ITC 4/51
Information Technology and Control
Vol. 51 / No. 4 / 2022
pp. 678-691
DOI 10.5755/j01.itc.51.4.31641

Prediction of Arterial Stiffness Risk in Diabetes Patients through Deep Learning Techniques

Received 2022/06/17 | Accepted after revision 2022/07/30

http://dx.doi.org/10.5755/j01.itc.51.4.31641

HOW TO CITE: Priya, A. M., Thilagamani, S. (2022). Prediction of Arterial Stiffness Risk in Diabetes Patients through Deep Learning Techniques. *Information Technology and Control*, 51(4), 678-691. http://dx.doi.org/10.5755/j01.itc.51.4.31641

Prediction of Arterial Stiffness Risk in Diabetes Patients through Deep Learning Techniques

A. Mohana Priya, S. Thilagamani

Department of Computer Science and Engineering, M. Kumarasamy College of Engineering, Karur, 639113, India; e-mail: mohanapriyaa21@outlook.com

Corresponding author: mohanapriyaa21@outlook.com

Diabetes and arterial stiffness are the primary health concerns related to each other. The understanding of both factors provides efficient disease prevention and avoidance. For the development of cardiovascular disease, arterial stiffness and Diabetes are pathological process considerations. The existing researchers reported the association between these two factors and the complications of arterial stiffness with Diabetes are still in research. Arterial stiffness is measured through pulse wave velocity (PWV), which influences cardiovascular disease in diabetic patients. Moreover, this study developed a medical prediction model for arterial stiffness through the machine and deep learning models to predict the patients who are high-risk factors. Brachial–ankle pulse wave velocity (baPWV) and fasting blood glucose (FBG) are the consideration of baseline. Gaussian-Least absolute shrinkage and selection operator (LASSO) with whale optimization is proposed for feature selection. Initially, key features are extracted from the wave measurement using LASSO, and Principal component analysis (PCA) has been used to remove the outliers. Second, Gaussian regression chooses the PWV-based relevant features from the LASSO-identified features. The parts are the critical points to increasing the accuracy of the prediction model. Hence, the selected features are further improved with an evolutionary algorithm called the cat optimization approach. Third, the prediction model is constructed using three machine and deep learning algorithms such as a Support vector machine (SVM), a convolution neural network (CNN), and Gated Recurrent Unit (GRU). The performance of these methods is compared through the area under the receiver operating characteristic curve metric in the dataset. The model with the best performance was selected and validated in an independent discovery dataset (n = 912) from the Dryad Digital Repository. From the experimental evaluation, LSTM performs better than other algorithms in classifying arterial stiffness with an AUROC of 0.985 and AUPRC of 0.976.

KEYWORDS: Diabetes, Arterial stiffness, machine, and deep learning, LASSO, cardiovascular, SVM, CNN, LSTM, AE.

1. Introduction

The occurrences of diabetes have been increasing day by day worldwide. Diabetes is the major risk factor for cardiovascular disease and renal dysfunction, and while diabetes is associated with Arterial Stiffness (AS), the risk of cardiovascular disease increases. As is the risk factor for Cardiovascular disease, and Fast blood glucose (FBG) concentration is the major risk factor for AS. The pulse wave velocity of diabetes and prediabetes patients is higher than that of patients with normal FBG [4, 8, 15]. The literature research found that AS is the risk factor for diabetes and the association between AS and diabetes [28]. The relationship and association between AS and diabetes are yet to be precise. FBG and AS assessed from brachial-ankle PWV (baPWV) were measured in 2010. This will increase our research focus on examining the relationship between AS and diabetes based on baPWV to find the risk factor for diabetes patients.

Arterial stiffness is the basis for cardiovascular diseases such as stroke, hypertension, and atherosclerosis and is associated with aging, the cardiovascular continuum [34]. Due to the loss of arterial elasticity and its conditions, AS increases the risk, including diabetes mellitus [6]. This motivates our research to identify the deep learning strategy for analyzing diabetes data that can acquire arterial risk. Clinically, baPWV is the measure of the AS, and the persons with ba PWV>1400 cm/s have vascular chance [24] with increased Framingham risk value and increased hypertension. The research focuses on reducing cardiovascular risk and the number of patients with AS risk. The need for a convenient tool with advanced technologies to predict the AS risk in day-to-day clinical practice to prevent the risks is highlighted. It is essential to diagnose and monitor AS with simple clinical data. SAGE with logistic regression is developed to predict AS [32]. Moreover, logistic regression-based methods failed to consider the association between the variables captured with various algorithms to increase the risk prediction accuracy. In recent research, Machine learning-based approaches were applied to developing clinical methods to diagnose diseases [23, 5, 33]. Compared to the logistic regression methods, machine and deep learning-based procedures recognize the non-linear association between the complex data and diagnose clinical results better [20]. The contribution of the proposed work is as follows:

- The biological characteristics of diabetes patients are gathered based on baPWV to identify Arterial Stiffness (AS) risk factors.
- The key characteristics (features) are selected using LASSO regression, and the outliers are removed using PCA.
- Among the key features, most optimized features are selected using Gaussian regression optimized by Cat Optimization Algorithm (COA).
- The proposed Gaussian-Lasso- COA selects the most relevant characteristics of the patients, which will increase the risk factor prediction.
- Four machine and deep learning algorithms such as SVM, CNN, LSTM, and GRU have been analyzed, and the best model is selected to predict risk factors as low, medium, and high-risk factors based on baPWV and its characteristics.
- This proposed feature selection with a deep learning model is evaluated using AUROC, AUPRC, and RMSE.
- The evaluated model can identify the arterial stiffness risk factors of diabetes patients that cause cardiovascular disease and is helpful to the clinical study for early treatment.

The remaining section of this paper is as follows: Section 2 discusses the related work. Section 3 introduced the materials and methods proposed for the prediction of risk factors. Section 4 discussed the results and evaluates the proposed model. Section 5 concludes the proposed model with merits, demerits, and future directions.

2. Review of Literature

This section discusses the related works of association between arterial stiffness and diabetes risk factors. Li et al., [14] developed a prediction model to recognize the elevated AS in diabetes patients using machine learning algorithms. They used LASSO and SVM with Recursive feature elimination for feature selection and used gradient boosting for classification. by comparing the various machine learning ap-

proaches, they conclude that gradient boosting outperforms detecting the non-elevated AS and elevated AS. Zheng et al., [34] studied the temporal association between AS and FBG status based on baPWV. The study experimented with using 14159 participants from the kailan survey from 2010 to 2015. Among the participants, path analysis was used to find the relationship between baPWV and FBG by repeating the assessment of baPWV and FBG in 2014 and 2017. Yaxin et al. [32] predicted the risk factor of diabetic retinopathy from type 2 diabetes mellitus to increase AS. They used baPWV as the indicator for AS. They investigated the relationship between baPWV and the severity of diabetic retinopathy based on 2473 Chinese patients as a cohort study. Multinomial logistic regressions have been used to analyze the risk factor baPWV. They conclude that baPWV is an independent predictor for finding the increasing AS involved in diabetic retinopathy development.

