arab Emirates
[f] School of Computer Science, Internet of Things Center of Excellence, University of Petroleum and Energy Studies, India
[g] Faculty of Medical & Health Sciences, SRM Institute of Science & Technology, India

ARTICLE INFO

Keywords:
Artificial intelligence (AI)
Blockchain
Diabetes mellitus type 2
Diagnosis
Digital health
eHealth
Logistic regression (LR)
Machine learning
Prognosis
Random forest (RF)
Risk factors
Smart connected healthcare
Support vector machine (SVM)

ABSTRACT

Diabetes Mellitus, one of the leading causes of death worldwide, has no cure to date and can lead to severe health complications, such as retinopathy, limb amputation, cardiovascular diseases, and neuronal disease, if left untreated. Consequently, it becomes crucial to be able to monitor and predict the incidence of diabetes. Machine learning approaches have been proposed and evaluated in the literature for diabetes prediction. This paper proposes an IoT-edge-Artificial Intelligence (AI)-blockchain system for diabetes prediction based on risk factors. The proposed system is underpinned by blockchain to obtain a cohesive view of the risk factors data from patients across different hospitals and ensure security and privacy of the user's data. We provide a comparative analysis of different medical sensors, devices, and methods to measure and collect the risk factors values in the system. Numerical experiments and comparative analysis were carried out within our proposed system, using the most accurate random forest (RF) model, and the two most used state-of-the-art machine learning approaches, Logistic Regression (LR) and Support Vector Machine (SVM), using three real-life diabetes datasets. The results show that the proposed system predicts diabetes using RF with 4.57% more accuracy on average in comparison with the other models LR and SVM, with 2.87 times more execution time. Data balancing without feature selection does not show significant improvement. When using feature selection, the performance is improved by 1.14% for PIMA Indian and 0.02% for Sylhet datasets, while it is reduced by 0.89% for MIMIC III.

1. Introduction

Diabetes Mellitus, commonly referred to as diabetes, is one of the top 10 leading causes of death globally [1]. It is a metabolic disease in which the body does not produce enough insulin or body cells do not appropriately respond to insulin, leading to increased blood sugar levels [2]. There are three main types of diabetes, namely type 1, type 2, and gestational diabetes [3]. According to a report by the International Diabetes Federation, 537 million adults (i.e., 1 in every 10 people), between the ages of 20-79 years, worldwide were having diabetes in 2021 [4]. Furthermore, this number is predicted to reach 643 million by 2030 and 783 million by 2045. In 2021, diabetes was responsible for 6.7 million deaths and caused at least USD 966 billion in health expenditure [4].

The etiopathology of type 2 diabetes mellitus has been linked to dynamic interactions between lifestyle, medical conditions, hereditary,

* Corresponding author at: Intelligent Distributed Computing and Systems Lab, Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University, United Arab Emirates.
E-mail address: leila@uaeu.ac.ae (L. Ismail).

https://doi.org/10.1016/j.csbj.2023.11.038
Received 11 July 2023; Received in revised form 20 November 2023; Accepted 20 November 2023
Available online 23 November 2023

psychosocial, and demographic risk factors [3]. Diabetes if left untreated might cause significant issues such as retinopathy, limb amputation, cardiovascular diseases, and neuronal disease [5,6]. In 2021, over 240 million adults with diabetes were undiagnosed (i.e., almost 1 in 2 diabetic) [7]. Consequently, machine learning-based diabetes prediction has gained increased attention in the literature [8–15] for better prognosis/diagnosis support to the medical health professionals and public health organizations [16]. Disparate work in literature focuses on evaluating machine learning algorithms for different diabetes datasets under non-unified experimental setups. However, to the best of our knowledge, no work proposes an end-to-end IoT-edge-Artificial Intelligence (AI)-blockchain integrated computing system for diabetes monitoring and prediction. This paper aims to address this void. The proposed system examines diabetes risk factors through medical sensors/devices and anticipates the likelihood of diabetes occurrence in an individual using the most accurate machine learning model. Furthermore, the proposed system employs edge computing to transform the risk factors data collected from IoT devices and send preprocessed data to the blockchain. Blockchain [17] stores the medical records of the patients as well as the machine learning model parameters and prediction results in a distributed and replicated ledger. This is based on the potential of blockchain in the healthcare industry [18–20]. The consensus, replication, traceability, and distributed features of blockchain aid in security, privacy, audit trail, transparency, and trust in the proposed system. In addition, the efficacy of the most commonly employed machine learning algorithms for predicting diabetes is assessed using PIMA Indian, Sylhet, and MIMIC III datasets.

The main contributions of this paper are as follows.

- We propose an end-to-end automated IoT-edge-AI-blockchain system for diabetes prediction based on risk factors.
- We present a comparative analysis of medical sensors, devices, and methods used to measure the values of diabetes risk factors; hypertension, obesity, cholesterol level, depression, serum uric acid, sleep duration, physical activity, and glucose level.
- We propose an implementation workflow for the proposed system.
- The performance of the proposed system is evaluated and compared with the most used machine learning approaches employing publicly available three real-life diabetes datasets, namely PIMA Indian, Sylhet, and MIMIC III.
- The machine learning prediction models are evaluated in terms of accuracy, precision, recall, F-measure, Area Under the Curve (AUC), Receiver Operating Characteristics (ROC), and execution time.

The rest of the paper is organized as follows. Section 2 provides an overview of the existing literature on machine learning-based diabetes prediction. The proposed automated end-to-end IoT-edge-AI-blockchain system for diabetes mellitus prediction is explained in Section 3. Section 4 discusses the implementation of the proposed system. Numerical experiments and comparative performance results are provided in Section 5. Finally, Section 6 concludes the paper with future research directions.

2. Related work

Several works in the literature have used machine and deep learning algorithms for diabetes prediction [8–15]. These works have described their experimental environment and experiments clearly for reproducibility and have compared the performance of different machine learning algorithms for diabetes prediction. Table 1 summarizes these works and presents the datasets, balancing techniques, feature selection approaches, and machine/deep learning algorithms used in each work. As stated in the table, [8–10] use private datasets for evaluating diabetes prediction models, whereas, [11–15] use publicly available datasets. In particular, [11] uses cross-section survey, NHANES, and PIMA Indian datasets, [12,13] use the PIMA Indian dataset, [14] employs a dataset from the Henan rural cohort study, and [15] uses a dataset from the CBHS health funds company. Furthermore, these studies exclusively concentrate on isolated diabetes prediction and do not propose an end-to-end diabetes prediction system. In contrast, we propose a secure and privacy-preserving end-to-end integrated IoT-edge-AI-blockchain monitoring system for diabetes prediction. We evaluate the performance of our proposed system and compare it with the most employed machine learning models using publicly available three real-life diabetes datasets, namely PIMA Indian, Sylhet, and MIMIC III.

3. Proposed automated end-to-end integrated IoT-edge-artificial intelligence-blockchain monitoring system for diabetes mellitus prediction

The overall architecture of our proposed end-to-end system for diabetes prediction is presented in Fig. 1. The main components of the architecture are explained in the following subsections.

3.1. User's diabetes risk factors monitoring

Diabetes, i.e., increased glucose levels, is associated with different demographic, psychosocial, hereditary, medical conditions, and lifestyle-related risk factors [2]. The values of these risk factors can be either self-reported by the users (i.e., patients/external participants) or measured using biosensors, wearable devices, or medical tests. The self-reported risk factors are age, gender, ethnicity, family history of diabetes, smoking, and alcohol consumption. Age, gender, and ethnicity are reported on the first visit to the hospital. A family history of diabetes is reported on every visit to the hospital. Smoking and alcohol consumption are reported daily. The measurable risk factors are hypertension, obesity, cholesterol level, depression, serum uric acid, sleep duration, physical activity, and glucose levels. Hypertension, obesity, serum uric acid, sleep duration, physical activity, and glucose levels are acquired daily. Cholesterol levels and depression are measured on every visit to the hospital. The measured risk factors data are sent to the mobile phone. A user communicates with the mobile application to identify the risk of incident diabetes.

Different sensors, devices, and methods can be used to acquire the values of measurable risk factors. In the following, we compare them based on their performances and costs.

- *Hypertension Monitoring:* Hypertension is a medical condition where the blood pressure in the arteries remains elevated, i.e., a systolic blood pressure greater than or equal to 140 mmHg and a diastolic blood pressure greater than or equal to 90 mmHg [21]. It increases the risk of developing diabetes. Table 2 lists different hypertension monitors along with their performances and approximate costs in US dollars. As shown in the table, Omron Evolv (HEM-7600T-E) [22] has the best performance whereas Omron M3 Comfort (HEM-7134-E) [23] has the least cost.
- *Obesity Monitoring:* Obesity is characterized by an excessive amount of body fat and is often defined in terms of Body Mass Index (BMI), waist circumference, and/or waist-hip ratio [31]. It is strongly associated with the prevalence of type 2 diabetes. Table 3 shows different methods and devices used to measure obesity with their strengths and weaknesses.
- *Cholesterol Level Monitoring:* Abnormal level of cholesterol and triglycerides increases the risk of type 2 diabetes prevalence. In particular, low level of high-density lipoproteins (HDL) and elevated level of low-density lipoproteins (LDL) leads to the development of diabetes [38]. The standard method to measure the cholesterol level is the lipid panel test (also known as lipid profile test) [39]. This test determines the levels of triglycerides, total, LDL, and HDL cholesterols in an individual. Recently, portable devices, such as EasyTouch [40] and BeneCheck Plus [41], have been developed

Table 1
Summary of related work on diabetes prediction.

Work	Dataset	Features[§]	Observations[§]	Data balancing	Feature selection	Algorithms	Evaluation metrics
[8]	Private: EHRs acquired from 5 hospitals in Saudi Arabia between 2016 – 2018	DOB, gender, height weight, hypertension, fasting plasma glucose, haemoglobin A1C, HDL, LDL, physical activity, diagnosis start date, and primary and secondary diagnosis codes and full names	3000 patients	Data is already balanced	Permutation importance and hierarchical clustering	LR, SVM, DT, RF♦, EMV*	Accuracy, precision, recall, and F-measure
[9]	Private: EHRs data collected at preventive healthcare examinations of healthy population in 10 Slovenian primary healthcare institutions	Related to FINDRISC questionnaire and medical history	27050 patients	✗	✗	Linear regression♦, Glmnet, RF, XGBoost, and lightGBM	AUC and RMSE
[10]	Private: EHRs collected between 2013 – 2018 from a private medical institute, Hanaro Medical Foundation, in Seoul (South Korea)	Related to blood test, anthropometric measurements, diagnostics results, and questionnaire answers	253359 subjects (68.1% normal, 4.3% diabetics, and 27.6% prediabetes)	Majority under-sampling and SMOTE	ANOVA, chi-squared test and recursive feature elimination	LR, RF♦, SVM, XGBoost, stacking†, soft voting†, and confusion matrix-based ensemble†	Accuracy, precision, recall, F-measure, MCC, and KC
[11]	D1: Cross-sectional diabetes survey in Saudi Arabia D2: NHANES D3: PIMA Indian	D1: region, age, gender, BMI, waist size, physical activity, diet, blood pressure, and family history of diabetes D2: smoking, diet, blood pressure, BMI, gender, and region D3: ‡	D1: 4896 (990 diabetics and 3906 non-diabetics) D2: 4918 (1709 prediabetes and 3209 diabetics) D3: 768 (268 diabetics and 500 non-diabetics)	SMOTE	Pearson chi-square test	BPM, AP, DF♦, LD-SVM, DJ, boosted DT, and NN	Accuracy, precision, recall, F-measure, and AUC
[12]	PIMA Indian	‡	768 (268 diabetics and 500 non-diabetics)	✗	Different combinations based on manual inspection	LR♦ and DT	Accuracy, error rate, AIC, BIC, R2, and log likelihood
[13]	PIMA Indian	‡	768 (268 diabetics and 500 non-diabetics)	✗	PCA, k-means clustering, and importance ranking	NB, RF♦, and DT	Accuracy, precision, sensitivity, specificity, F-measure, and AUC
[14]	Henan rural cohort study: participants aged between 18 – 79 years were recruited from five rural areas in Henan province of China between July 2015 and September 2017	Related to socio-demographic characteristics, information on physical examination, and laboratory tests	39259 participants	SMOTE	Iterative approach	LR, CART, ANN, SVM, RF♦, and GBM	AUC, sensitivity, specificity, positive prediction value, negative prediction value, and area under precision-recall curve
[15]	CBHS health funds company in Australia: hospital admissions data between 1995 – 2018	Age, gender, and smoking status	2056 (1028 diabetics and 1028 non-diabetics)	Data is already balanced	✗	LR, kNN, SVM, NB, DT, RF♦, XGBoost, and ANN	Accuracy, precision, recall, F-measure, and AUC

EHRs – Electronic Health Records; ♦ - outperforming model; LR – Logistic Regression; SVM – Support Vector Machine; DT – Decision Tree; RF – Random Forest; EMV – Ensemble Majority Voting; * - EMV consists of LR, SVM, and DT; Glmnet – Regularized Generalized Linear Model; XGBoost – Extreme Gradient Boosting; lightGBM – light Gradient Boosting Machine; † - ensemble algorithms use LR, RF, SVM, and XGBoost; BPM – Bayes Point Machine; AP – Average Perceptron; DF – Decision Forest; LD-SVM – Locally Deep SVM; DJ – Decision Jungle; NN – Neural Network; NB – Naïve Bayes; CART – Classification and Regression Tree; GBM – Gradient Boosting Machine; kNN – k Nearest Neighbor; AUC – Area Under the ROC Curve; RMSE – Root Mean Squared Error; MCC – Mathews Correlation Coefficient; KC – Kappa's Coefficient; AIC – Akaike's Information Criteria; BIC – Bayesian Information Criteria; PCA – Principal Component Analysis; SMOTE – Synthetic Minority Oversampling Technique; NHANES – National Health and Nutrition Examination Survey; HDL – High Density Lipoprotein; LDL – Low Density Lipoprotein; ‡ - Number of times pregnant, plasma glucose concentration at 2 h oral glucose tolerance test, diastolic pressure, triceps skin fold thickness, 2-h serum insulin, BMI, diabetes pedigree function, and age; BMI – Body Mass Index; DOB – Date of Birth; D – Dataset; § - Before data preprocessing; ✗- Not performed.

to measure cholesterol levels. The approximate costs of EasyTouch and BeneCheck Plus are 60USD and 136USD respectively.

- *Depression Monitoring:* Depression is a medical condition that negatively affects the feelings, thoughts, and actions of an individual. It has a strong association with the prevalence of type 2 diabetes [42]. Depression is generally measured using clinical rating scales such as Beck's Depression Inventory (BDI), Center for Epidemiological Studies – Depression scale (CES-D), and Zung Self-Rating Depression Scale (SDS) [43]. A BDI score ≥ 11, CES-D score ≥ 8, or SDS score > 39 increases the risk of developing diabetes.
- *Serum Uric Acid Monitoring:* Serum uric acid is a waste product generated by the body during the purines breakdown process. A serum

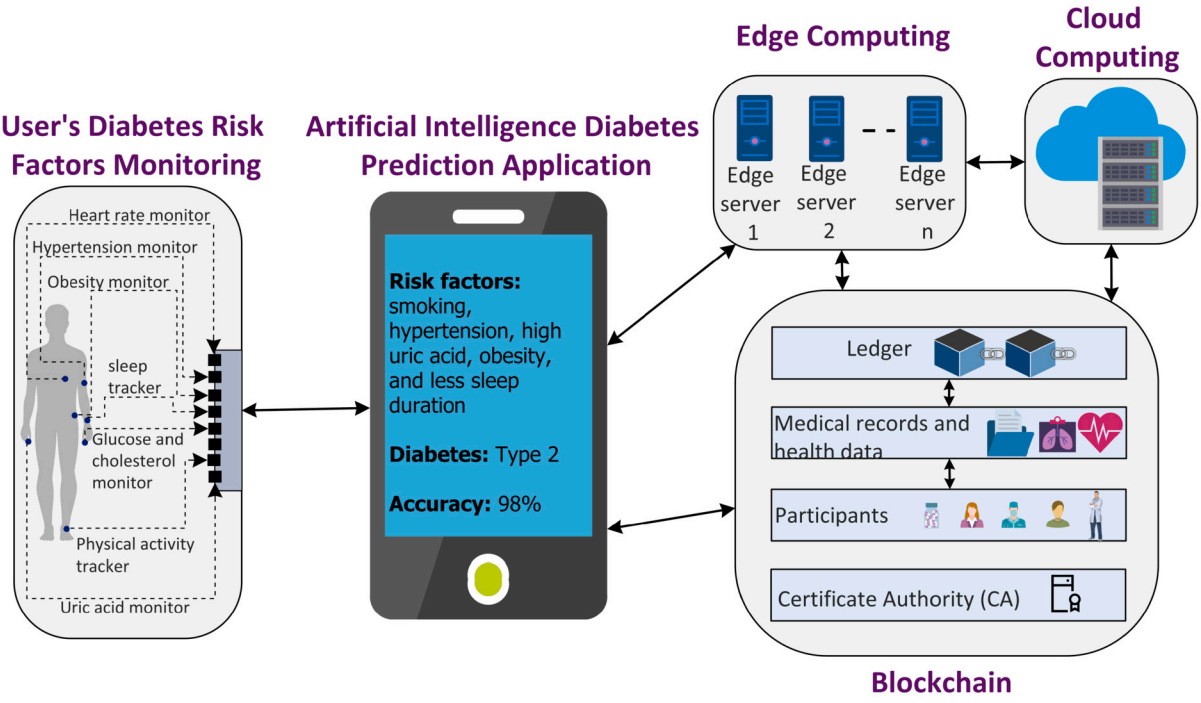


Fig. 1. Architecture overview of proposed automated end-to-end integrated IoT-edge-artificial intelligence-blockchain monitoring system for diabetes mellitus prediction.

Table 2
Comparison between different hypertension monitoring devices.

Device	Performance	Approximate cost (in US Dollars)
Omron Evolv (HEM-7600T-E) [22]	Mean difference compared to standard mercury sphygmomanometer test [24]: -0.1 ± 5.0 mmHg (for systolic blood pressure) -0.2 ± 4.1 mmHg (for diastolic blood pressure)	136 [25]
Omron M3 Comfort (HEM-7134-E) [23]	Mean difference compared to standard mercury sphygmomanometer test [24]: -0.9 ± 5.4 mmHg (for systolic blood pressure) -0.6 ± 4.7 mmHg (for diastolic blood pressure)	63.16 [26]
Omron (HEM-9210T) [27]	Mean difference compared to standard mercury sphygmomanometer test [24]: -2.1 ± 4.7 mmHg (for systolic blood pressure) -1.2 ± 4.1 mmHg (for diastolic blood pressure)	Not available
Mobil-O-Graph [28]	Mean difference compared to standard mercury sphygmomanometer test [29]: -2.2 ± 7.3 mmHg (for systolic blood pressure) -0.4 ± 6.1 mmHg (for diastolic blood pressure)	1365.86 [30]

Table 3
Comparison between different obesity monitoring methods and devices.

Method/device	Strengths	Weaknesses
Statistical BMI calculation [32]	Quick, cost-effective, and easy	Not accurate for elderly, muscular, and pregnant individuals
Skinfold calipers [33]	Easy to use, portable, and cost-effective	Accuracy depends on the skill of the person using the caliper
Smart weighing scales [34]	Quick and easy	Reliability of the result depends on the condition of the individual whose measurement is taken (for instance, hydrated or dehydrated), some accurate scales are costly
Hydrodensitometry [35]	Accurate and reliable	Costly and not suitable for children and elderly people as it requires the individual to be submerged in water for 5-7 seconds repeatedly 2-3 times
Air displacement plethysmography [36]	Quick, accurate, reliable, and suitable for any age	Costly
Dual energy x-ray absorptiometry [37]	Quick, precise, and reliable	Costly

Table 4
Comparison between different serum uric acid monitoring methods and devices.

Method/device	Performance (Coefficient of variation)	Approximate cost (in US Dollars)
Smartphone as electro-chemical analyzer [47]	Low concentration: 4.1%* Mid concentration: 2.47%* High concentration: 1.87%* [47]	Not available
EasyTouch [40]	27.2% [48] (Not acceptable)†	60 [49]
UAsure [50]	25.9% [48] (Not acceptable)†	64 [51]
BeneCheck Plus [41]	9.5% [48] (Acceptable)†	136 [52]
HumaSens^{plus} [46]	11.5% [48] (Acceptable)†	52 [53]
Liquid chromatography mass spectrometry [54]	0.01 – 3.37%* [54]	Not available

*Average; †According to College of American Pathologists

Table 5
Comparison between different sleep duration monitoring tests and devices.

Type	Test/device	Performance	Approximate cost (in US Dollars)
Non-invasive	Polysomnography test [57]	Sensitivity: 0.957* Specificity: 0.532* Accuracy: 0.904* Cohen's kappa: 0.495* [57]	943 – 2,798 [58]
Wearable	ŌURA ring [59]	Sensitivity (to detect sleep): 96%* Specificity (to detect wake): 48%* [60]	299-399 [59]
Wearable	Fitbit Flex [61]	97.46% accuracy [62]	100 [62]
Wearable	Fitbit Charge HR [55]	Overestimates the sleep duration [63]	65.39 [64]
Wearable	Polar A370 fitness tracker [65]	**Age group (mean ± SD): 11 ± 0.8** Sensitivity*: 0.93 Specificity*: 0.77 Accuracy*: 0.91 **Age group (mean ± SD): 17.8 ± 1.8** Sensitivity*: 0.91 Specificity*: 0.83 Accuracy*: 0.90 [66]	163 [65]
Wearable	Actiwatch 2 [67]	**Age group (mean ± SD): 11 ± 0.8** Sensitivity*: 0.93 Specificity*: 0.68 Accuracy*: 0.90 **Age group (mean ± SD): 17.8 ± 1.8** Sensitivity*: 0.93 Specificity*: 0.58 Accuracy*: 0.89 [66]	Not available
Wearable	Fitbit Alta HR [68]	All sleep Sensitivity: 0.96 ± 0.02 Specificity: 0.58 ± 0.16 Accuracy: 0.90 ± 0.04 [69]	270 [70]
Wearable	Withings Pulse [71]	98.1% accuracy [62]	100 [71]
Wearable	Misfit Shine [62]]	96% accuracy [62]	100 [62]
Wearable	Jawbone Up24 [72]	97.23% accuracy [62]	100 [62]
Non-wearable	EMFIT Quantified Sleep [73]	Overestimates total sleep time and underestimates wake after sleep [74]	Not available

*Average

uric acid level >370 μmol/l is associated with a risk of developing type 2 diabetes [44]. A uric acid test is commonly used to measure the amount of uric acid either using blood or urine samples [45]. Recently, several devices have been introduced to measure serum uric acid levels. Table 4 shows a comparison between different methods and devices. As shown in the table, HumaSens^{plus} [46] is the most economical compared to other methods/devices for serum uric acid monitoring.

- *Sleep Duration Monitoring:* The quantity of sleep during night time is highly associated with the prevalence of type 2 diabetes [2]. Compared to 6-8 hours of nighttime sleep, a shorter sleep duration (<6 hours/night) and a longer sleep duration (>8 hours/night) are associated with diabetes. In addition, daytime napping can lead to the prevalence of diabetes. Table 5 summarizes different tests, devices, and applications used for tracking sleep. It shows the performance of each test/device/application along with its cost in USD. As shown in the table, Fitbit Charge HR [55] costs the least compared to other tests/devices for monitoring sleep duration. In addition to these tests and devices, Sleep Cycle [56] is a free mobile application that is used to monitor sleep duration.
- *Physical Activity Monitoring:* Physical inactivity can lead to obesity and depression, resulting in the prevalence of type 2 diabetes

Table 6
Comparison between different physical activity monitoring devices.

Type	Device	Performance (Accuracy)	Approximate cost (in US Dollars)
Waist-based	Fitbit One [76]	>90% [77]	70 [78]
Waist-based	Omron HJ-321 [79]	>90% [77]	67.25 [80]
Waist-based	Sportline 340 Strider [75]	>90% [77]	22 [81]
Wrist-based	Fitbit Force [82]	<90% [77]	Not available
Ankle-based	StepWatch activity monitor [83]	Non-running activities: >95% Running activities: 74.4% [77]	Not available
Mobile phone	Apple iPhone 5 [84]	<90% [77]	Obsolete
Mobile phone	Samsung Galaxy S4 [85]	<90% [77]	405 [86]

Table 7
Comparison between different glucose level monitoring devices.

Type	Device	Performance	Approximate cost (in US Dollars)
Non-invasive	Wearable-band type visible-near infrared optical [88]	Average correlation coefficient between actual and measured glucose: 0.86 [88]	Not available
Non-invasive	Triple-pole complementary split ring resonator-based microwave bio-sensor [89]	Sensitivity: 6.2 dB/(mg/ml) [89]	Not available
Invasive	EasyTouch [40]	Not reported	60 [49]
Invasive	BeneCheck Plus [41]	Not reported	136 [52]

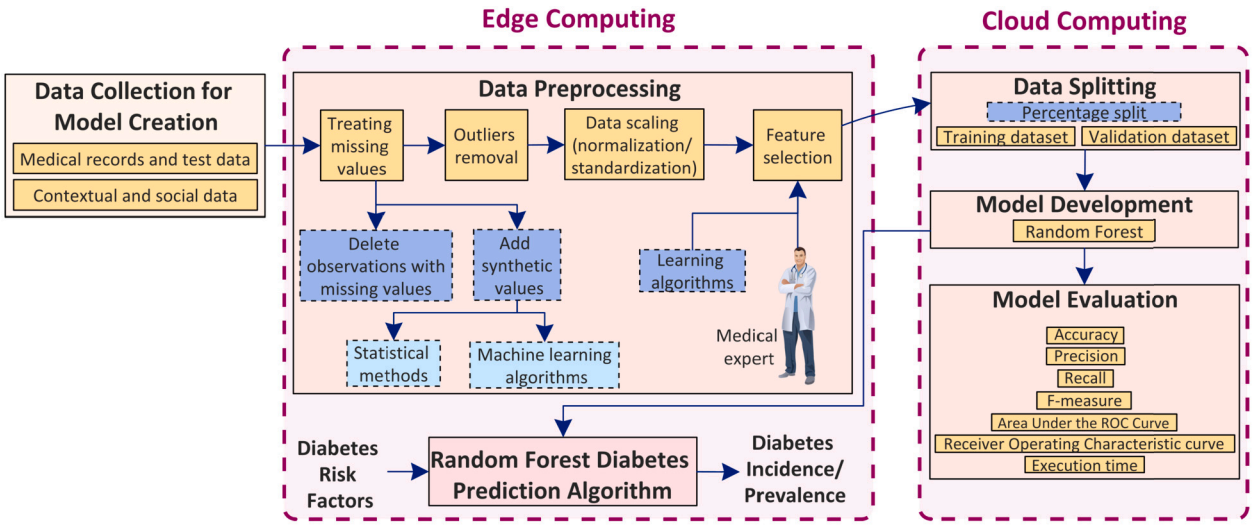


Fig. 2. Stages of artificial intelligence-based diabetes mellitus prediction system.

[2]. An individual performing 30-60 minutes of exercise 3 – 4 times/week can be considered physically active. Table 6 summarizes different devices for tracking physical activity. As shown in the table, Sportline 340 Strider [75] outperforms other physical activity monitors in terms of performance and cost.

- *Glucose Level Monitoring:* Diabetes is characterized by elevated glucose levels. For instance, an individual having a fasting plasma glucose level of less than 100 mg/dl is non-diabetic, whereas one having a level between 100-125 mg/dl is considered pre-diabetic and an individual having a fasting plasma glucose level greater than 125 mg/dl is diabetic [87]. Table 7 compares different invasive and non-invasive glucose monitoring devices.

3.2. Artificial intelligence diabetes prediction application

As shown in Fig. 1, the mobile application plays the role of a gateway between the sensors connected to the users and the edge computing devices for uploading risk factors data. A user can also communicate with the mobile application to identify the risk of diabetes based on risk factors data. In that case, the mobile application will communicate with the edge server to retrieve the prediction results.

3.3. Edge and cloud computing

The risk factors data, collected using different medical sensors and devices, is sent to edge servers to transform it into a format that can be used by a machine learning algorithm. Edge servers in proximity to the users, compared to the cloud, aid in real-time data acquisition. The gathered data is subsequently preprocessed by the edge servers. However, edge servers are not capable of training compute-intensive machine learning models due to low processing and storage capacity. Consequently, the preprocessed data is transmitted to the cloud for storage and the development and validation of machine learning model(s). The developed model is then sent back to the edge for predicting the risk

Table 8
Security and privacy analysis using blockchain.

Issue	Blockchain solution
Data confidentiality	The private and sensitive health data records can only be accessed by authorized network participants based on access control rights defined in the blockchain. A transaction for unauthorized access will not be validated by the network participants.
Data integrity	Health data records are stored in blocks and each block is linked to the previous one using a cryptography mechanism. Modifying existing data in a block is computationally very expensive as the attacker has to change all the subsequent blocks in each copy of the ledger. Furthermore, any modification if performed will be logged in the ledger and can be easily traced.
Data repudiation	Data update and query events are recorded in an immutable ledger after validation ensuring fraud denials.
Data audit	The replicated, time-stamped and immutable ledger ensures efficient, trusted, and integral auditing.
Data access control	Access control rights for health data records in the blockchain can be defined using smart contracts for secure access by authorized participants.

of developing diabetes based on risk factors data. Fig. 2 shows the machine learning operations pipeline used in our framework for diabetes prediction. The following explains the different operations involved in the machine learning pipeline for AI-based diabetes mellitus prediction.

- *Data collection for model creation:* In this stage, medical records, laboratory results, and contextual and social data are collected. The inclusion of the risk factors in the dataset should be verified. The collected data is then required to be aggregated. The diabetes class labels should be defined. For instance, all the observations in the dataset having fasting plasma glucose levels less than 100 mg/dl can be labeled as a non-diabetic class, whereas all those having levels between 100-125 mg/dl can be labeled as a pre-diabetic class and all those having fasting plasma glucose level greater than or equal to 125 mg/dl can be labeled as a diabetic class. This can be done with the help of an expert's advice.
- *Data preprocessing:* This stage involves handling missing values, removal of outliers, data scaling, and feature selection. The missing values can be treated by either removing the corresponding observations or adding synthetic values. Synthetic values can be generated using statistical (mean/mode/median) or machine learning (kNN imputation and rpart) approaches [90]. Data scaling is achieved through normalization and/or standardization. The numerical features having varying ranges should be normalized. This is because the model could be biased towards the feature with a bigger range [91]. For example, the range for BMI is 18.2-67.1, whereas that for plasma glucose is 44-199. In feature selection, the features that do not contribute to diabetes are excluded to avoid overfitting the model at its development stage. For instance, features, such as data sequence number, hospital ID, time, and date should be removed. All the features (diabetes risk factors) available in the dataset can be used or a subset of features can be selected by applying feature selection algorithms [92] or taking an expert's advice or using a hybrid approach. In our proposed system we use Recursive Feature Elimination [93] which selects the set of features that are more relevant to the incidence of diabetes.
- *Data splitting:* In this stage, data is split for training (model development) and testing. This is done by dividing the dataset into 70% and 30% for training and validation respectively.
- *Model Development:* In this stage, k-fold cross-validation technique [94] is used to develop the model with the preprocessed training data. In the proposed system, we use decision-tree random forest (RF) classification model [95] as it is the most frequently used algorithm in the diabetes literature [8–15].
- *Model Evaluation:* In this stage, the developed model is evaluated using validation data in terms of accuracy, precision, recall, F-measure, ROC, AUC, and execution time. F-measure is an important metric to evaluate the performance of a machine learning model when trained using an imbalanced dataset. This is because F-measure can reveal the ability of the model to detect both majority and minority classes [96].
- *Diabetes Prediction:* In this stage, the evaluated machine learning algorithm is used to predict the incidence or diagnose the prevalence of diabetes based on the risk factors data.