Jin et al., [9] estimated the PWV from radial pressure waves with the help of machine learning approaches. They used two ML approaches as a pipeline to measure the carotid-femoral PWV (cfPWV) from the pulse wave measured from the radial pressure wave. The first approach is Gaussian process regression, which estimates the cfPWV from the extracted features using pulse wave analysis. The second approach is a recurrent neural network used to estimate the cfPWV from the whole radial pressure wave. The used study population consists of 4374 subjects from the Twin UK cohort aged 25 to 75 years. They conclude that the probability of vascular aging assessment using a single pulse wave than cfPWV is higher, leading to cardiovascular disease risk. Usman et al., [26] predict the abnormalities of diabetes patients by monitoring the arterial stiffness using logistic regression. They used three characteristics of the patients: age, HbA1C, and AUC-PPG. The correlation between these three features is analyzed to find the risk of diabetes patients with increased arterial stiffness.

Lim et al., [13] studied the increased arterial stiffness leading to renal diseases in Type 2 diabetes mellitus using baPWV. Based on the EMPA-REG-Outcome trial, the increased risk of arterial stiffness is defined as baPWV>=1800 cm/s. They analyzed that the increasing macroalbuminuria is higher in baPWV>=1800 cm/sf, with causes an increased risk of renal disease. They conclude that patients with baPWV>=1800 cm/s are a higher risk factor for renal disease progression in type 2 diabetes patients. Oliveira Alvim et al., [22] studied the impact of AS in diabetes patients using an urban Brazilian study population. This study used 1415 individuals randomly selected from the Brazilian community. Using the non-invasive automatic device, PWV was measured, and the increasing AS is defined as PWV>=12 m/s. They conclude that increasing AS in diabetes patients in the Brazilian population increases cardiovascular risk in diabetes patients.

Elias et al., [18] analyzed the association between type 2 diabetes mellitus and AS using Maine-Syracuse Study based on cfPWV.

They employed around 508 participants from the Maine Syracuse study and compared PWV between type 2 diabetes at six and non-diabetes at wave 6 for finding the risk factors. They conclude that type 2 diabetes mellitus has increased PWV, leading to cardiovascular risk factors. Prenner and Chirinos [21] reviewed the relationship between AS in diabetes mellitus. Growing AS is the predictor of cardiovascular events both in diabetes and non-diabetes. They conclude that maximum research demonstrated the association between AS and DM, which is cross-sectional. Cecilia et al. [16] developed a prediction model to find the relationship between blood pressure and Aortic stiffness using the characteristics such as heritability, blood pressure polymorphisms, and randomization of mendelian. They used LASSO and generalized equation regression for the analysis to find the association between BP-linked SNP. The conclusion states that the genetic BP basis is regulated by BP and influenced by AS, and there is a bi-directional association between BP and AS. The existing approaches still lack in finding the prediction of risk factors of diabetes patients with improved accuracy and reduced error. improved accuracy.

In the existing models, the data analysis is done with good machine learning algorithms. The accuracy of the results is good. To acquire the best result, a deep learning model is implemented in this article.

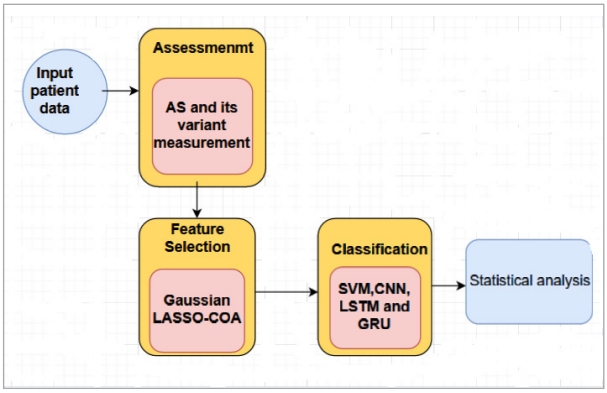
3. Materials and Methods

This section discusses the system overview and the methods used to develop the arterial stiffness prediction of diabetes patients. The proposed model struc-

ture is shown in Figure 1, which consists of four stages of processing. First, the input data is assessed, and the factors such as PWV and fast glucose level with its variants are measured. Second, the key features are extracted using LASSO from the measured data. Then the outliers are removed through PCA for better prediction. The relevant parts from the LASSO are selected through Gaussian regression and optimized with cat optimization to reduce the network loss. Third, machine and deep learning algorithms such as SVM, CNN, LSTM, and GRU are used for prediction. Fourth, the model is statistically analyzed to know the impact of arterial stiffness on diabetes patients' risk factors.

Figure 1

Overview of proposed Arterial Stiffness prediction



3.1. Data for the Study

This paper uses a discovery dataset to predict and analyze arterial stiffness and its impact on diabetes patients. This dataset consists of 760 patient records recruited from Fujian Medical University Hospital from 2017 to 2019. The criteria for their consideration are as follows: patients with diabetes mellitus older than 18 years and were visited the hospital for the first time for their baPWV test. They are not considered as follows: patients with ABI less than 0.9, patients with highly affected arrhythmia, renal, pulmonary, heart valve, autophagy, rheumatic disease, and myocarditis; the usage of antibiotics in the last three months. The validation dataset is taken from the current study from Dryad digital repository [2], which can be used to analyze the prediction model. The cohort details are given in Fukuda et al., [2]

3.2. Arterial Stiffness Assessment and Measures of Effective Factors

The artery tester BP203RPE-II has been used to measure the baPWV, blood pressure, ABI (ankle-brachial index), and Body Mass Index (BMI). The patients are divided into three groups: baPWV <1400 cm/s for low risk of AS, 1800 <baPWV>1400 for medium risk of AS, and baPWV ≥1800 cm/s for high risk of AS.

- **BMI:** Body Mass Index is computed as the weight divided by the square of height (kg/m2). This measurement also includes the serum glucose level, HbA1C (glycated hemoglobin total protein), cholesterol, high and low-density lipoprotein cholesterol (Hdl and Ldl), and alanine aminotransferase (Alt), blood urea nitrogen (bun), aspartate aminotransferase (Ast) and creatinine. The data on the duration of diabetes, comorbidities such as stroke, hypertension, coronary artery disease, and renin-angiotensin inhibitors were also collected. Different blood samples were collected from the antecubital vein after overnight fasting.
- **baPWV measurement:** baPWV, Blood pressure and ABI were measured through a non-invasive vascular device called BP-203RPEIII, which was measured by a trained person who used pneumatic cuffs placed on each ankle and patient each upper arm. We measured and grouped the patients according to their baPWV values into three such as the person with baPWV<1400 cm/s has normal arterial stiffness, a person with baPWV as 1800<baPWV>1400 cm/s has borderline arterial stiffness and the person with baPWV≥1800 cm/s is considered as the threshold of having a high risk of arterial stiffness which leads to cardiovascular disease and heart failure-related issues [23,24].