3.4. Blockchain

Security and privacy of healthcare data are the main requirements for a trustworthy and patient-centric system [16]. The cloud provides scalable computing and storage facilities for healthcare data. However, the involvement of third-party cloud service providers leads to increased security and privacy threats due to a lack of transparency and data integrity. Blockchain eliminates a centralized authority and ensures trust and transparency among the network participants [97]. The blockchain component in our proposed framework connects all the network participants in a peer-to-peer manner. The network participants involve allied health professionals, patients, pharmacies, medical experts, and hospitals. Each participant is authenticated by a certificate authority. Table 8 shows how blockchain addresses different security and privacy issues that prevail in an only cloud-based system.

We use a non-encapsulated integrated blockchain-cloud architecture [17], in which the diabetes risk factors data are stored in the cloud database and the associated meta-data is recorded in the blockchain, such as the hash of the risk factors data, update and query events, access control policy, and diabetes prediction results. Storing data in the cloud aids in system scalability, whereas recording meta-data in the blockchain ledger enables security and privacy. The hash of risk factors data and prediction results in the ledger ensures data integrity. In addition, recording data update and query events in the ledger discourages unauthorized access, leading to enhanced privacy. Furthermore, we employ a multi-ledger-based permissioned blockchain architecture that provides configurable access control rights and facilitates the development of a separate ledger for collaborating allied health professionals [98]. The selection of permissioned blockchain over permissionless [99] is due to the following disadvantages of the latter: 1) unauthorized participation in the network leading to impersonate account holders, 2) clear transaction data in the ledger accessible to each network participant revealing sensitive patients' data, 3) slow network throughput hindering real-time patient's treatment, and 4) the need to pay transaction execution fees and mining rewards limiting the usability of the network.

The blockchain component consists of participants, assets, transactions, and events. Table 9 shows the different types of participants, assets, transactions, and events that will be used in our system along with their descriptions.

Fig. 3 shows the blockchain usage in our end-to-end AI-based prognosis/diagnosis support system for healthcare management. In addition to the network participants described in Table 9, the system consists of a certificate authority (CA) and a medical expert. The CA works as both

Table 9
Description of participants, assets, transactions, and events for the proposed blockchain network.

	Name	Description
Participants	Hospitals	Responsible for uploading medical records to the cloud, validating healthcare transactions, and responding to the data retrieval query. They store a copy of the ledger.
Participants	Allied health professionals	They are the doctors and nurses registered with the hospitals. They are responsible for updating patients' medical records based on symptoms, diagnoses, treatments, and medications. They can also update the laboratory and pathological results. In addition, they can query the medical records from the cloud by performing query transactions.
Participants	Pharmacists	They are responsible for updating the information related to medications, bills, and insurance claims to the cloud. This is by performing update transactions. In addition, they can query a patient's records.
Participants	Patients	They are the diabetic and pre-diabetic patients registered with the hospitals. They can query their medical data and update contextual and lifestyle data to the cloud by performing update transactions. The patients can enter the data into the network using mobile phones.
Participants	External users	They are the participants not necessarily registered with the hospitals. They can insert their lifestyle, medical conditions, hereditary, psychosocial, and demographic data to predict the development of diabetes.
Assets	Laboratory and pathological data (by hospitals)	This asset includes laboratory and pathological test data such as blood and urine reports, x-rays, MRIs, ultrasound, endoscopy, fasting plasma glucose, uric acid level, etc. These data are updated by the hospitals to the cloud with the hash of the data being recorded in the blockchain. The data is made available to the corresponding patient upon a data retrieval query.
Assets	Medical condition data (by hospitals)	This asset includes the medical condition data such as symptoms, diagnosis, medications, treatments, and vitals, i.e., heart rate, blood pressure, oxygen level, cholesterol level, and BMI. These data are sent to the cloud for storage with meta-data recorded in the blockchain.
Assets	Social and contextual data (by patients)	This asset includes the social and contextual data such as age, gender, family history of diabetes, history of heart disease, depression, ethnicity, geographical location, smoking habits, alcohol consumption, diet, sleep duration, physical activity, educational level, and socioeconomic status. These data are sent as transactions by the patients for ledger updates.
Assets	Risk factors data (by external users)	This asset includes the diabetes risk factors data such as lifestyle, medical condition, hereditary, psychosocial, and demographic. These data are sent by external users as transactions for the prediction of diabetes incidence.
Transactions	Medical records update (by hospitals)	This transaction involves the update of the patient's medical records by the hospitals to the cloud. The meta-data is recorded in the ledger.
Transactions	Laboratory and pathological results update (by hospitals)	This transaction involves the update of the patient's laboratory and pathological results by the hospitals to the cloud. The meta-data is recorded in the ledger.
Transactions	Social and contextual data update (by patients)	This transaction involves the update of the social and contextual data by the patients to the cloud. The meta-data is recorded in the ledger.
Transactions	Query (from patients to hospitals)	This transaction involves the data retrieval request by the patient to the registered hospital for his/her medical data.
Transactions	Response to a query (from hospitals to the patients	This transaction involves the response from the hospital to the data retrieval query made by the patient.
Transactions	Risk factors data (from external users to AI-based prediction system)	This transaction involves the risk factors data sent by the external user as transactions for the prediction of diabetes incidence. The prediction request to the AI-based system will be recorded as a transaction in the ledger.
Transactions	Risk of diabetes incidence (from AI-based prediction system to the external users)	This transaction involves the prediction result regarding the development of diabetes. This data is used by the hospitals to develop a prevention plan.
Events	Patient's medical records update (to patients)	The patient is notified about his/her records being added to the cloud by the corresponding hospital. This notification helps the patient to be up-to-date with his/her records.
Events	Patient's laboratory and pathological results up-date (to patients)	The patient is notified about his/her laboratory and pathological results being added to the cloud by the corresponding hospital. This notification helps the patient to be up-to-date with his/her results.
Events	Patient's social and con-textual data update (to hospitals)	The hospital receives the social and contextual data transaction from the patient requesting to be added to the cloud. Upon validation, the data is added to the ledger.
Events	External user's risk factors data update (to hospitals)	The hospital receives the risk factors data transactions by the external to update the ledger.
Events	External user's prediction update (to hospitals)	The hospital updates the prediction results of the AI-based prognosis/diagnosis system in the blockchain ledger. This will aid in the development of a nationwide prevention plan.

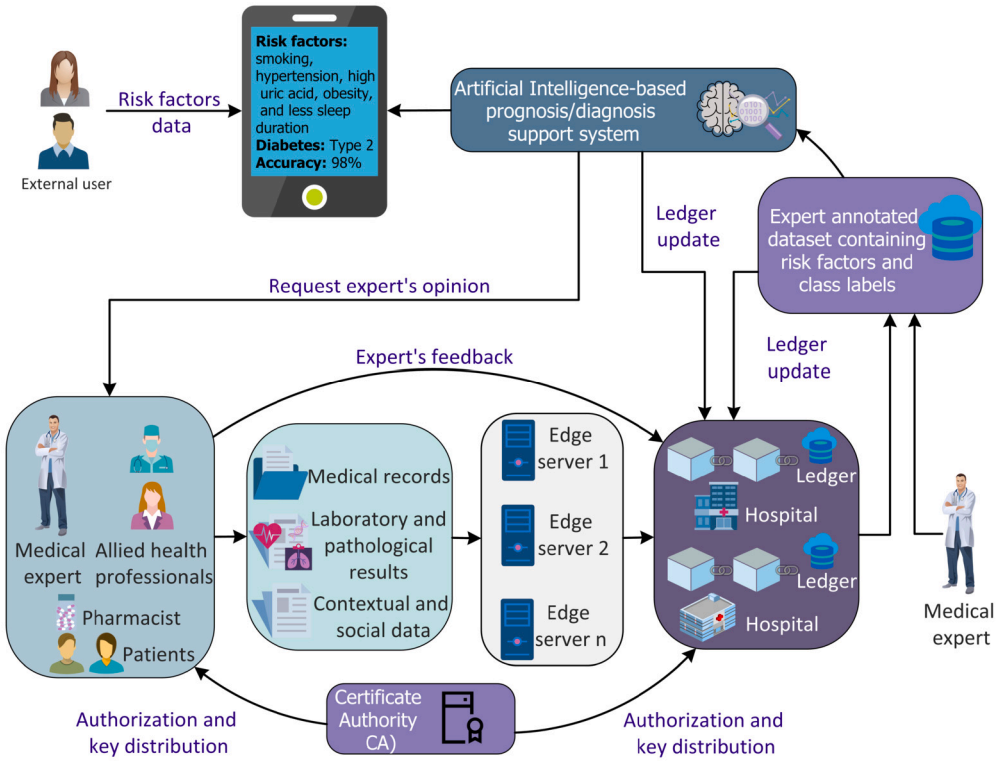


Fig. 3. Proposed blockchain and artificial intelligence integrated monitoring system for prediction of diabetes mellitus.

Table 10
Health data and corresponding attributes used in the proposed system.

Data	Attributes
Laboratory and pathological results	X-rays, MRIs, CT scans, blood report, and urine report
Medical records	File number, patient ID, patient name, age, gender, nationality, national identity number, medical insurance number, contact details, patient name, height, weight, waist circumference, body temperature, blood pressure, the reason for attendance, patient medical history, family medical history, allergies, symptoms, diagnosis, point of care testing (random blood sugar, urine dip, pregnancy test), medications
Social and contextual data	Age, diet, sleeping pattern, heart rate, physical activity, smoking habits, alcohol consumption
Risk factors data	High-level serum uric acid, sleep quality/quantity, smoking, depression, cardiovascular disease, dyslipidemia, hypertension, aging, ethnicity, family history of diabetes, physical inactivity, and obesity

a system administrator by removing malicious nodes from the network and an authority management entity by generating and distributing digital certificates. A participant's public-private key pair is also generated by the CA. The public-private key pair for each participant is linked to the participant ID, a secret PIN code set by the participant, and the participant identity proof. In a situation where the participant loses his/her public-private key pair, a new pair is generated by the CA after authenticating the participant ID, secret PIN code, and identity proof. Each network participant, i.e., patient, allied health professionals, pharmacists, and external users, is identified using an identity number. For instance, a patient is identified by the patient ID whereas a doctor is identified by the doctor ID. A medical expert is responsible for annotating the diabetes risk factors and class labels to the medical records data present in the cloud. The hash of annotated data is stored in the blockchain ledger. In addition, the medical expert will give feedback on the performance of the AI system when asked for an opinion. The expert's feedback is recorded as a transaction in the cloud with its hash in the blockchain. The AI-based prognosis/diagnosis support system consists of classification learning models for diabetes prediction. The prediction query and results are stored as transactions in the blockchain. Table 10 shows the data and the corresponding attributes used in the proposed system.

4. Implementation of proposed automated end-to-end blockchain artificial intelligence-system for diabetes mellitus prediction

In this section, the implementation of the system is discussed. The system operates through two main functions: 1) *DP(user_risk)* which allows end-users to get diabetes prediction from the system through a front-end device (e.g. smartphone), and 2) *DPMT(df_{risk})*, the diabetes prediction model trainer, which trains or updates the system's AI model by using new labeled data.

For the first operative function, *DP(user_risk)*, the implementation diagram is shown in Fig. 4(a). The risk factor data (which is unlabeled) is collected from a data source D_{src}. The users' health records including diabetes risk factors data are stored in the cloud with meta-data recorded in the blockchain. The raw risk factor data df_{risk} is fed as an input to the data transformation component for preprocessing. Data transformation is performed by edge servers. The preprocessed data frame df'_{risk} is then passed as input to the current Machine Learning model for diabetes prediction. The result of the prediction is sent back to the end-user device and the Blockchain ledger.

For the second operative function, *DPMT(df_{risk})*, the system is upgraded using previous modeling data and the new data generated by users and/or health professionals, which is already labeled by the health professionals and stored in the cloud, as shown in Fig. 4(b). This new

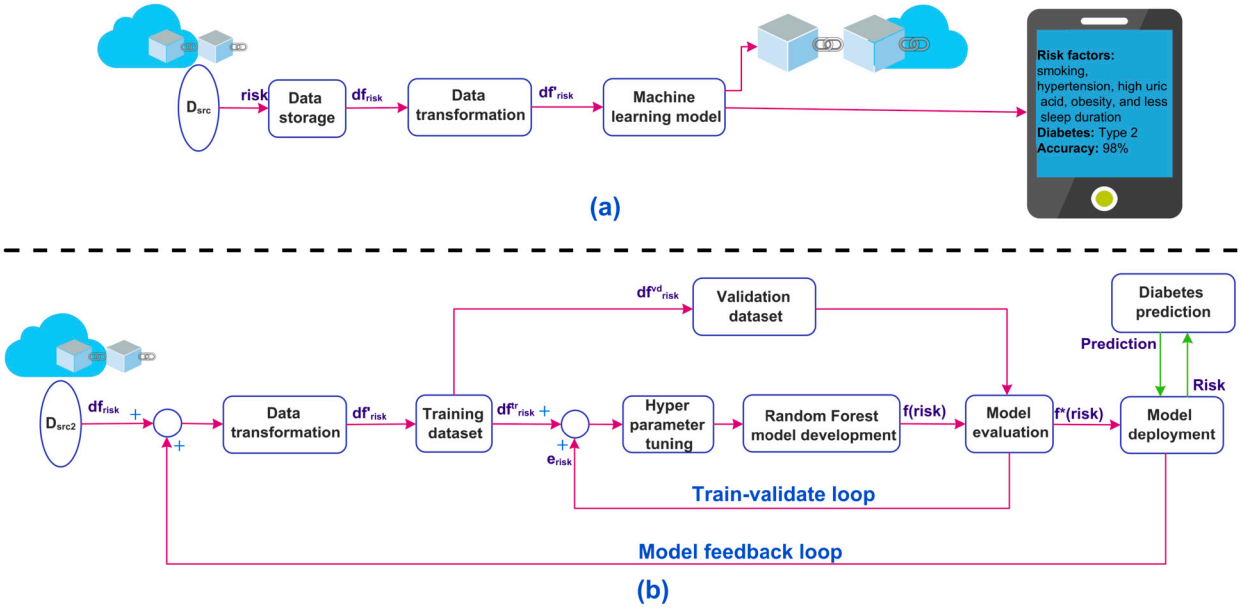


Fig. 4. Implementation of the proposed end-to-end automated artificial intelligence (AI)-blockchain systems for diabetes monitoring.

training data, D_{src2}, is extracted from the cloud data source to be fed as input to the data transformation component for preprocessing. The data extraction and preprocessing operations along with the metadata are recorded in the blockchain ledger. The preprocessed data is then divided into training df^{tr}_{risk} and validation df^{vd}_{risk} datasets. The selected Random Forest model *f(risk)* is trained again using df^{tr}_{risk}. The performance of the model is evaluated using df^{vd}_{risk}. The model development is a feedback control process where the model is tuned using hyperparameter tuning unless the desired performance is obtained. The diabetes prediction error e_{risk} obtained from the evaluation of the prediction model is fed back to tune the hyperparameters. The tuned model $f^*(risk)$ is deployed in the system for predicting accurately the risk of diabetes occurrence in users. Consequently, the diabetes prediction function, *DP(user_risk)*, uses the deployed model to predict the risk of diabetes.

5. Method

The proposed system consists of four components: 1) data collection and storage, 2) data preprocessing, 3) machine learning model development and validation, and 4) machine learning model deployment. In this paper, we evaluate the data preprocessing, machine learning, model development, and validation components. The measurable risk factors data can be collected using the medical tests and devices that are mentioned in Section 3.1. However, in this paper, we utilize publicly available diabetes datasets, such as the PIMA Indian dataset [100] and MIMIC III [101] which include diabetes risk factors data, and Sylhet [102] dataset which consists of diabetes-related symptoms. These datasets contain observations for patients who are admitted to a critical care unit, with a number of patients having diabetes. Each dataset contains two class labels, i.e., diabetic or non-diabetic. Consequently, these labeled datasets are used to train supervised classification algorithms for diabetes prediction. The datasets are selected in a way that they include as many diabetes risk factors as possible. The experiments are performed on separate datasets with no integration among them. This is because these datasets have different sets of features associated with patients. We assess the performance of the most utilized machine learning models in the literature for diabetes prediction, specifically Random Forest (RF), Logistic Regression (LR) [103], and Support Vector Machine (SVM) [104], with and without feature selection, before and after data balancing. The models are evaluated using different metrics such as Accuracy, F-measure, precision, recall, and AUC.

5.1. Datasets

The three datasets used in this paper are: 1) PIMA India, made available by the National Institute of Diabetes and Digestive and Kidney Diseases, 2) Sylhet, collected using direct questionnaires from the patients of Sylhet Diabetes Hospital in Sylhet, Bangladesh, and 3) MIMIC III, a large dataset which contains information of over 40,000 patients who stayed in critical care units of the Berth Israel Deaconess Medical Center between 2001 and 2012. Table 11 shows the characteristics of the datasets used to evaluate our proposed system. The information about the prevalence of diabetes for patients in MIMIC III datasets is not explicitly mentioned. Consequently, we presented the Diabeteic/Non-Diabetic class as 'Not Explicitly Mentioned' in Table 11. However, we extracted the class for each patient during the pre-processing stage (Section 5.3, Table 13) by using the ICD9 coding scheme of the dataset.

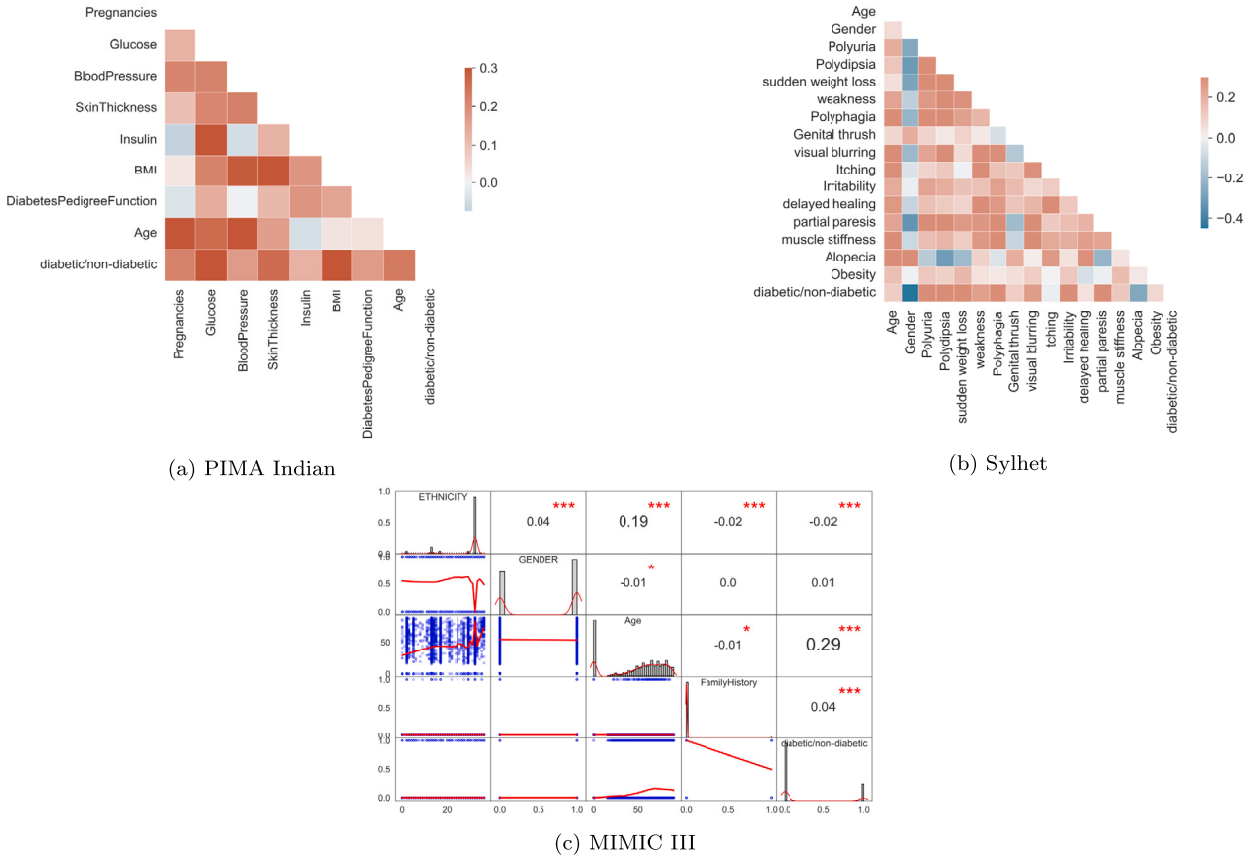
5.2. Data exploration

For each dataset, we investigate the correlations between features and the diabetic/non-diabetic class. We choose the Phik (Φk) correlation coefficient because that works consistently between categorical, ordinal, and interval variables. It captures non-linear dependency and reverts to the Pearson correlation coefficient in the case of bi-variate normal input distribution [105]. So, it encompasses multiple types of correlations. As shown in Fig. 5a, PIMA India presents a logical correlation between Age and Number of Pregnancies. Regarding diabetes detection, the features that are correlated with the diabetic/non-diabetic outcome of the patient are Glucose, Age, BMI, Insulin, and Skin Thickness. In addition, BMI is correlated with Blood Pressure. Correlations for Sylhet are displayed in Fig. 5b. In this dataset, the diabetic/non-diabetic outcome is highly correlated with Polydipsia and Polyuria and in a lower manner with partial paresis, gender, and sudden weight loss. Furthermore, Polydipsia and Polyuria are highly correlated with each other. Similar to PIMA India, MIMIC III dataset shows high correlations between the class (diabetic/non-diabetic) outcome and the Age feature (Fig. 5c). However, Fig. 5c shows a correlation between Age and Ethnicity that may indicate the randomness in the MIMIC III dataset under study.

Table 11
Original datasets characteristics.

Dataset	Features	Positive Classes	Negative Classes	Total Records
PIMA Indian	Pregnancies, Glucose, Blood pressure, Skin thickness, Insulin, BMI, Diabetes pedigree[1], and Age	268 (34.9%)	500 (65.1%)	768
Sylhet	Age, Gender, Polyuria[2], Polydipsia[3], sudden weight loss, weakness, Polyphagia[4], Genital thrush[5], visual blurring, Itching, Irritability, delayed healing, partial paresis[6], muscle stiffness, Alopecia[7], and Obesity	320 (61.5%)	200 (38.5%)	520
MIMIC III	Ethnicity, Gender, Age, and Family history of diabetes	Not Explicitly Mentioned	Not Explicitly Mentioned	46,520

[1]Diabetes pedigree provides a synthesis of diabetes history in relatives and the genetic relationship of those relatives to the subject [100].
[2]Polyuria is a condition where the body urinates more than usual and passes excessive or abnormally large amounts of urine each time you urinate [102].
[3]Polydipsia is the feeling of extreme thirstiness [102].
[4]Polyphagia, also known as hyperphagia, is the medical term for excessive or extreme hunger [102].
[5]Genital thrush is a common infection caused by an overgrowth of yeast [102].
[6]Paresis involves the weakening of a muscle or group of muscles. It may also be referred to as partial or mild paralysis. Unlike paralysis, people with paresis can still move their muscles. These movements are just weaker than normal [102].
[7]Alopecia areata is an autoimmune disorder that causes your hair to come out, often in clumps the size and shape of a quarter [102].



(a) PIMA Indian

(b) Sylhet

(c) MIMIC III

Fig. 5. Correlation between features and diabetic/non-diabetic class for PIMA Indian, Sylhet, and MIMIC III datasets.

5.3. Data preprocessing

PIMA Indian dataset shows a number of missing values in some numerical features. As shown in Fig. 6, blood pressure, skin thickness, and BMI features contains observations having zero value. Consequently, we remove these observations. The Sylhet dataset does not have any missing values. Regarding the MIMIC III dataset, there is a need to extract the available risk factor features from the raw data. MIMIC III raw data are split into different tables. The data of interest in MIMIC III for bringing out risk factors and diabetic/non-diabetic class is shown in Table 12. We build the preprocessed MIMIC III dataset by joining information of the patients from the different data tables (Table 12). For each patient, we retrieved age and ethnicity data from PATIENTS and ADMISSIONS tables respectively. Similarly, for the family history of diabetes, we determined if a patient has ICD9 diagnostic code V180 (Family history of diabetes mellitus). The categorical values of 'UNKNOWN/NOT SPECI-

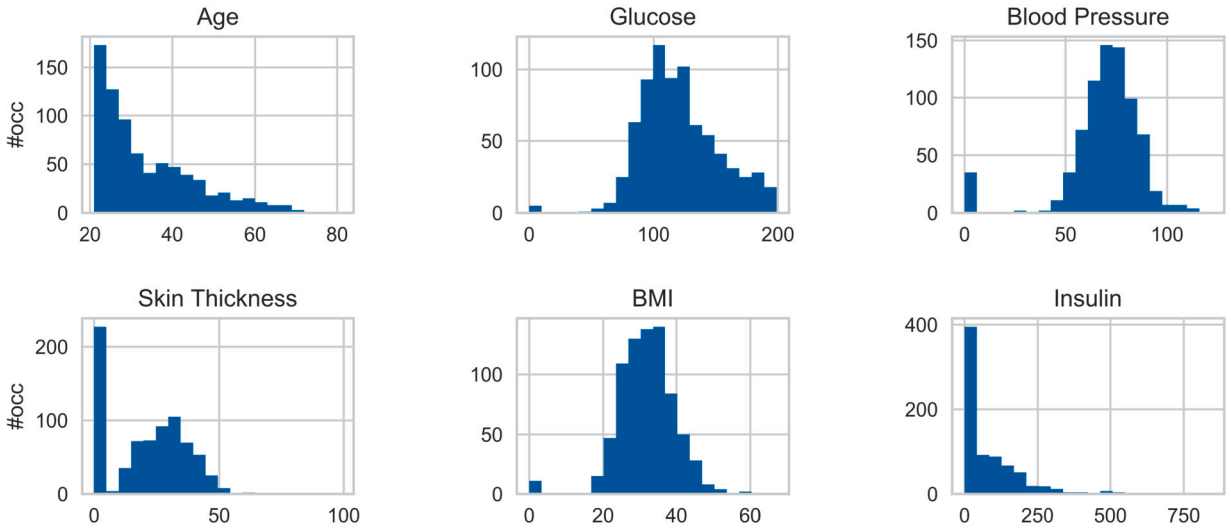


Fig. 6. Data exploration histograms for PIMA Indian dataset numerical features.

Table 12
Data tables used from MIMIC III dataset.

Table Name	Available data and purpose
PATIENTS	Subject ID, Gender, Date of Birth
ADMISSIONS	Subject ID, visits of a patient, start and end of the patient visit, other demographic data (Ethnicity)
DIAGNOSES_ICD	Subject ID, Association of ICD9 diagnostics with patients
D_ICD_DIAGNOSES	Dictionary of ICD9 codes associated with their description

Table 13
Dataset characteristics after preprocessing.

Dataset	# of Features	Positive Classes	Negative Classes	Total Records
PIMA Indian	8	177 (33.3%)	355 (66.7%)	532
Sylhet	16	320 (61.5%)	200 (38.5%)	520
MIMIC III	4	8,820 (22.5%)	30,469 (77.5%)	39,289

FIED', 'PATIENT DECLINED TO ANSWER', and 'UNABLE TO OBTAIN' for Ethnicity in the MIMIC III dataset are interpreted as missing values. Consequently, patients with such values for Ethnicity are removed from the dataset. The information about diabetic/non-diabetic outcomes is retrieved from the ICD9 diagnostics associated with the patient in table DIAGNOSES_ICD. If one of the diagnostics is for diabetes mellitus, then the patient is set to have diabetes. Table 13 shows the characteristics of the resulting datasets after preprocessing.

5.4. Feature selection

We use Recursive Feature Elimination, Cross-Validated (RFECV) [93] feature selection algorithm with RF [95] as a cross-validation evaluator. The selection of the RFECV algorithm is based on its performance in the literature [93]. The RF model is a type of bagging algorithm that combines a specific number of decision trees and is used to detect feature importance in learning. The tree-based random forest ranks the features according to how well the purity of the feature is improved, that is, a decrease in the impurity (Gini impurity) over all the trees. Features with the greatest decrease in impurity happen at the start of the trees, while features with the least decrease in impurity occur at the end of the trees. Therefore, by pruning trees below a particular feature, one can create a subset of the most important features. Recursive Feature Elimination works by searching for a subset of features by starting with all features in the training dataset and successfully removing features until the desired number remains. This is achieved by fitting random forest, ranking features by importance, discarding the least important features, and re-fitting the model. This process is repeated until a specified number of features remains.

5.5. Balancing data augmentation

The datasets under study are imbalanced towards the negative class. The number of diabetes class observations is roughly 30% for PIMA Indian and Sylhet datasets and 22% for MIMIC. To reduce the biases in the created models, the synthetic minority oversampling technique (SMOTE) is used as a data balancing technique based on its performance in literature [106]. SMOTE is an oversampling technique that increases the number of minority class samples in the dataset, by generating new samples from existing minority class samples. The application of SMOTE to clinical datasets can improve model performance by reducing the negative effects of imbalanced data as observed in recent literature. SMOTE is only applied on the training/validation split (70%) of the data samples so that the model sees equal numbers of both class types, and the test split (30%) is not modified.

5.6. Machine learning models for diabetes prediction

We implement our proposed automated end-to-end blockchain AI-system for diabetes prediction using the most popular and accurate RF

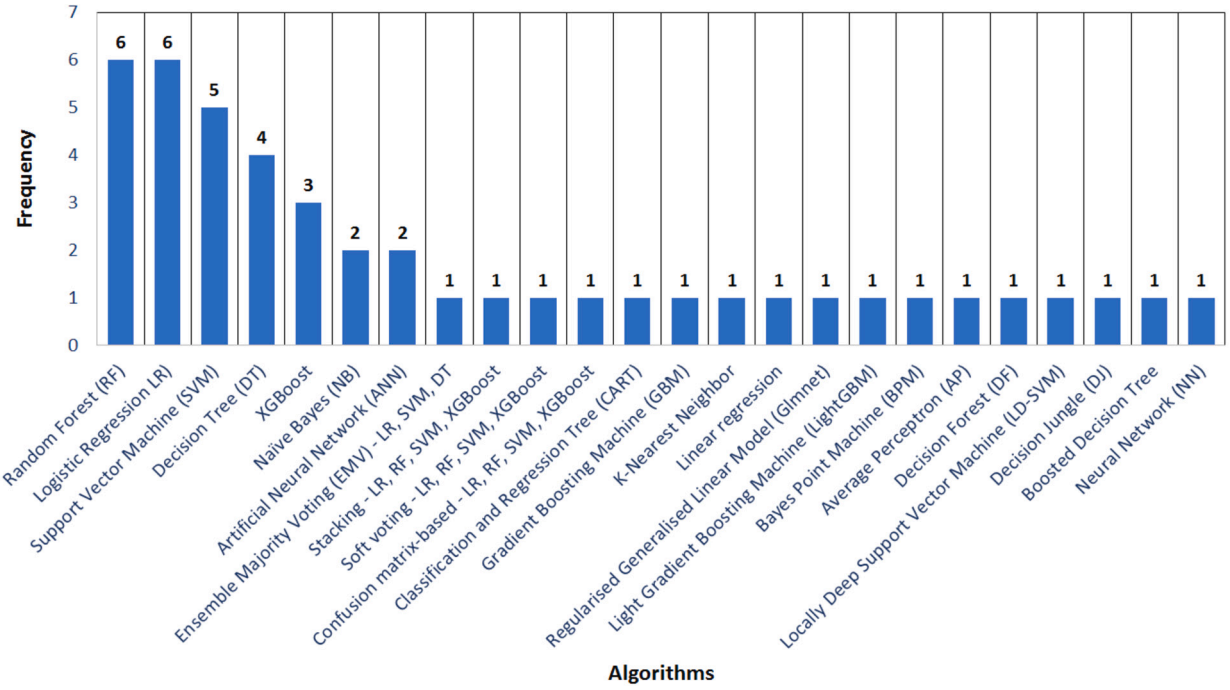


Fig. 7. Frequency of classification algorithms used in literature for diabetes prediction.

model. Fig. 7 shows the relative usage frequency of different Machine Learning algorithms in current research papers on Diabetes prediction [8–15]. Furthermore, we compare the performance of RF with LR and SVM algorithms for diabetes prediction. The selection of LR and SVM is based on their popularity as shown in Fig. 7.