3.2.1. FBG (Fasting Blood Glucose) Measurement

The blood samples of 5 mL during the fasting hours of 8-10 hours were taken from the empty stomach patient elbow vein. This sample is injected to the vacuum tube having EDTA. It is centrifuged in room temperature for four hours. FBG is measured through hexokinase / glucose 6 phosphate dehydrogenase approaches with the maximum limit detection of 30.07 mmol/L [8]. FBG≥ 7 mmol/L is defined as diabetes and pre-diabetes is defined in the FBG value of 5.6>FBG<6.9 mmol>L were not used antibiotic medication [27].

3.2.2. Another Covariant Measurement

With the use of blood samples, the factors such as uric acid, and high sensitivity C reactive protein (hs-CRP) were measured through an auto analyzer. From creatinine, the estimated glomerular filtration rate (eGFR) is computed. Chronic kidney disease is declared with positive proteinuria or eGFR<60 mL/(min.m2) [12]. Blood pressure is measured during times from 7 and 9. Before the measure, the drinking habit of tea or coffee and smoking is allowed within 30 minutes. Pulse pressure is also the marker of arterial stiffness, measured with the difference between systolic blood pressure (SBP) and diastolic blood pressure (DBP). SBP reading is taken as the first time of the Hirschmann sound, and DBP reading is taken as the fifth time of the kischmann sound. These measurements are repeated three times, and the mean value is taken. The date of birth, smoking and alcohol consumption, sex, and previous medical status with its history is taken from the self-reported questionnaire [30]. A smoker is defined as a person a smokes abbot of one cigarette per day. A drinking habit is declared based on the average alcohol consumption of previous years. [7]

3.3. Feature Selection Using Proposed Gaussian-LASSO Regression with Cat Optimization Approach

This section extracts the features from the measured data to improve the accuracy of the prediction model. LASSO regression is used to select the candidate characteristics, and PCA removes the outliers. Gaussian regression has been used to extract the relevant parts from LASSO regression. Finally, an optimization algorithm enhances the feature extraction process by reducing the loss.

3.3.1. LASSO Regression

It is an compression algorithm which is used to select the candidate variable set using the penalty parameter. It uses the cost function as denoted in Equation (1) for estimating the candidate variable

$$l = \frac{1}{2n}\sum_{i=1}^{n}(Y_{real}^{i} - Y_{predication}^{i})^2 + \alpha \sum_{j=1}^{n}\left|f_j\right|, \quad (1)$$

where, f_j – coefficient of jth feature, α – hyper parameter that tunes the penalty term. Higher the coefficient value, higher the cost function. The cost function is optimized to reduce the coefficient absolute values. LASSO regression automatically chosen the relevant features and removing the features coefficient value is 0. It is executed using the scikit learn library from python. The hyperparameter of the LASSO is selected from 5 fold cross validation using the library called GridSearch CV.

3.3.2. PCA

It is performed on the features extracted from LASSO using the scikit learn library PCA function. From the origin, the distance of the data points are considered as outliers and those points are removed from classification.The high dimensional feature space are mapped into the kernel function to transform the data into low dimensional feature space. For distribution of n data points which $x_i \in R^d$, kernel matrix K is mapped to high dimensional feature space F as per the following steps:

Step 1: Dot product of the kernel matrix is determined using the kernel function as denoted in Equation

$$K_{ij} = k(x_i, x_j) \quad (2)$$

Step 2: from the resultant matrix K, eigenvectors are calculated with the normalization function such as

$$\gamma k(\alpha k \alpha k) = 1 \quad (3)$$

Step 3: test point projection to eigen vectors v_k with kernel function is calculated as

$$kPCA\big(k(x)\big) = \sum_{i=1}^{m} \alpha k_i . k(x_i, x) \quad (4)$$

3.3.3. Gaussian LASSO Regression

It is used to estimate the PWV key features from the outcome of LASSO regression. The primary vital factors of using Gaussian regression are (i) this can provides the uncertainty estimation, which is not suitable for other regression methods (ii) with the maximization of the log-likelihood. This model hyperparameters can be identified. This consumes less time than cross-validation. It can be executed through the Gaussian Process Regressor library from the scikit learns package in python. The kernels functions such as radial basis, Matern kernel, and the rational quadratic kernel were also selected from the library. For this analysis, the rational quadratic kernel function has been chosen.

3.3.4. Cat Optimization Algorithm

The subset of features selected from Gaussian Lasso regression is forward as input to COA to select the optimal subset of features. This optimal feature subset will increase the classification accuracy as higher. The cat swarm optimization approach [10] works based on cat behavior called the seeking and tracking method. With the use of fitness and flag values, this method examines the present feature in the search space in different directions. The cat rest time is inspected in seeking mode that consists of various parameters including seeking rank and memory pool, dimension, dimension change, and self-position [19]. While analyzing the feature, the memory pool is engaged initially and each feature position is examined. According to the change of position, each feature is assigned with seeking rank. This rank is changed based on position and the cat is moved in the search space by updating the bias and weights to reduce the error. The updating process is performed as in Equation (5)

$$V_{i,c} = V_{i,c} + r_1 ac_1 (X_{best,c} - X_{i,c}) \quad (5)$$

Where, $V_{i,c}$ – ith position of new cat velocity, r_1 – rank, ac_1 – acceleration coefficient, $X_{best,c}$ – best fitness value of ith cat, and $X_{i,c}$ –cat position. This process is repeated until the optimal subset features seeking and tracking in the feature space is found. This will reduce the network error on classification process. The proposed optimal feature extraction algorithm is as follows:

Algorithm 1: Proposed Optimal feature Extraction using Gaussian Lasso regression with COA

Step1: From the initial feature set, effective features are extracted using LASSO

Step 2: Outliers are removed using PCA

Step 3: relevant features from LASSO is selected using Gaussian regression

Step4: the optimal set of best attributes are selected based on the swarm attractiveness and brightness depending on movement computed as in Equation (6)

$$X_p(t+1) = X_p(t) + \frac{\beta_0}{1+\gamma r^n}\left[X_q(t) - X_p(t)\right] + R(\alpha) \quad (6)$$

Step 5: the output of each feature is computed with the objective function as in Equation (7)

$$f(x) = \frac{L}{1+e^{-i(x-x_0)}} \quad (7)$$

Step 6: the weight and bias of the optimization process is updated using Equation (5).

Step 7: repeat until seeking and tracking of feature space is found

Step 8: return X as the optimal subset of feature

3.4. Classification Using ML and DL Approaches

The prediction of arterial stiffness accuracy is essential for the decision of clinical treatment, which will reduce the treatments predicted from false-positive assessments. Thus, this paper aimed to obtain the improved accuracy model using machine and deep learning approaches such as support vector machine (SVM), convolution neural network (CNN), and Gated Recurrent Unit (GRU) was used. The performance of these models is compared to find the best model.