5.6.1. Random forest (RF)

This algorithm is based on Decision Tree (DT), which constructs a tree structure to define the sequences of decisions and outcomes, and to use it for prediction. At each node of the tree, the algorithm selects the branch having the maximum information gain.

Random Forest is a set of decision trees constructed using randomly selected samples of the dataset [95]. It performs voting on the output of each decision tree and classifies an observation into diabetes or non-diabetes depending on the majority of the decision trees' output.

5.6.2. Logistic regression (LR)

This algorithm predicts the probability that a given observation belongs to the diabetes or non-diabetes class using a sigmoid function [103] as stated in Equation (1).

$$P(diabetes) = \frac{e^{\beta_0} + \sum_{i=1}^{n} \beta_i R_i}{1 + e^{\beta_0 + \sum_{i=1}^{n} \beta_i R_i}} \tag{1}$$

where p(diabetes) represents the probability of having diabetes, R is the set of risk factors, and β_0 and β_i are the regression coefficients representing the intercept and the slope respectively. The values of regression coefficients are calculated using maximum likelihood estimation such that the value of Equation (2) is the maximum.

$$l(\beta_0, ..., \beta_1) = \prod_{i, y_i=1} P(diabetes) \prod_{i, y_i=0} (1 - P(diabetes)) \tag{2}$$

5.6.3. Support vector machine (SVM)

This algorithm aims to create a decision boundary known as a hyperplane that can separate n-dimensional instance space into diabetes and non-diabetes classes. The hyperplane is created using the extreme points (support vectors) of the dataset. The generation of a hyperplane is an iterative process to find the maximum possible margin between the support vectors of the opposite classes. Let $r^{(i)}$ and $y^{(i)}$ represent the risk factors and classes in the dataset and there exists a hyperplane that separates diabetes and non-diabetes classes as stated in Equation (3).

$$\begin{gathered} w^T r + b = 0 \\ w^T r^{(i)} + b > 0, if\, y^{(i)} = +1 \; and \; w^T r^{(i)} + b < 0, if\, y^{(i)} = -1 \end{gathered} \tag{3}$$

where w is the normal of the hyperplane and b is the bias. The minimization problem to obtain the optimal hyperplane that maximizes the margin can be formulated using.

$$Minimize \Phi(W) = \frac{1}{2}||W||^2, such\ that\ y_i(W.r_i + b) \geq 1 \tag{4}$$

6. Performance evaluation

We evaluate the models under study with and without feature selection, before and after balancing, using the tenfold cross-validation method where the dataset is divided into k (k = 10) partitions. One partition is for validation data and k-1 partitions are for training with replacement. This is repeated until each partition is used for training and validation. The resultant model is then obtained by averaging the result of each iteration. For SVM, we use the polynomial kernels. Each model is executed 10 times on each dataset and the average for accuracy, F-measure, precision, recall, AUC, and execution time (for training and validation) is calculated. The use of Accuracy as a comparative metric between the models is justified because the datasets are not heavily imbalanced. AUC is the measure of the ability of a classifier to distinguish between classes and is used as a summary of the ROC curve. The higher the AUC, the better the performance of the model at distinguishing between the positive and negative classes. The accuracy, F-measure, recall, and precision are calculated using Equations (5) and (6) respectively. Recall and precision for the positive (negative) class are calculated using Equations (7) and (8) respectively.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{5}$$

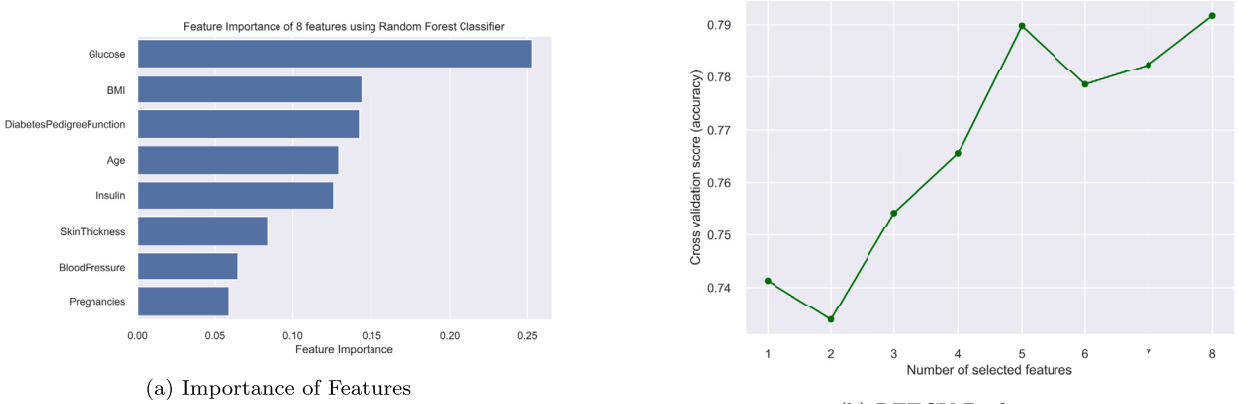
$$F - measure = \frac{2(Recall \times Precision)}{Recall + Precision} \tag{6}$$

$$Recall = \frac{TP(TN)}{TP(TN) + FN(FP)} \tag{7}$$

Table 14
Value(s) of hyperparameters used and optimal values for hyperparameters obtained in our experiments.

Algorithm	Hyperparameter	Values used in our experiments	Optimal values											
			1	2	3	4	5	6	7	8	9	10	11	12
Random Forest	Number of estimators/trees	100 [7,13], (300, 500, 1000) [13], 20, 40, 60, 80, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000	50	40	50	50	20	100	50	50	20	100	50	50
	Splitting criteria	entropy and Gini	entropy				Gini			entropy	Gini			entropy
	Maximum features	Nmax*, sqrt, and log2	None	sqrt		log2	sqrt							
	Max depth	None, 2, 5, 8	5		None	8	None							
Support Vector Machine	Regularization parameter	(0.001, 0.01, 0.1, 1, 2, 3, 5, 7, 10) [13], 4, 6, 8, 9, 10	1			7	1							
Logistic Regression	Regularization parameter	2^{-6}, 2^{-4}, 2^{-2}, 2^{0}, 2^{2}, 2^{4}, 2^{6}	16		0.25	4	16	4			16	4		
	Solver	Newton-cg, lbfgs, liblinear, sag, and saga	lbfgs	liblinear		lbfgs								

Nmax*: Number of features in the dataset, Newton-cg: Newton Conjugate Gradient,
1 - PIMA Indian: no feature selection and no balancing, 2 - PIMA Indian: feature selection and no balancing,
3 - PIMA Indian: feature selection and balancing, 4 - PIMA India: no feature selection and balancing,
5 - Sylhet: no feature selection and no balancing, 6 - Sylhet: feature selection and no balancing,
7 - Sylhet: feature selection and balancing, 8 - Sylhet: no feature selection and balancing,
9 - MIMIC IIII: no feature selection and no balancing, 10 - MIMIC III: feature selection and no balancing,
11 - MIMIC III: feature selection and balancing, 12: MIMIC III - no feature selection and balancing.



(a) Importance of Features

(b) RFECV Performance

Fig. 8. Performance of feature selection algorithms for PIMA Indian dataset.

$$Precision = \frac{TP(TN)}{TP(TN) + FP(FN)} \quad (8)$$

where TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative. TP (TN) represents the number of observations in the positive (negative) class that are classified as positive (negative), and FP (FN) represents the number of observations in the negative (positive) class that are classified as positive (negative).

6.1. Hyperparameter tuning

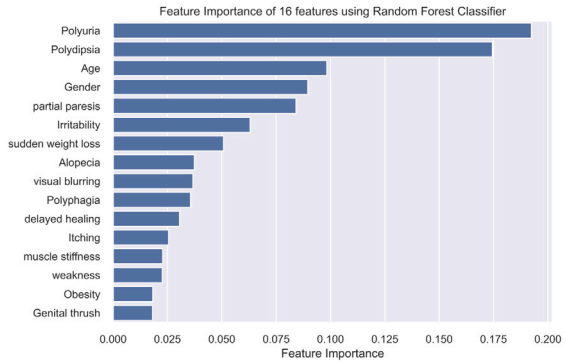
To achieve the best performance possible with the end-to-end system for diabetes prediction, we perform hyperparameter tuning, that determines the optimal values for machine learning models' parameters. Evaluating each model only on the training set can lead to overfitting. Consequently, to reduce the effect of overfitting we perform again stratified k-fold cross validation with k =10. The parameters we study for each algorithm, their ranges, and their optimal values are described in Table 14. The ranges are selected in a way that they include the values considered in literature. To perform the search for the best parameters, we use GridSearchCV from the Python library sklearn.model_selection module.

6.2. Feature selection

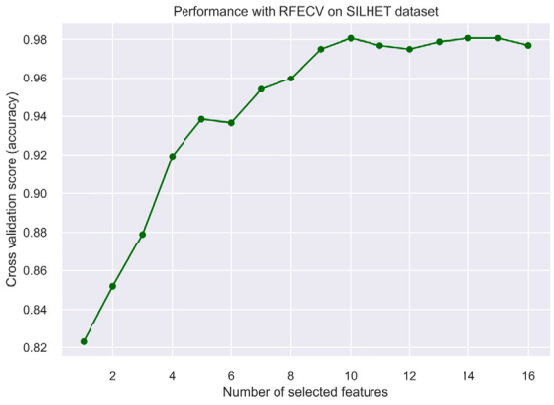
Following the feature selection method described in 5.4, we give the results for the three datasets.

For the PIMA Indian dataset, Fig. 8a shows the importance of each feature to prediction. It shows that glucose is the most important feature for the prevalence/incidence of diabetes in users, followed by BMI, insulin, age, and diabetes pedigree function. This is confirmed by studies in literature [107,108] and type 2 diabetes risk assessment form by the Finnish Diabetes Association [109]. There are 5 selected features: glucose, BMI, insulin, age, and diabetes pedigree function (Fig. 8b).

For the Sylhet dataset, Fig. 9a shows the importance of each feature to prediction. It shows that Polyuria and Polydipsia are the most important features in the prevalence/incidence of diabetes in users. This

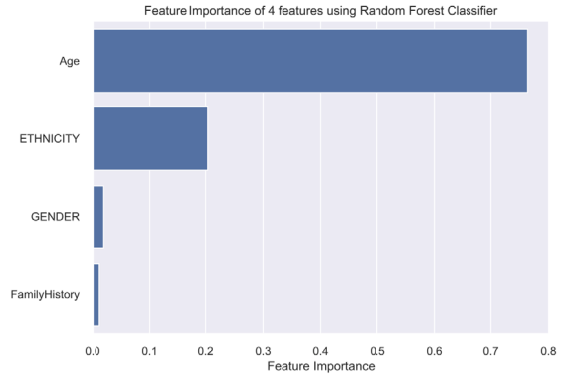


(a) Importance of Features

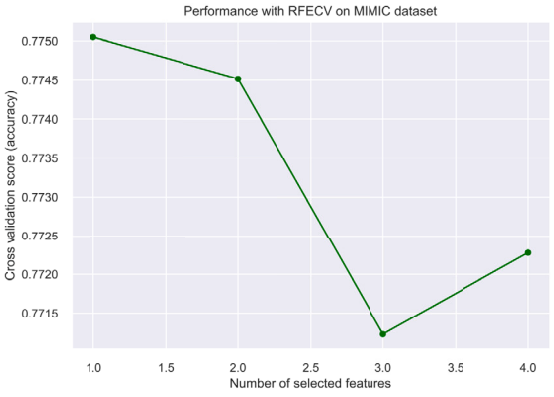


(b) RFECV Performance

Fig. 9. Performance of feature selection algorithms for Sylhet dataset.

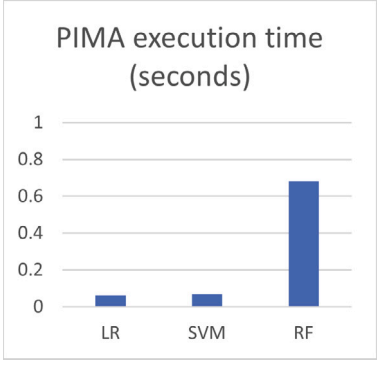


(a) Importance of Features

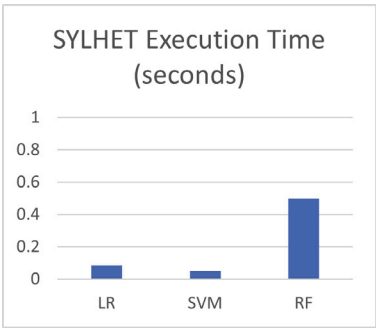


(b) RFECV Performance

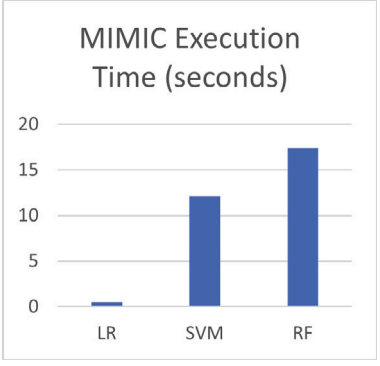
Fig. 10. Performance of feature selection algorithms for MIMIC III dataset.



(a) PIMA Indian



(b) Sylhet



(c) MIMIC III

Fig. 11. Execution time for logistic regression, support vector machine, and random forest algorithms for the datasets under study.

is in alignment with the result obtained in the literature [110]. In the context of gender, Fig. 9a reveals that men are more correlated with the prevalence/incidence of diabetes. This is confirmed by the American Diabetes Association's type 2 diabetes risk test [111]. There are 10 selected features: Polyuria, Polydipsia, Age, Gender, partial paresis, irritability, sudden weight loss, Alopecia, visual blurring, and Polyphagia (Fig. 9b).

In the MIMIC III dataset, there are 4 attributes which all related to diabetes risk factors: gender, age, ethnicity, and family history of diabetes. The feature importance graph, Fig. 10a, shows that age has the highest importance for the prevalence/incidence of type 2 diabetes in the population. This is in alignment with the American Diabetes Association's type 2 diabetes risk test [111]. The second important feature is the Black/African American ethnicity. This is also confirmed by studies in literature [112–115]. Furthermore, the gender feature has low importance and consequently, it was not selected. Fig. 10b shows that two features are selected as significant.

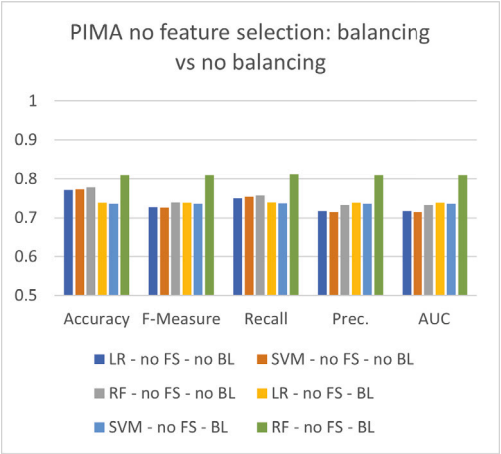
6.3. Hardware and execution time

The Hardware used for the performance analysis is Intel(R) Core (TM) i7-9700, with 32 Kilobytes of L1 data-cache, 32 Kilobytes of L1 instruction-cache, 256 Kilobytes of L2 Cache, 12 Megabytes of L3 Cache. The total execution times of each machine learning model under study for PIMA Indian, Sylhet, and MIMIC III datasets are shown in Figs. 11a, 11b, and 11c respectively. The measurements have been done using the tuned parameters for each model. They consist of the total time for training and validating the model. The time to make a prediction using a trained model is insignificant. The model needs to be re-trained only when there is a change in the previously used training dataset. For scalable training and re-training, a cloud server with higher computing capabilities can be used. We can observe that RF has the highest execution time. The main reason is that the number of estimators (n_estimator) is the principal parameter driving computational usage.

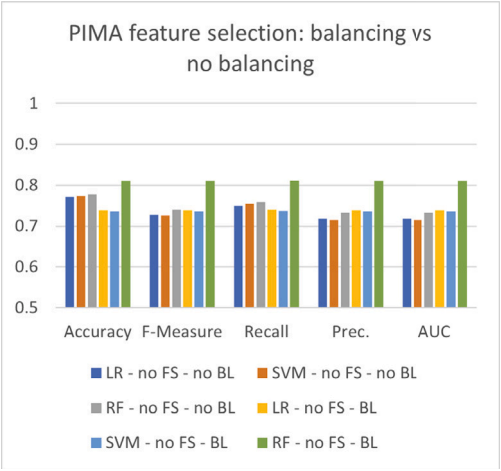
6.4. Experimental results analysis

In this section, our experimental results are analyzed and we provide insights into the reasons for the performance exhibited by the system. To compare the different metrics for the models under study Figs. 12, 13, and 14 display the Recall, Precision, Accuracy, and F-measure that was obtained when applying the models to the different datasets. The results are available in different settings: with feature selection or without and before data balancing or after. When focusing on accuracy, the results show that feature selection always improves this metric. The impact of balancing data is not always positive. The impact on accuracy depends on how imbalanced were the initial data. For the RF algorithm, there is an improvement in accuracy when balancing the data, the increase is 0.97 to 0.98 for the Sylhet dataset and 0.77 to 0.81 for the PIMA dataset. For the MIMIC III dataset, accuracy decreases from 0.77 to 0.66 with data balancing, but the F-measure increases from 0.51 to 0.66. This conclusion can be drawn for all datasets and algorithms, analysis of confusion matrices can give an insight on this. As shown by the confusion matrices (Figs. 15, 16, and 17) before and after feature selection and balancing, there is a better detection of the minority class after balancing. For the RF algorithm, the increase in detection of the minority class is 70% less false negative for PIMA Indian and 80% less false negative for the MIMIC III dataset. For the Sylhet dataset, there is no significant improvement because it is already balanced and there was no false negative before balancing.

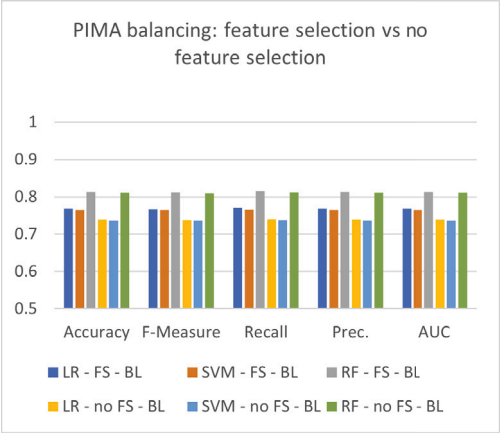
For PIMA India and Sylhet, execution times are very low. However, the execution time is much longer with the MIMIC III dataset. It can be explained because execution time increases with the number of features and observations. The datasets utilized in this study exhibit no significant imbalance, as a result, balancing the data does not improve



(a) Comparison between Balancing and No Balancing: No Feature Selection



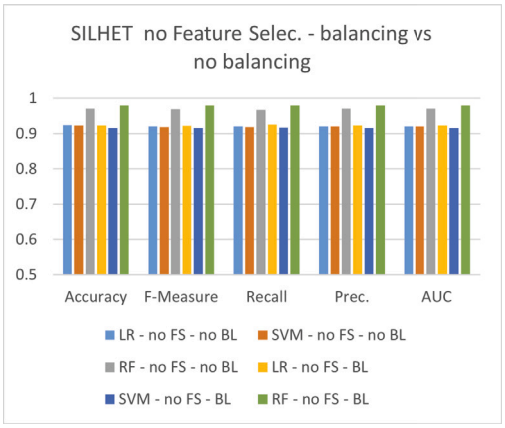
(b) Comparison between Balancing and No Balancing: Feature Selection



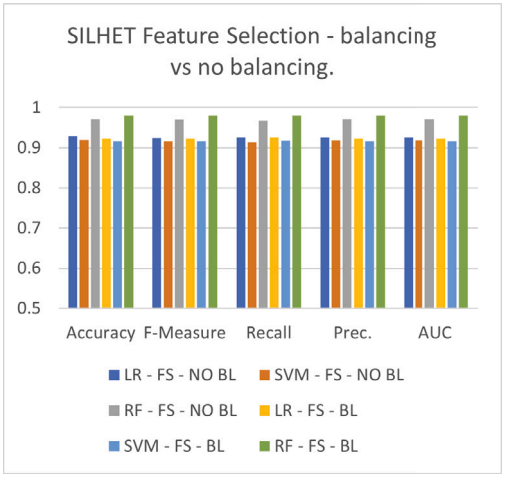
(c) Comparison between Feature Selection and No Feature Selection: Balancing

Fig. 12. Comparison of accuracy, F-measure, recall, precision, and AUC for the algorithms under study on PIMA Indian dataset (FS: Feature Selection, BL: Data Balancing).

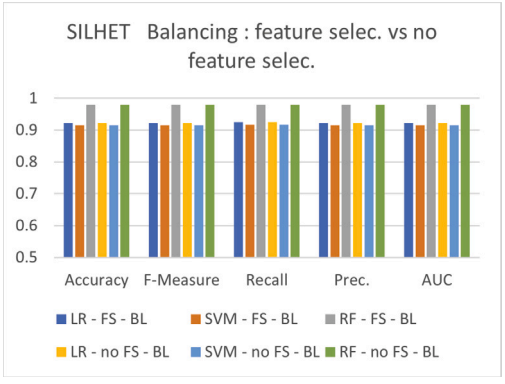
accuracy (except for the Sylhet dataset where a slight increase in accuracy is noted). This is particularly true for the MIMIC III dataset because it does not present enough risk factor features and it is also the most unbalanced. In this context, there is no proof of a better performance using balancing with the SMOTE algorithm. A root cause is that the



(a) Comparison between Balancing and No Balancing: No Feature Selection

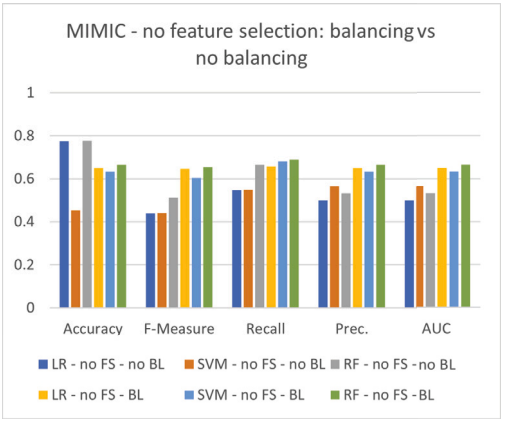


(b) Comparison between Balancing and No Balancing: Feature Selection

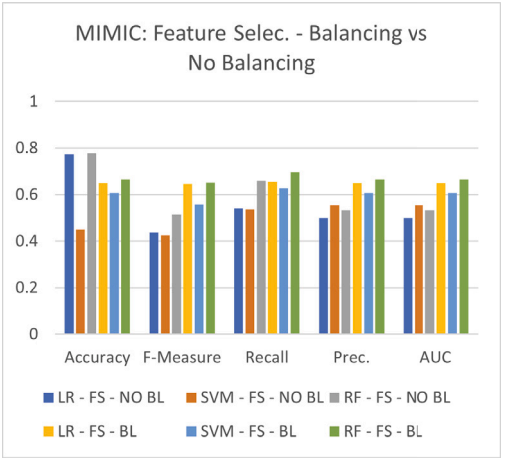


(c) Comparison between Feature Selection and No Feature Selection: Balancing

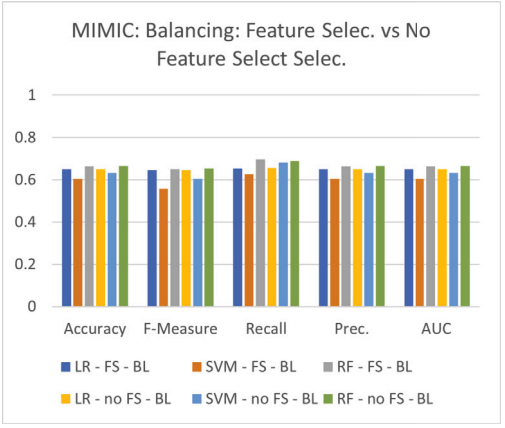
Fig. 13. Comparison of accuracy, F-measure, recall, precision, and AUC for the algorithms under study on Sylhet dataset (FS: Feature Selection, BL: Data Balancing).



(a) Comparison between Balancing and No Balancing: No Feature Selection



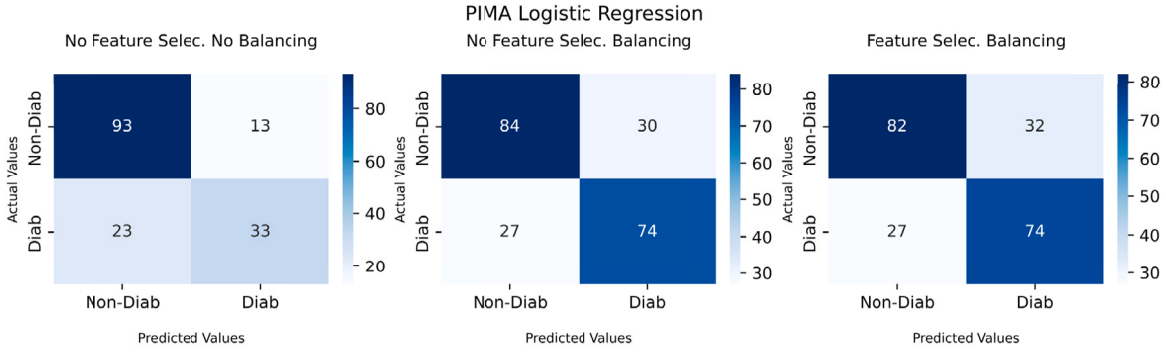
(b) Comparison between Balancing and No Balancing: Feature Selection



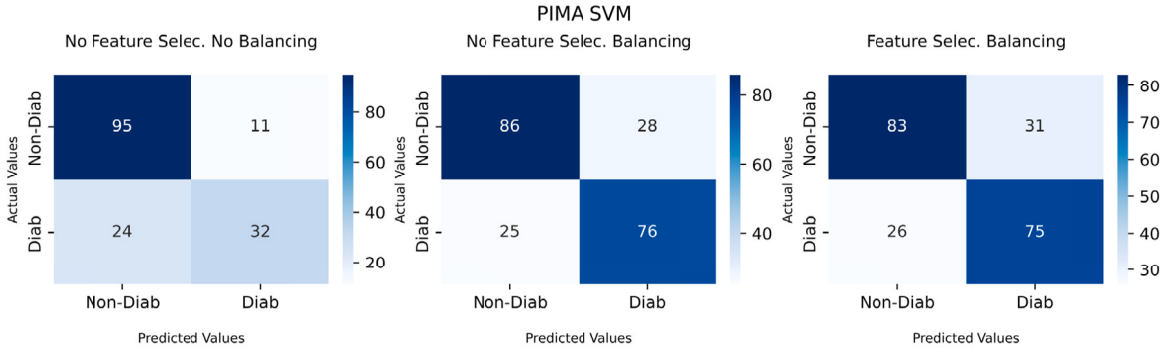
(c) Comparison between Feature Selection and No Feature Selection: Balancing

Fig. 14. Comparison of accuracy, F-measure, recall, precision, and AUC for the algorithms under study on MIMIC III dataset (FS: Feature Selection, BL: Data Balancing).

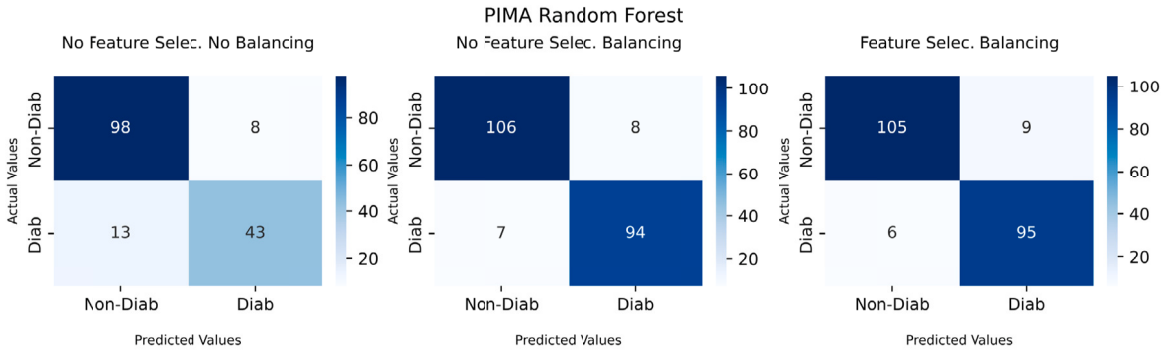
SMOTE algorithm augments the dataset with non-informative observations. A common observation is that the feature selection chosen in our system always improves the accuracy and also reduces the execution time. The most effective model with Feature Selection for the PIMA Indian dataset is RF demonstrating an accuracy score of 0.7827. The most effective model for Sylhet is RF with an accuracy score of 0.9723, adding Feature Selection results in a modest reduction in processing time, and there is no adverse impact on accuracy. The best model for the MIMIC III dataset is Logistic Regression, but Random Forest is very near. The best accuracy for LR is 0.7734 and the best for RF is 0.7703 and a difference of only 0.4%. ROC curve analysis shows that for the Sylhet dataset, the Random Forest classifier is very good. The performance outcomes for the Sylhet dataset significantly surpass those for the PIMA and MIMIC III datasets. The primary distinction among these datasets lies in the number of features available, which can be consid-



(a) PIMA Indian: Logistic Regression



(b) PIMA Indian: Support Vector Machine



(c) PIMA Indian: Random Forest

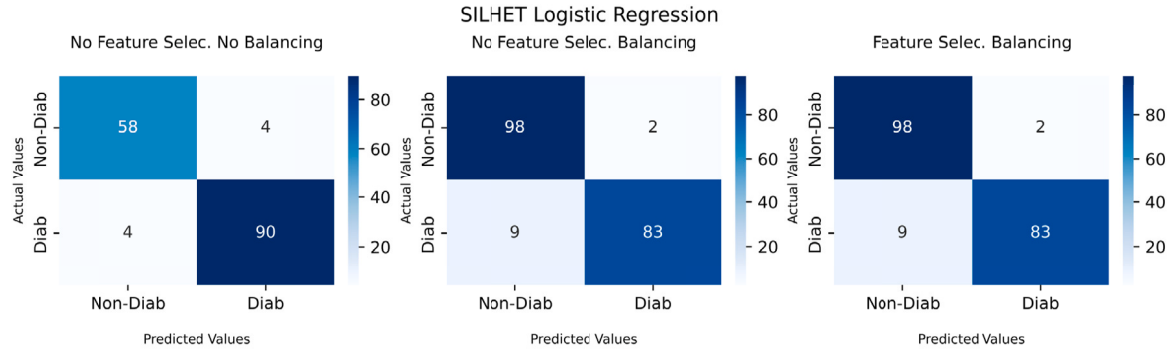
Fig. 15. Confusion matrices for PIMA Indian dataset.

ered as potential risk factors for diabetes. In conclusion, it is advisable to seek datasets with a comprehensive range of risk factors, as exemplified by the Sylhet dataset. The datasets utilized in our analysis lacked sufficient diversity to evaluate the necessity for data balancing. Using RF with feature selection adds value to the system because it can reduce training time.