3.4.1. SVM

The arterial stiffness prediction using SVM with Gaussian kernel method helps to produce better performance. SVM with kernel function is producing the better outcome and treated as best classifier. It divides the hyperplane into two components which will transform the input into its high dimensional using Equation (8). The input samples are assigned to the feature space and the reduced samples has been used as the input to the classification problem.

$$f(X,Y) = \sum_{i=1}^{n} y_i l_i(x) + k, \quad (8)$$

where, k-bias, $f(X, Y)$ = linear model of the input X and Y, $l_i(x)$ = non-linear transformation, i=1...n using the kernel function, the classification problem is controlled and this paper used Gaussian radial basis function as stated in the Equation (9)

$$k(X,Y) = \sum_{i=1}^{n} e^{-\left(\frac{\|X-Y\|^2}{2\sigma^2}\right)}. \quad (9)$$

Then the output of the prediction is either class 1 or class 0.

$$= \begin{cases} 1 & if\ \|\bar{X}\| \le 1 \\ 0 & otherwise \end{cases}. \quad (10)$$

The hyperplane divides the two classes and the margin divides the hyperplane. The predicted result which is closest to the best margin will fall into any one of the classes. These classes predict the impact of arterial stiffness risk factors for diabetes patients.

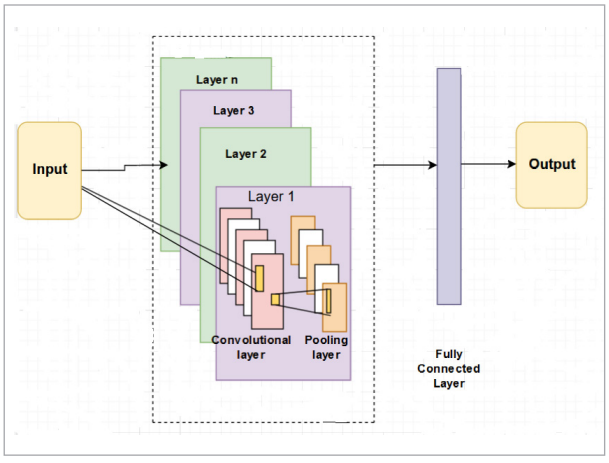
3.4.2. CNN

Convolution neural network is the widely used deep learning based classification model which consists of convolution layer to extract the features and fully connected layer for deciding the output. CNN has been implemented for image recognition as maximum [3]. The convolution layer can extract the unique features and the size of the features is reduced with pooling layer. The processing of CNN is denoted in Equation (11)

$$I' = \frac{I-K+2D}{C} + 1, \tag{11}$$

where I refers to the size of the input, K refers the kernel size, D filled with zero for the both end dimension and C denotes the kernel stride in the convolution layer. The performance of the convolution layer is not proportional to the number of convolution layers [28]. Multiple convolution layer may learn complex features effectively. The correlation between convolution layer number and its efficiency is based on the input and the optimal output is computed by learning. The number of convolution layers, kernel size, weights used for hidden layer is the parameters of CNN. The model is implemented using python programming. Figure 2 illustrates the CNN structure.

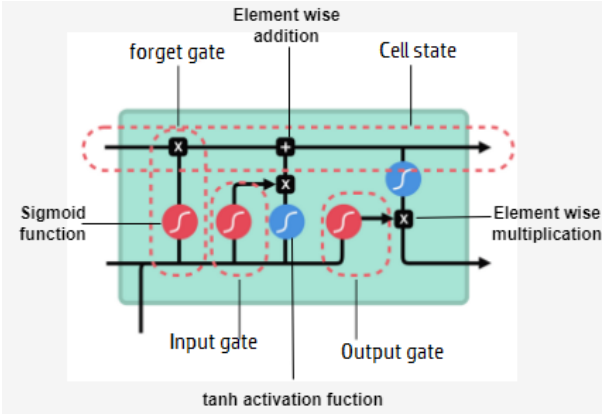
Figure 2
CNN model structure



3.4.3. LSTM

LSTM is a kind of recurrent neural network (RNN) form. The issue behind RNN is while handling the long term dependencies, it affects the gradient. The gradient can be fixed through gradient clipping method, but it still limits the gradient with threshold. To handle this gradient, gates have been suggested which is implemented in LSTM which use three gates such as input, output and forget gate in addition to RNN and also it consists of memory for storage as in Figure 3.

Figure 3
LSTM model



The LSTM inputs are as follows:

X_t – Input at time t

H_t – Current hidden state or output

H_{t-1} – Previous hidden state

$\tilde{C}_{t-1}$ – Cell sate activation vector

C_t – Current memory state

C_{t-1} – previous memory state

w_f, w_c, w_i, w_o– weights of each forget, cell, input and output gates current state

u_f, u_c, u_i, u_o– weights of each forget, cell, input and output gates previous hidden state

b_f, b_c, b_i, b_o– each gate bias vectors

σ_s – activation function (sigmoid)

σ_{ht} – activation function (hyperbolic activation function)

$\odot$- Hadamard product (element wise multiplication)

The expression of each gate is denoted in Equation (12) to (17)

$$f_t = \sigma_s(w_f X_t + u_f H_{t-1} + b_f) \quad (12)$$

$$i_t = \sigma_s(w_i X_t + u_i H_{t-1} + b_i) \quad (13)$$

$$\tilde{C}_t = \sigma_{ht}(w_c X_t + u_c H_{t-1} + b_c) \quad (14)$$

$$C_t = f_t \odot C_{t-1} + i_t * \tilde{C}_t) \quad (15)$$

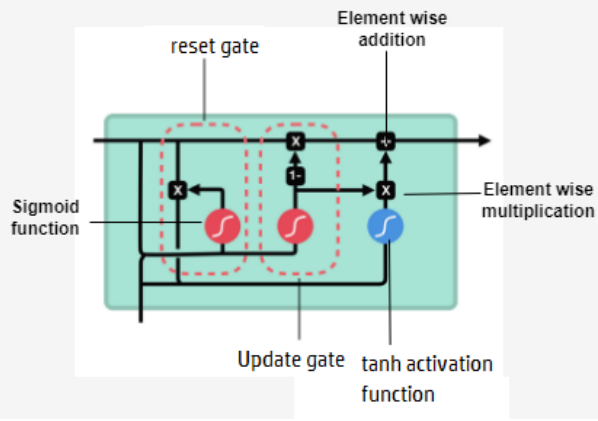
$$o_t = \sigma_s(w_O X_t + u_o H_{t-1} + b_O) \quad (16)$$

$$H_t = o_t \odot \tanh(C_t) \quad (17)$$

3.4.4. GRU

GRU is an improved version of RNN which consists of two gates such as update gate and reset gate to finalize the output. The update gate (ug) is responsible to decide how much of data are from previous step will pass to next step at time. The reset gate (rg) is responsible to decide how much of data from previous step to forget at time as shown in Figure 4. The activation vector $\tilde{H}_t$ compute the current memory state using the current input and previous hidden state from reset gate.