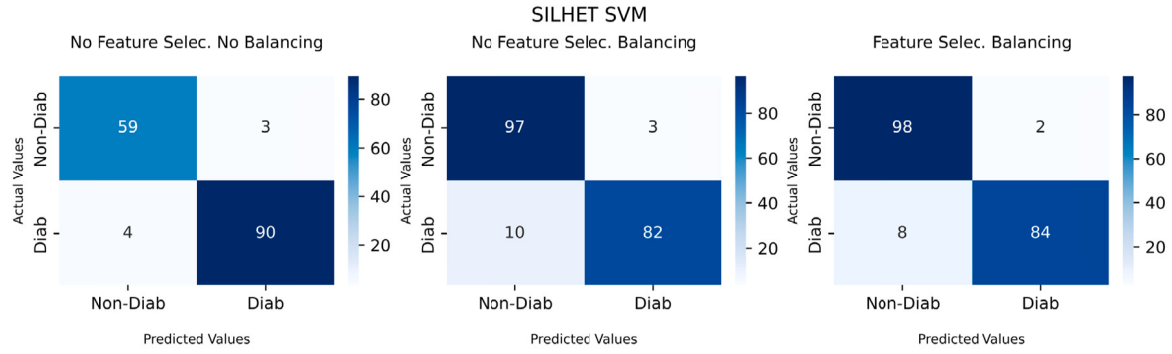
7. Conclusions

In this paper, we propose a novel end-to-end IoT-edge-AI-blockchain system for diabetes prediction. The proposed system consists of medical devices and sensors to obtain the value of different diabetes risk factors. A comparative analysis, in terms of performance and cost, of these devices and sensors is presented to aid readers in the selection of the most optimal test/device/sensor for measuring each risk factor. The proposed system employs edge computing to preprocess the diabetes risk factors data in terms of feature selection and data balancing. The preprocessed data is employed to train and validate machine learning model(s) in the cloud. Subsequently, the trained model is utilized at the edge to predict the risk of diabetes for users, based on the risk factors data. Furthermore, to ensure the privacy and security of the data, the proposed system is underpinned by the blockchain. The hash of medical records, risk factors data, machine learning model parameters, and diabetes prediction results are stored in a blockchain ledger. In addition, we identify the most significant risk factors for diabetes prediction and compare the performance of the most commonly employed machine learning algorithms in the literature, using three real-world diabetes datasets, in a unified setup. Our experimental results show that the RF is the most accurate compared to LR and SVM. The following requirements should be considered at the time of choosing a classification model in the context of type 2 diabetes prediction.

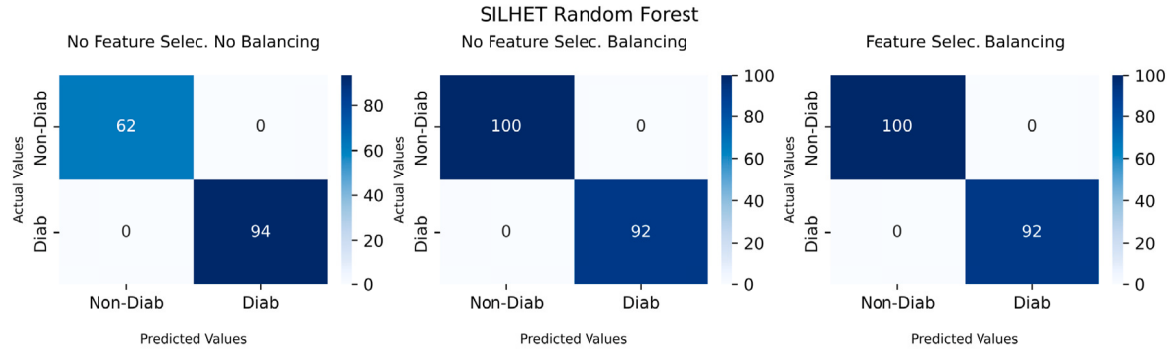
1. *Accuracy and F-measure:* High accuracy can be achieved by a large number of algorithms. However, if we evaluate the performance of a classification algorithm only using the accuracy, then misinterpretations of the results can occur. Because even if the algorithm achieves high accuracy, it might not be able to classify correctly observations that belong to the minority classes, as the F-measure



(a) Sylhet: Logistic Regression



(b) Sylhet: Support Vector Machine



(c) Sylhet: Random Forest

Fig. 16. Confusion matrices for Sylhet dataset.

would reveal, This is particularly true with imbalanced datasets, which are very frequent in the health domain. In such a context, there might be a risk of predicting a wrong result with serious medical consequences, because a patient who is diabetic could be classified as non-diabetic. As a result, we highly recommend selecting F-measure as one of the evaluation metrics.

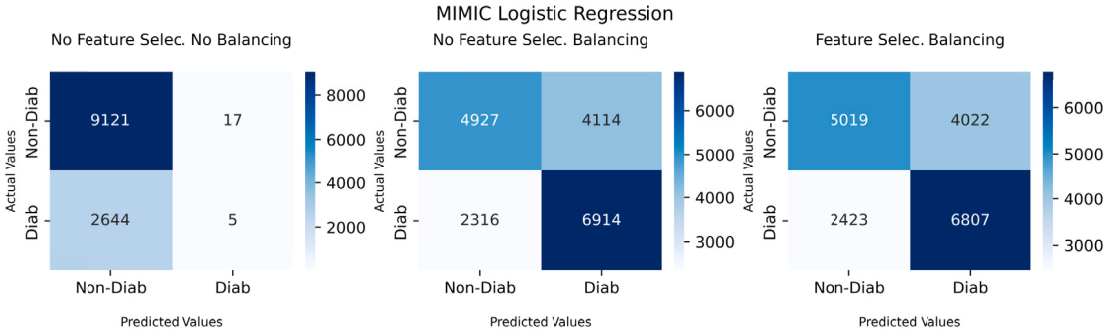
2. *Feature selection:* Feature selection algorithms should be used on the dataset before training the classification model. This can avoid overfitting and reduce execution time. The experiments we conducted show that feature selection does not incur accuracy degradation.
3. *Significant features:* As a recommendation, we propose to use age, ethnicity, glucose, family history of diabetes, and obesity for the prediction of type 2 diabetes. This is based on our experimental results obtained by implementing the RFECV feature selection algorithm. As revealed by our experimental results, using these selected features for machine learning model development, the performance of learning models remains the same or improves compared to using all the features. The selection of significant features in our results is in alignment with the Finnish Diabetes Association's type 2 diabetes risk assessment form [109] and the American Diabetes Association's type 2 diabetes risk test [111].

CRediT authorship contribution statement

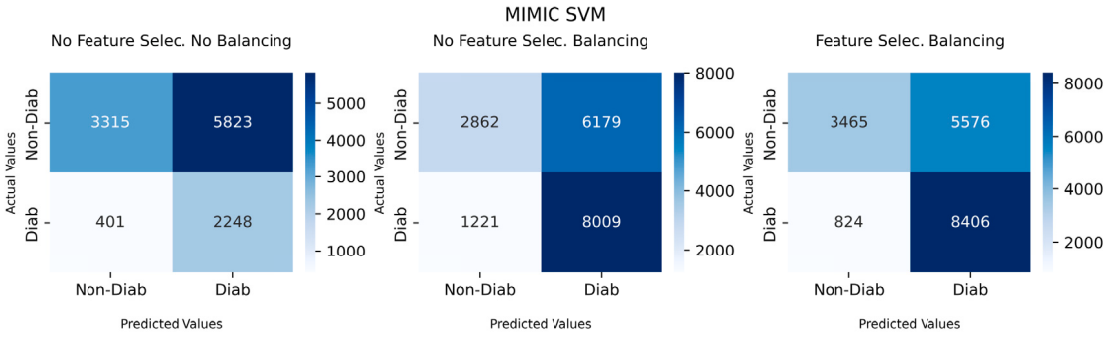
Alain Hennebelle: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **Leila Ismail:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing. **Huned Materwala:** Methodology, Writing – review & editing. **Juma Al Kaabi:** Writing – review & editing. **Priya Ranjan:** Writing – review & editing. **Rajiv Janardhanan:** Writing – review & editing.

Declaration of competing interest

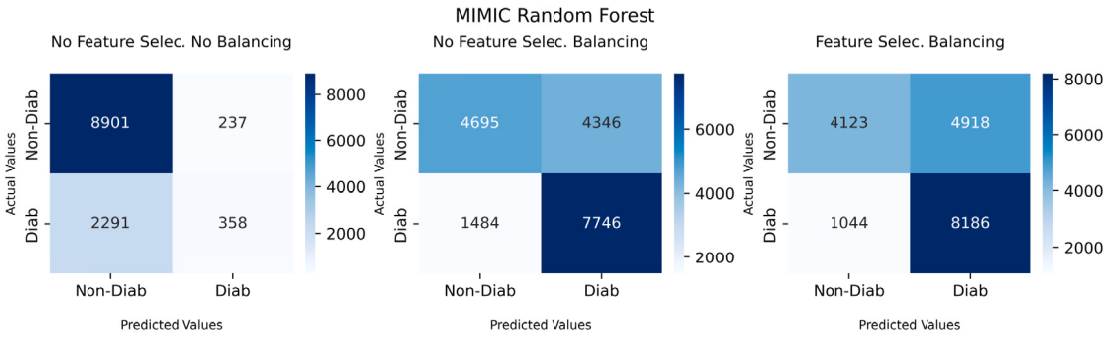
None.



(a) MIMIC III: Logistic Regression



(b) MIMIC III: Support Vector Machine



(c) MIMIC III: Random Forest

Fig. 17. Confusion matrices for MIMIC III dataset.

Acknowledgements

This research was funded by the National Water and Energy Center of the United Arab Emirates University (Grant 31R215).

References

[1] The top 10 causes of death. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. [Accessed 14 October 2022].

[2] Ismail L, Materwala H, Kaabi J Al. Association of risk factors with type 2 diabetes: a systematic review. Comput Struct Biotechnol J 2021;19:1759–85.

[3] Ismail L, Materwala H, Tayefi M, Ngo P, Karduck AP. Type 2 diabetes with artificial intelligence machine learning: methods and evaluation. Arch Comput Methods Eng 2022;29(1):313–33.

[4] Diabetes around the world in 2021. https://diabetesatlas.org/#:~:text=Diabetes%20around%20the%20world%20in%202021%3A,-%20and%20middle-income%20countries. [Accessed 14 October 2022].

[5] Diabetes - symptoms and causes - Mayo clinic. https://www.mayoclinic.org/diseases-conditions/diabetes/symptoms-causes/syc-20371444. [Accessed 3 October 2023].

[6] Diabetes complications | ada. https://diabetes.org/about-diabetes/complications. [Accessed 3 October 2023].

[7] Diabetes facts and figures. https://www.idf.org/aboutdiabetes/what-is-diabetes/facts-figures.html. [Accessed 14 October 2022].

[8] Ahmad HF, Mukhtar H, Alaqail H, Seliaman M, Alhumam A. Investigating health-related features and their impact on the prediction of diabetes using machine learning. Appl Sci 2021;11(3):1173.

[9] Kopitar L, Kocbek P, Cilar L, Sheikh A, Stiglic G. Early detection of type 2 diabetes mellitus using machine learning-based prediction models. Sci Rep 2020;10(1):1–12.

[10] Deberneh HM, Kim I. Prediction of type 2 diabetes based on machine learning algorithm. Int J Environ Res Public Health 2021;18(6):3317.

[11] Syed AH, Khan T. Machine learning-based application for predicting risk of type 2 diabetes mellitus (t2dm) in Saudi Arabia: a retrospective cross-sectional study. IEEE Access 2020;8:199539–61.

[12] Joshi RD, Dhakal CK. Predicting type 2 diabetes using logistic regression and machine learning approaches. Int J Environ Res Public Health 2021;18(14):7346.

[13] Chang V, Bailey J, Xu QA, Sun Z. Pima Indians diabetes mellitus classification based on machine learning (ml) algorithms. Neural Comput Appl 2022:1–17.

[14] Zhang L, Wang Y, Niu M, Wang C, Wang Z. Machine learning for characterizing risk of type 2 diabetes mellitus in a rural Chinese population: the Henan rural cohort study. Sci Rep 2020;10(1):1–10.

[15] Lu H, Uddin S, Hajati F, Moni MA, Khushi M. A patient network-based machine learning model for disease prediction: the case of type 2 diabetes mellitus. Appl Intell 2022;52(3):2411–22.

[16] Ismail L, Materwala H, Karduck AP, Adem A, et al. Requirements of health data management systems for biomedical care and research: scoping review. J Med Internet Res 2020;22(7):e17508.