Figure 4
GRU model



The update and reset gate are denoted in Equation (18 to 21)

$$ug_t = \sigma_s(w_{ug} X_t + u_{ug} H_{t-1} + b_{ug}) \quad (18)$$

$$rg_t = \sigma_s(w_{rg} X_t + u_{rg} H_{t-1} + b_{rg}) \quad (19)$$

$$\hat{H}_t = \sigma_{ht}(w_H X_t + u_H(rg_t \odot H_{t-1}) + b_H) \quad (20)$$

$$H_t = (1 - ug_t) \odot H_{t-1} + (1 - ug_t) \odot \hat{H}_t), \quad (21)$$

where, the variables are same as in LSTM except $\tilde{H}_t$ – candidate activation vector. The optimal hyper parameters are obtained using the AUROC based on 10 fold cross validation in discovery dataset.

3.5. Statistical Analysis

The discovery data has been split into two groups: training and testing data 70% and 30%, respectively. The machine and deep learning algorithms were executed on training data based on the tuning hyperparameters. AUROC and AUPRC (Area under the precision-recall curve) have been computed on testing data for all the classification algorithms. Based on the comparison of AUROC and AUPRC best model will be found. Numerical data in the dataset is expressed in terms of mean ± standard deviation. The categorical data in the discovery dataset are expressed in terms of percentages and numbers. The continuous variables are compared with two groups using a t-test. Categorical data are compared using the chi-square test. All the evaluations are performed using python, scikit learn package. The features are normalized using the Standard Scaler python library. Normality of the data are assessed using K-s test (Kolmogorov-smirnov). To evaluate the machine learning algorithms, Root Mean Square Error (RMSE) is used as denoted in Equation (22)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(baP\acute{W}V_i - baPWV_i)^2}{n}}, \quad (22)$$

where, n-test dataset size, $baP\acute{W}V_i$ – ith estimated baPWV, $baPWV_i$ – ith measured baPWV. Error percentage based on RMSE is computed as in Equation (23).

$$\varepsilon = \frac{RMSE}{baPWV} \times 100\ \%, \quad (23)$$

where $\overline{baPWV}$ – mean value of the study population baPWV

4. Result and Discussions

In this paper, 760 subjects from the discovery dataset is used for evaluation with the age of 56 ± 12 years in a male and female ratio of 60:40. Based on the baPWV value of 1800 cm/s, the samples are divided into three

Table 1

Baseline characteristics of the patients based on arterial stiffness status

	Low risk of arterial stiffness (baPWV ≤1400)	Medium risk of arterial stiffness (1400≥ baPWV ≤1800)	High risk of arterial stiffness (baPWV ≥1800)
n	230	160	370
Male	153	126	171
Female	36	28	180
Age (years)	49 ± 14	55 ± 8	61± 9
Height (cm)	166.41 ± 9.28	164.71 ± 7.89	162.12 ± 8.11
Weight (kg)	67.78 ± 13.81	66.17 ± 12.03	65.19 ± 11.42
BMI (kg/m^2)	24.45 ± 3.82	24.34 ± 3.72	24.64 ± 3.62
FBG (mmol/L)	5.13±0.63	5.33±0.76	5.45±0.93
Heart rate (bpm)	71.32±9.17	73.18±10.62	74.8±11.8
Waist (cm)	83.21±10.61	86.1±10.16	89.14±10.42
Smoking Status			
Never	154	86	207
Former smoker	39	35	51
Current smoker	25	28	88
Alcohol consumption			
Never	147	84	196
Former	1	3	3
Light	57	49	107
Moderate	5	6	22
Heavy	4	5	7
Diabetes types			
Type 1	36	12	16
Type 2	188	200	301
Others	3	2	2
Complication of diabetes			
Nephropathy	13	54	83
Retinopathy	26	47	94
Peripheral neuropathy	75	102	163
LDL-C (mmol/L)	2.43±0.62	2.72±1.11	2.81±0.83
HDL-C (mmol/L)	1.58±0.56	1.47±0.53	1.62±0.63
Uric acid (μmol/L)	283±92.8	329±97.6	332±92.8
hs-CRP (mg/L)	0.91 (0.37–1.91)	1.14 (0.45–2.31)	1.25 (0.60–2.71)
eGFR (mL/(min-1.73m^2))	110±21.5	98.1±18.4	91.2±18.5
ALT (U/L)	31.7 ± 24:5	26.1 ± 18.6	27.1 ± 18.5
AST (U/L)	20.8 ± 12.7	21.8±8.2	22.4 ± 11.8
Creatinine (μmol/L)	67.2 ± 13.8	66.2 ± 18.2	67.6 ± 14.3
BUN, mmol/L	5.21 ± 1.73	5.67 ± 2.91	6.24 ± 3.12
Fasting plasma glucose, mmol/L	7.82 ± 3.19	8.35 ± 2.82	9.35 ± 3.02
Total cholesterol (mmol/L)	4.08 ± 1.01	4.12 ± 1.27	4.68 ± 1.27
History of CVD (n, %)	23	26	36
HbA1c (%)	9.1±3.3	9.3±4.2	9.6±2.7
baPWV (cm/s)	1219 ± 129	1453± 131	1767 ± 305

groups such as low, medium and high risk. The baseline characteristics are shown in Table 1. A total of 230 patients with an age and standard deviation of 49 ± 14 are at low risk. Whereas 160 patients with an age and standard deviation of 55 ± 8 are in medium risk and 370 patients with a period and standard deviation of 61± nine are in the high-risk group. Significant baPWV differences are observed between these groups. The patient data from baPWV is listed in Table 1.

The proposed feature selection algorithm called Gaussian-Lasso-COA has been applied on these patient data to select the features for prediction. Lasso regression selects the top most 15 features are relevant for prediction. PCA removes the outliers of 2 features that are not correlated to each other. Gaussian with COA improved the efficiency of feature selection approach and select most important five features such as age, BMI, FBG, diastolic blood pressure (DBP), systolic blood pressure (SBP) are selected for further process. These features considered as the risk factors of arterial stiffness. The significant increase in the features age, SBP,DBP and FBG are in the high risk arterial stiffness group. The proposed algorithm also selects BMI also significant on predicting the AS risk factors. The increase in FBG also increases the baPWV which leads to the high risk factor of AS. Hence, the five important features are selected from our proposed optimal feature selection model.

Next, the dataset is divided randomly into training (70%) and testing (30%) datasets. Each time, the four machine and deep learning algorithms with the optimized hyper parameters are used on training data. The best model is assessed using the probability of AS on testing data using ROC, PRC analyses and AUC values computation. The comparative analysis shows that SVM obtained significantly minimum AUROC and AUPRC values. Whereas LSTM secured maximum AUROC and AUPRC value as shown in Figures 6-7. The LSTM outperformed than other evaluated machine and deep learning algorithms based on the capacity of the classification of AS.