[17] Ismail L, Materwala H, Hennebelle A. A scoping review of integrated blockchain-cloud (bcc) architecture for healthcare: applications, challenges and solutions. Sensors 2021;21(11):3753.
[18] Ismail L, Materwala H, Khan MA. Performance evaluation of a patient-centric blockchain-based healthcare records management framework. In: Proceedings of the 2020 2nd international electronics communication conference; 2020. p. 39–50.
[19] Ismail L, Materwala H, Sharaf Y. Blockhr–a blockchain-based healthcare records management framework: performance evaluation and comparison with client/server architecture. In: 2020 International symposium on networks, computers and communications (ISNCC). IEEE; 2020. p. 1–8.
[20] Ismail L, Materwala H. Blockhr: a blockchain-based framework for health records management. In: Proceedings of the 12th international conference on computer modeling and simulation; 2020. p. 164–8.
[21] Kim M-J, Lim N-K, Choi S-J, Park H-Y. Hypertension is an independent risk factor for type 2 diabetes: the Korean genome and epidemiology study. Hypertens Res 2015;38(11):783–9.
[22] Omron evolv hem-7600t-e blood pressure monitor. https://www.omron-healthcare.com/eu/blood-pressure-monitors/EVOLV.html. [Accessed 14 October 2022].
[23] Omron m3 comfort (hem-7134-e) blood pressure monitor. https://medaval.ie/device/omron-m3-comfort-hem-7134-e/. [Accessed 14 October 2022].
[24] Takahashi H. Validation of home blood pressure-monitoring devices omron evolv (hem-7600t-e), hem-9210t, and m3 comfort (hem-7134-e) according to European society of hypertension international protocol (esh-ip) revision 2010. J Clin Physiol 2019;49(1):37–46.
[25] Omron evolv hem-7600t-e blood pressure monitor price. https://www.amazon.ae/Omron-evolv-pressure-monitor-wireless/dp/B07NQSMWZP. [Accessed 14 October 2022].
[26] Omron m3 comfort (hem-7134-e) blood pressure monitor price. https://www.amazon.ae/Omron-Upper-Blood-Pressure-Monitor/dp/B07NQSQ7ZH. [Accessed 14 October 2022].
[27] Omron hem-9210t blood pressure monitor. https://www.omronhealthcare-ap.com/ap/product/143-hem-9210t--9200t. [Accessed 14 October 2022].
[28] Mobil-o-grpah blood pressure monitor. https://www.iem.de/en/produkt/mobil-o-graph-pwa/. [Accessed 14 October 2022].
[29] Wei W, Tölle M, Zidek W, van der Giet M. Validation of the mobil-o-graph: 24 h-blood pressure measurement device. Blood Press Monit 2010;15(4):225–8.
[30] Mobil-o-graph blood pressure monitor price. https://www.numed.co.uk/products/mobil-o-graph-ambulatory-blood-pressure-monitor. [Accessed 14 October 2022].
[31] Neovius M, Linne Y, Rossner S. Bmi, waist-circumference and waist-hip-ratio as diagnostic tests for fatness in adolescents. Int J Obes 2005;29(2):163–9.
[32] Bjelica D, Gardasevic J, Milosevic Z, Bozic PR, Masanovic B. Trajectories of body height, body weight, bmi, and nutrition status from 1979 to 1987: a measurement-based analysis of 8740 Montenegrin male adolescents from the municipality of Berane. Int J Environ Res Public Health 2021;18(10):5490.
[33] Martin A, Ross W, Drinkwater D, Clarys J. Prediction of body fat by skinfold caliper: assumptions and cadaver evidence. Int J Obes 1985;9:31–9.
[34] Frija-Masson J, Mullaert J, Vidal-Petiot E, Pons-Kerjean N, Flamant M, d'Ortho M-P, et al. Accuracy of smart scales on weight and body composition: observational study. JMIR mHealth uHealth 2021;9(4):e22487.
[35] Brodie D, Vicki Moscrip MSC H, Hutcheon R. Body composition measurement: a review of hydrodensitometry, anthropometry, and impedance methods. Nutrition 1998;14(3):296–310.
[36] Ginde SR, Geliebter A, Rubiano F, Silva AM, Wang J, Heshka S, et al. Air displacement plethysmography: validation in overweight and obese subjects. Obes Res 2005;13(7):1232–7.
[37] Kaul S, Rothney MP, Peters DM, Wacker WK, Davis CE, Shapiro MD, et al. Dual-energy X-ray absorptiometry for quantification of visceral fat. Obesity 2012;20(6):1313–8.
[38] Kawamoto R, Tabara Y, Kohara K, Miki T, Kusunoki T, Takayama S, et al. Relationships between lipid profiles and metabolic syndrome, insulin resistance and serum high molecular adiponectin in Japanese community-dwelling adults. Lipids Health Dis 2011;10(1):1–7.
[39] Birtcher KK, Ballantyne CM. Measurement of cholesterol: a patient perspective. Circulation 2004;110(11):e296–7.
[40] Easytouch blood glucose/cholesterol/uric acid monitoring system. https://5.imimg.com/data5/VP/VX/MY-28906720/glucometer-cholesterol-uric-acid-test-easy-touch-meter.pdf. [Accessed 14 October 2022].
[41] Benecheck blood glucose/total cholesterol/uric acid monitoring. https://www.smithbiomed.com/wp-content/uploads/2015/08/BeneCheck-Questions-and-Answers.pdf. [Accessed 14 October 2022].
[42] Golden SH, Williams JE, Ford DE, Yeh H-C, Paton Sanford C, Nieto FJ, et al. Depressive symptoms and the risk of type 2 diabetes: the atherosclerosis risk in communities study. Diabetes Care 2004;27(2):429–35.
[43] Shafer AB. Meta-analysis of the factor structures of four depression questionnaires: Beck, ces-d, hamilton, and zung. J Clin Psychol 2006;62(1):123–46.
[44] Dehghan A, Van Hoek M, Sijbrands EJ, Hofman A, Witteman JC. High serum uric acid as a novel risk factor for type 2 diabetes. Diabetes Care 2008;31(2):361–2.
[45] Uric acid test. https://medlineplus.gov/lab-tests/uric-acid-test/#:~:text=A%20uric%20acid%20test%20can,a%20test%20tube%20or%20vial. [Accessed 14 October 2022].
[46] Fabre S, Clerson P, Launay J-M, Gautier J-F, Vidal-Trecan T, Riveline J-P, et al. Accuracy of the humasensplus point-of-care uric acid meter using capillary blood obtained by fingertip puncture. Arthritis Res Ther 2018;20(1):1–9.
[47] Guo J. Uric acid monitoring with a smartphone as the electrochemical analyzer. Anal Chem 2016;88(24):11986–9.
[48] Paraskos J, Berke Z, Cook J, Miner JN, Braddock M, Platt A, et al. An analytical comparison between point-of-care uric acid testing meters. Expert Rev Mol Diagn 2016;16(3):373–82.
[49] Easytouch price. https://www.amazon.com/gp/product/B07PWYKFJC/ref=ox_sc_act_title_1?smid=A1JTEEWGBZR44P&psc=1. [Accessed 14 October 2022].
[50] Kuo C-S, Hwu C-M, Lin Y-H, Huang Y-H, Kao W-Y, Weih M-J, et al. Portable electrochemical blood uric acid meter. J Clin Lab Anal 2002;16(2):109–14.
[51] Uasure uric acid meter price. https://www.ubuy.ae/en/product/RN9U17E-uasure-uric-acid-meter-ua-sure-test-kit-gout-monitor-tester-complete-monitoring-kit. [Accessed 14 October 2022].
[52] Benecheck plus price. https://www.amazon.ae/Benecheck-Plus-Meter-Cholesterol-Glucose/dp/B07F758J7T. [Accessed 14 October 2022].
[53] Humasens 2.0 plus price. https://www.labmark.eu/humasens-2-0plus. [Accessed 14 October 2022].
[54] Kim KM, Henderson GN, Ouyang X, Frye RF, Sautin YY, Feig DI, et al. A sensitive and specific liquid chromatography–tandem mass spectrometry method for the determination of intracellular and extracellular uric acid. J Chromatogr B 2009;877(22):2032–8.
[55] Fitbit charge hr activity tracker. https://www.fitbit.com/c/chargehr/chargehr-101. [Accessed 14 October 2022].
[56] Robbins R, Affouf M, Seixas A, Beaugris L, Avirappattu G, Jean-Louis G, et al. Four-year trends in sleep duration and quality: a longitudinal study using data from a commercially available sleep tracker. J Med Internet Res 2020;22(2):e14735.
[57] Kang S-G, Kang JM, Ko K-P, Park S-C, Mariani S, Weng J. Validity of a commercial wearable sleep tracker in adult insomnia disorder patients and good sleepers. J Psychosom Res 2017;97:38–44.
[58] Polysomnography test price. https://www.mdsave.com/procedures/sleep-study-polysomnography/d782f4c8. [Accessed 14 October 2022].
[59] Ōura ring sleep tracker. https://ouraring.com/. [Accessed 14 October 2022].
[60] de Zambotti M, Rosas L, Colrain IM, Baker FC. The sleep of the ring: comparison of the ōura sleep tracker against polysomnography. Behav Sleep Med 2019;17(2):124–36.
[61] Fitbit flex fitness band. https://www.fitbit.com/pl/shop/flex. [Accessed 14 October 2022].
[62] Kaewkannate K, Kim S. A comparison of wearable fitness devices. BMC Public Health 2016;16(1):1–16.
[63] Lee H-A, Lee H-J, Moon J-H, Lee T, Kim M-G, In H, et al. Comparison of wearable activity tracker with actigraphy for sleep evaluation and circadian rest-activity rhythm measurement in healthy young adults. Psychiatry Invest 2017;14(2):179.
[64] Fitbit charge hr activity tracker price. https://www.walmart.com/ip/Fitbit-Charge-HR-Heart-Rate-Activity-Wristband/39877708. [Accessed 14 October 2022].
[65] Polar a370 fitness tracker. https://www.polar.com/ae-en/products/sport/A370-fitness-tracker. [Accessed 14 October 2022].
[66] Pesonen A-K, Kuula L. The validity of a new consumer-targeted wrist device in sleep measurement: an overnight comparison against polysomnography in children and adolescents. J Clin Sleep Med 2018;14(4):585–91.
[67] Actiwatch 2. https://www.usa.philips.com/healthcare/product/HC1044809/actiwatch-2-activity-monitor. [Accessed 14 October 2022].
[68] Fitbit alta hr activity tracker. https://www.fitbit.com/gb/shop/altahr. [Accessed 14 October 2022].
[69] Cook JD, Eftekari SC, Dallmann E, Sippy M, Plante DT. Ability of the fitbit alta hr to quantify and classify sleep in patients with suspected central disorders of hypersomnolence: a comparison against polysomnography. J Sleep Res 2019;28(4):e12789.
[70] Fitbit alta hr activity tracker price. https://www.amazon.com/Fitbit-Alta-Black-Small-Version/dp/B06W58QKH6. [Accessed 14 October 2022].
[71] Withings pulse hr. https://www.withings.com/us/en/pulse-hr. [Accessed 14 October 2022].
[72] Jawbone up24. https://www.wareable.com/fitness-trackers/remembering-the-jawbone-up24-7320. [Accessed 14 October 2022].
[73] Emfit quantified sleep tracker. https://emfit.com/. [Accessed 14 October 2022].
[74] Kholghi M, Szollosi I, Hollamby M, Bradford D, Zhang Q. A validation study of a ballistocardiograph sleep tracker emfit qs against polysomnography. J Clin Sleep Med 2021.
[75] Sportline 340 strider pedometer. https://www.manualsdir.com/manuals/711593/sportline-340-step.html. [Accessed 14 October 2022].
[76] Fitbit one activity tracker. https://www.fitbit.com/pl/shop/one. [Accessed 14 October 2022].
[77] Battenberg AK, Donohoe S, Robertson N, Schmalzried TP. The accuracy of personal activity monitoring devices. Seminars in arthroplasty, vol. 28. Elsevier; 2017. p. 71–5.
[78] Fitbit one activity tracker price. https://www.walmart.com/ip/Fitbit-One-Wireless-Act-Tracker/24726546. [Accessed 14 October 2022].
[79] Omron hj-321 pedometer. https://omronhealthcare.com/wp-content/uploads/HJ-321-IM-WEB-03212012.pdf. [Accessed 14 October 2022].

[80] Omron hj-321 pedometer price. https://www.amazon.ae/Omron-HJ-321-Pedometer/dp/B007ZWIJR2. [Accessed 14 October 2022].
[81] Sportline 340 strider pedometer price. https://www.whizz.ae/product/547908/sportline-340-strider-pedometer/. [Accessed 14 October 2022].
[82] Fitbit force activity tracker. https://www.manualsdir.com/manuals/601824/fitbit-force.html. [Accessed 14 October 2022].
[83] Stepwatch activity monitor. https://www.cymatech.com/. [Accessed 14 October 2022].
[84] iphone 5. https://support.apple.com/kb/sp655?locale=en_US. [Accessed 14 October 2022].
[85] Samsung galaxy s4. https://en.wikipedia.org/wiki/Samsung_Galaxy_S4. [Accessed 14 October 2022].
[86] Samsung galaxy s4 price. https://www.amazon.ae/Samsung-Galaxy-Black-Mist-Unlocked/dp/B00HXRE7PW. [Accessed 14 October 2022].
[87] Yu W, Liu T, Valdez R, Gwinn M, Khoury MJ. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. BMC Med Inform Decis Mak 2010;10(1):1–7.
[88] Rachim VP, Chung W-Y. Wearable-band type visible-near infrared optical biosensor for non-invasive blood glucose monitoring. Sens Actuators B, Chem 2019;286:173–80.
[89] Omer AE, Shaker G, Safavi-Naeini S, Alquié G, Deshours F, Kokabi H, et al. Non-invasive real-time monitoring of glucose level using novel microwave biosensor based on triple-pole csrr. IEEE Trans Biomed Circuits Syst 2020;14(6):1407–20.
[90] Jerez JM, Molina I, García-Laencina PJ, Alba E, Ribelles N, Martín M, et al. Missing data imputation using statistical and machine learning methods in a real breast cancer problem. Artif Intell Med 2010;50(2):105–15.
[91] Larose DT. Data mining methods & models. John Wiley & Sons; 2006.
[92] Chandrashekar G, Sahin F. A survey on feature selection methods. Comput Electr Eng 2014;40(1):16–28.
[93] Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn 2002;46(1):389–422.
[94] Fushiki T. Estimation of prediction error by using k-fold cross-validation. Stat Comput 2011;21(2):137–46.
[95] Cutler A, Cutler DR, Stevens JR. Random forests. In: Ensemble machine learning. Springer; 2012. p. 157–75.
[96] Ismail L, Materwala H. Idmpf: intelligent diabetes mellitus prediction framework using machine learning. Appl Comput Inform 2021.
[97] Cichosz SL, Stausholm MN, Kronborg T, Vestergaard P, Hejlesen O. How to use blockchain for diabetes health care data and access management: an operational concept. J Diabetes Sci Technol 2019;13(2):248–53.
[98] Ismail L, Materwala H. A review of blockchain architecture and consensus protocols: use cases, challenges, and solutions. Symmetry 2019;11(10):1198.
[99] Zheng Z, Xie S, Dai H-N, Chen X, Wang H. Blockchain challenges and opportunities: a survey. Int J Web Grid Serv 2018;14(4):352–75.
[100] Smith JW, Everhart JE, Dickson W, Knowler WC, Johannes RS. Using the adap learning algorithm to forecast the onset of diabetes mellitus. In: Proceedings of the annual symposium on computer application in medical care. American Medical Informatics Association; 1988. p. 261.
[101] Johnson AE, Pollard TJ, Shen L, Lehman L-wH, Feng M, Ghassemi M, et al. Mimic-iii, a freely accessible critical care database. Sci Data 2016;3(1):1–9.
[102] Islam M, Ferdousi R, Rahman S, Bushra HY. Likelihood prediction of diabetes at early stage using data mining techniques. In: Computer vision and machine intelligence in medical image analysis. Springer; 2020. p. 113–25.
[103] Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression, vol. 398. John Wiley & Sons; 2013.
[104] Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intell Syst Appl 1998;13(4):18–28.
[105] Baak M, Koopman R, Snoek H, Klous S. A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics. Comput Stat Data Anal 2020;152:107043.
[106] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. Smote: synthetic minority oversampling technique. J Artif Intell Res 2002;16:321–57.
[107] Boden G, Laakso M. Lipids and glucose in type 2 diabetes: what is the cause and effect? Diabetes Care 2004;27(9):2253–9.
[108] Gerstein HC, Miller ME, Byington RP, Goff DC, Bigger JT, Buse JB, et al. Effects of intensive glucose lowering in type 2 diabetes - the action to control cardiovascular risk in diabetes study group. N Engl J Med 2008;358(24):2545–59.
[109] F.D. Association. Type 2 diabetes risk assessment form. https://www.diabetes.fi/files/502/eRiskitestilomake.pdf. [Accessed 8 November 2022].
[110] Kumar A, Bharti SK, Kumar A. Type 2 diabetes mellitus: the concerned complications and target organs. Apollo Medicine 2014;11(3):161–6.
[111] Clinical protocol for the prevention and treatment of diabetes in adults. https://championprovider.ucsf.edu/sites/champion.ucsf.edu/files/images/DIABETES%20PROTOCOL.pdf. [Accessed 8 November 2022].
[112] Cichosz SL, Johansen MD, Ejskjaer N, Hansen TK, Hejlesen OK. Improved diabetes screening using an extended predictive feature search. Diabetes Technol Ther 2014;16(3):166–71.
[113] Shai I, Jiang R, Manson JE, Stampfer MJ, Willett WC, Colditz GA, et al. Ethnicity, obesity, and risk of type 2 diabetes in women: a 20-year follow-up study. Diabetes Care 2006;29(7):1585–90.
[114] Jackson CL, Redline S, Kawachi I, Hu FB. Association between sleep duration and diabetes in black and white adults. Diabetes Care 2013;36(11):3557–65.
[115] Zizi F, Pandey A, Murrray-Bachmann R, Vincent M, McFarlane S, Ogedegbe G, et al. Race/ethnicity, sleep duration, and diabetes mellitus: analysis of the national health interview survey. Am J Med 2012;125(2):162–7.