Based on Figures 6-7, LSTM outperforms with improved AUROC and AUPRC values and it is finalized for our prediction model. The data is trained using LSTM on the discovery data with proposed feature selection model to predict the risk factors of arterial stiffness. The predictors with its information on discovery and validation data are shown in Table 2. For both cohorts, the AUROC and AUPRC values are as-

Figure 5
AUROC on testing data using four ML and DL algorithms

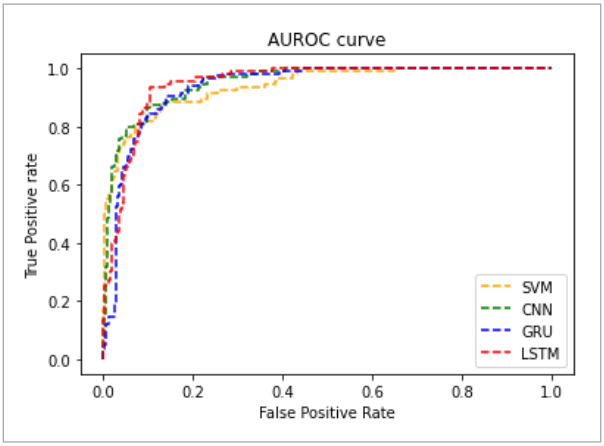


Figure 6
AUPRC on testing data using four ML and DL algorithms

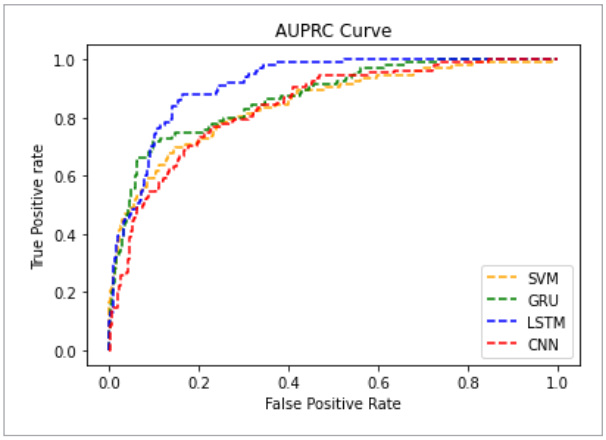


Figure 7
RMSE comparison of prediction methods

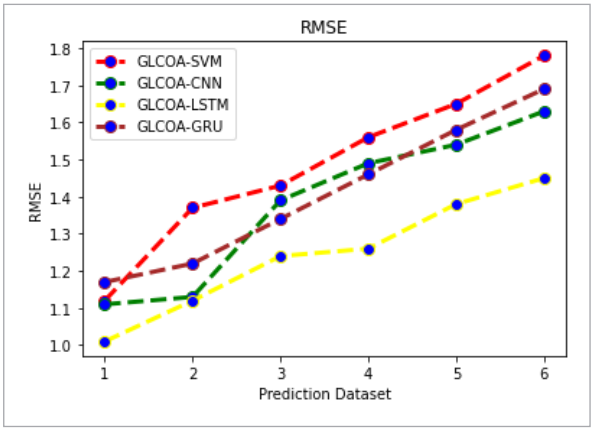


Table 2
Comparison of selected clinical characteristics between discovery and validation dataset

	Discovery dataset			Validation dataset		
	Low risk	Medium risk	High risk	Low risk	Medium risk	High risk
Age	49 ± 14	55 ± 8	61± 9	49 ± 7	54 ± 18	59± 29
BMI	24.45 ± 3.82	24.34 ± 3.72	24.64 ± 3.62	23.11 ± 3.21	24.12 ± 2.78	25.04 ± 3.91
FBG	5.13± 0.63	5.33± 0.76	5.45± 0.93	5.43± 0.81	5.33±0.06	5.45±0.98
SBP	115 ± 16	141 ± 21	142 ± 91	112 ± 17	125 ± 09	139 ± 82
DBP	71 ± 07	82 ± 32	83 ± 41	74 ± 08	82 ± 11	83 ± 52
baPWV	1219 ± 129	1453± 131	1767 ± 305	1219 ± 141	1453± 156	1767 ± 386

Figure 8
Error percentage of prediction methods

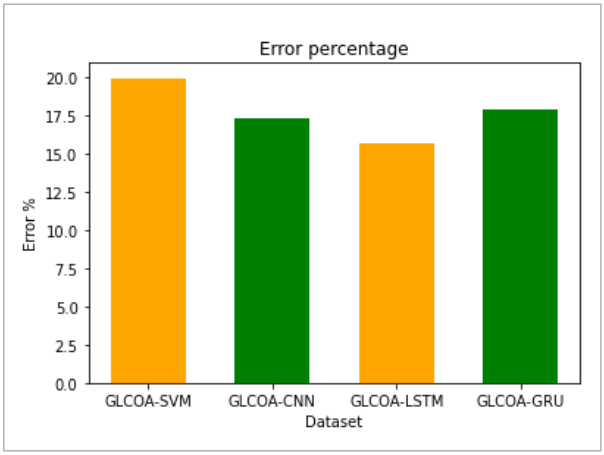
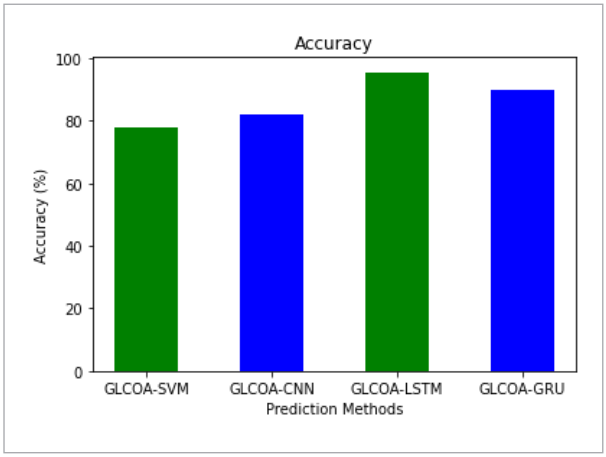


Figure 9
Accuracy



sessed and the results proves that the improved AUROC values of 0.985 and 0.916 and AUPRC value of 0.976 and 0.901 for discovery dataset and validation dataset respectively for the classification of low risk, medium risk and high risk of arterial stiffness.

The accuracy of the proposed model has been estimated with the proposed Gaussian-Lasso-COA (GLCOA) and four machine and deep learning models are applied on the training data. Figure 7 and Figure 8 shows the evaluation of the classification models in terms of error. With the proposed feature selection model, LSTM secured minimum RMSE value of 1.45 with the percentage 15.6%. The other approaches such as SVM secured 1.78 and 18.0%, CNN secured 1.63 and 15.6% and GRU secured 1.69 and 17.9% are shown in Figures 9-11, respectively.

Figure 10
Execution Time

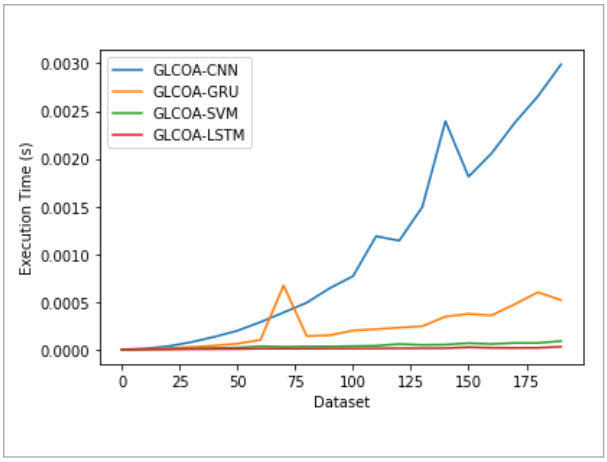
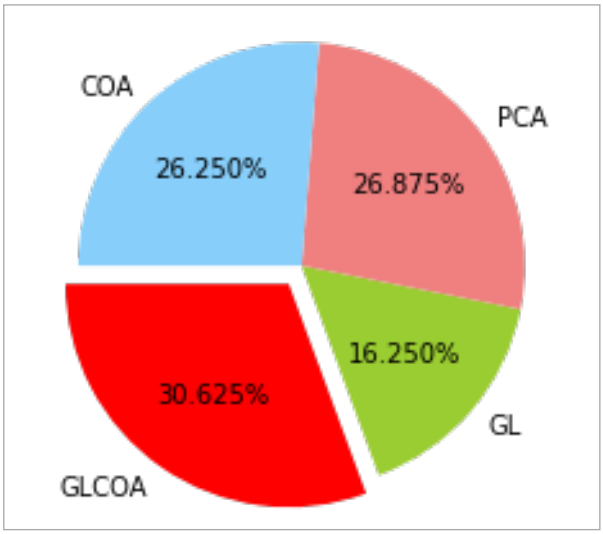


Figure 11
Feature Selection



5. Conclusion

This paper developed a machine and deep learning-based arterial stiffness prediction model to find the risk factors of diabetes patients based on the association between baPWV and FBG. The essential key features of the diabetes patient's biological characteristics are selected using LASSO regression. PCA has been used in this study to remove the outliers, which will increase the prediction accuracy. From the features chosen, Gaussian regression with Cat optimization will choose the most optimized five characteristics such as age, BMI, FBG, SBP, and DBP, which will produce the accurate classification of the arterial stiffness risk factors. Then, four machine and deep learning algorithms such as SVM, CNN, LSTM, and GRU were evaluated on predicting the arterial stiffness risk factors. The AUROC and AUPRC evaluation prove that LSTM performs better in predicting the risk as low, medium, and high-risk aspects with the proposed GLCOA-based feature selection. The proposed model secured the prediction AUROC of 0.985 and AUPRC of 0.976. The RMSE value and percentage in the proposed model are 1.45 and 15.6%.

In conclusion, arterial stiffness is related to the lead risk of diabetes patients. The high FBG and baPWV increase the risk of arterial stiffness and cause cardiovascular diseases in diabetes patients. The limitation of this study is the reduced training set size, and the model is suited and validated for discovery data set only. In the future, it could be evaluated in various datasets for result validation with an increased number of patient records.

References

1. Chung, Y. K., Lee, Y. J., Kim, K. W., Cho, R. K., Chung, S. M., Moon, J. S. Serum Cystatin C is Associated with Subclinical Atherosclerosis in Patients with Type 2 Diabetes: A Retrospective Study. Diabetes and Vascular Disease Research, 2018, 15, 24-30. https://doi.org/10.1177/1479164117738156
2. Fukuda, T., Hamaguchi, M., Kojima, T., Ohshima, Y., Ohbora, A., Kato, T. Data from: Association Between Serum γ-glutamyl-Transpeptidase and Atherosclerosis: a Population-based Cross-Sectional Study, 2014, Dryad Digital Repository. https://doi.org/10.1136/bmjopen-2014-005413
3. Girčys, R., Kazanavičius, E., Maskeliūnas, R., Damaševičius, R., Woźniak, M. Wearable System for Real-Time Monitoring of Hemodynamic Parameters: Implementation and Evaluation. Biomedical Signal Processing and Control, 2020, 59. https://doi.org/10.1016/j.bspc.2020.101873
4. Gomez-Sanchez, L., Garcia-Ortiz, L., Patino-Alonso, M.C., Recio-Rodriguez, J.I., Feuerbach, N., Marti, R., Agudo-Conde, C., Rodriguez-Sanchez, E., Maderuelo-Fernandez, J.A., Ramos, R. MARK Group Glycemic Markers and Relation with Arterial Stiffness in Caucasian Subjects of the MARK Study. PLoS One, 2017, 12, e0175982. https://doi.org/10.1371/journal.pone.0175982
5. Gonzalo-Calvo, De, D., Martínez-Camblor, P., Bär, C., Duarte, K., Girerd, N., Fellström, B. Improved Cardiovascular Risk Prediction in Patients with End-Stage Renal Disease on Hemodialysis Using Machine Learning Modeling and Circulating Micro Ribonucleic Acids Theranostics, 2020, 10, 8665-8676. https://doi.org/10.7150/thno.46123
6. Horton, W. B., Jahn, L. A., Hartline, L. M., Aylor, K. W., Patrie, J. T., Barrett, E. J. Insulin Increases Central Aortic Stiffness in Response to Hyperglycemia in Healthy

Humans: A Randomized Four-Arm Study. Diabetes and Vascular Disease Research, 2021, 18, 14791641211011009. https://doi.org/10.1177/14791641211011009

7. Huang, S., Li, J., Shearer, G.C., Lichtenstein, A. H., Zheng, X., Wu, Y., Jin, C., Wu, S., Gao, X. Longitudinal Study of Alcohol Consumption and HDL Concentrations: A Community-Based Study. The American Journal of Clinical Nutrition, 2017, 105, 905-912. https://doi.org/10.3945/ajcn.116.144832
8. Jagadeesh Kumar, S. J. K., Parthasarathi, P., Mehedi, M., Jehad, F., Al-Amri, M., Abouhawwash, Butterfly Optimized Feature Selection with Fuzzy C-Means Classifier for Thyroid Prediction. Intelligent Automation and Soft Computing, 2022, 35(3), 2909-2924. https://doi.org/10.32604/iasc.2023.030335
9. Jin, C., Chen, S., Vaidya, A., Wu, Y., Wu, Z., Hu, F. B., Kris-Etherton, P., Wu, S., Gao, X. Longitudinal Change in Fasting Blood Glucose and Myocardial Infarction Risk in A Population Without Diabetes. Diabetes Care, 2017, 40, 1565-1572. https://doi.org/10.2337/dc17-0610
10. Jin, W., Chowienczyk, P., Alastruey, J. Estimating Pulse Wave Velocity from the Radial Pressure Wave using Machine Learning Algorithms. PLoS ONE, 2021, 16(6). https://doi.org/10.1371/journal.pone.0245026
11. Kumar, A., Mishra, D., Cat Swarm Based Optimization of Gene Expression Data Classification. International Journal of Computer Trends and Technology (IJCTT), 2013, 4, 1185.
12. Levey, A. S., Stevens, L. A., Schmid, C. H., Zhang, Y. L., Castro, A. F., Feldman, H. I., Kusek, J. W., Eggers, P., Van, Lente, F., Greene, T. CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A New Equation to Estimate Glomerular Filtration Rate. Ann Intern Med, 2009, 150, 604-612. https://doi.org/10.7326/0003-4819-150-9-200905050-00006
13. Lim, T. H., Chung, S. M., Lee, D. S., Choi, S. R., Moon, J. S., Yoon, J. S., Won, K. C., Lee, H. W. Peripheral Arterial Stiffness Increases the Risk of Progression of Renal Disease in Type 2 Diabetic Patients. Frontiers in Medicine, 2020, 7, 588967. https://doi.org/10.3389/fmed.2020.588967
14. Li, Q., Xie, W., Li, L., Wang, L., You, Q., Chen, L., Li, J., Ke, Y., Fang, J., Liu, L., Hong, H., Development and Validation of a Prediction Model for Elevated Arterial Stiffness in Chinese Patients with Diabetes Using Machine Learning. Frontiers in Physiology, 2021, 12, 714195. https://doi.org/10.3389/fphys.2021.714195
15. Loehr, L. R., Meyer, M. L., Poon, A. K., Selvin, E., Palta, P., Tanaka, H., Pankow, J. S., Wright, J. D., Griswold, M. E., Wagenknecht L. E., Prediabetes and Diabetes are Associated with Arterial Stiffness in Older Adults: The ARIC Study. American Journal of Hypertension, 2016, 29, 1038-1045. https://doi.org/10.1093/ajh/hpw036
16. Marina, C., Keehn, L., Ye, L., Spector, T. D., Hughes, A. D., Chowienczyk, P. Genetic Aetiology of Blood Pressure Relates to Aortic Stiffness with Bi-Directional Causality: Evidence from Heritability, Blood Pressure Polymorphisms, and Mendelian Randomization. European Heart Journal, 41(35), 3314-3322. https://doi.org/10.1093/eurheartj/ehaa238
17. Mengyi, Z., Zhang, X., Chen, S., Song, Y., Zhao, Q., Gao, X., Wu, S. Arterial Stiffness Preceding Diabetes: A Longitudinal Study. Circulation Research, 2020, 127, 1491-1498. https://doi.org/10.1093/eurheartj/ehaa238
18. Merril, E. F., Crichton, G. E., Dearborn, P. J., Robbins, M. A., Abhayaratna, W. P. Associations between Type 2 Diabetes Mellitus and Arterial Stiffness: A Prospective Analysis Based on the Maine-Syracuse Study. Pulse, 2017, 5(1-4), 88-98. https://doi.org/10.1159/000479560
19. Meysam, O., Yasin, O., Mohammad, M., Mohammad, T., A Novel Cat Swarm Optimization Algorithm for Unconstrained Optimization Problems. International Journal of Computer Science and Information Technologies, 2013, 11, 32-41. https://doi.org/10.5815/ijitcs.2013.11.04
20. Myszczynska, M. A., Ojamies, P. N., Lacoste, A. M. B., Neil, D., Saffari, A., Mead, R. Applications of Machine Learning to Diagnosis and Treatment of Neurodegenerative Diseases. Nat. Rev. Neurol, 2020, 16, 440-456. https://doi.org/10.1038/s41582-020-0377-8
21. Prenner, S. B., Chirinos, J. A., Arterial Stiffness in Diabetes Mellitus. Atherosclerosis, 2015, 238, 370-379. https://doi.org/10.1016/j.atherosclerosis.2014.12.023
22. Rafael, D. O. A., Santos, P. C. J. L., Musso, M. M., Cunha, R. D. S., Krieger, J. E., Mill, J. G., Pereira, A. C. Impact of Diabetes Mellitus on Arterial Stiffness in a Representative Sample of an Urban Brazilian Population. Diabetology and Metabolic Syndrome, 2013, 5, 45. https://doi.org/10.1186/1758-5996-5-45
23. Rajkomar, A., Dean, J., Kohane, I., Machine Learning in Medicine. Journal of the American Medical Association, 2019, 380, 1347-1358. https://doi.org/10.1056/NEJMra1814259
24. Sang, Y., Wu X., Miao, J., Cao M., Ruan, L., Zhang C., Determinants of Brachial-Ankle Pulse Wave Velocity

and Vascular Aging in Healthy Older Subjects. Medical Science Monitor: International Medical Journal of Experimental and Clinical Research, 2020, 26:e923112. https://doi.org/10.12659/MSM.923112

25. Takae, M., Yamamoto, E., Tokitsu, T., Oike, F., Nishihara, T., Fujisue, K. Clinical Significance of Brachial-Ankle Pulse Wave Velocity in Patients with Heart Failure with Reduced Left Ventricular Ejection Fraction. American Journal of Hypertension, 2019, 32, 657-67. https://doi.org/10.1093/ajh/hpz048
26. Usman, S.M., Bin, Ibne, Reaz, M., A. M. Ali, Risk Prediction of Having Increased Arterial Stiffness Among Diabetic Patients Using Logistic Regression. 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2016, 699-701. https://doi.org/10.1109/IECBES.2016.7843540
27. Vaidya, A., Cui, L., Sun, L., Lu, B., Chen, S., Liu, X., Zhou, Y., Liu, X., Xie, X., Hu, F. B. A Prospective Study of Impaired Fasting Glucose and Type 2 Diabetes in China: The Kailuan Study. Medicine, 2016, 95. https://doi.org/10.1097/MD.0000000000005350
28. Weber, T. Arterial Stiffness, Wave Reflections, and Diabetes: A Bidirectional Relationship? American Journal of Hypertension, 2010, 23, 1047-1048. https://doi.org/10.1038/ajh.2010.163
29. Wu, J., Wei, W., Zhang, L., Wang, J., Damasevicius, R., Li, J., Wozniak, M. Risk Assessment of Hypertension in Steel Workers Based on LVQ and Fisher-SVM Deep Excavation. IEEE Access, 2019, 7, 23109-23119. https://doi.org/10.1109/ACCESS.2019.2899625
30. Wu, S., An, S., Li, W., Lichtenstein, A.H., Gao, J., Kris-Etherton, P. M., Wu, Y., Jin, C., Huang, S., Hu, F. B. Association of Trajectory of Cardiovascular Health Score and Incident Cardiovascular Disease. JAMA Network Open, 2019, 2. https://doi.org/10.1001/jamanetworkopen.2019.4758
31. Xaplanteris, P., Vlachopoulos, C., Protogerou, A.D., Aznaouridis, K., TerentesPrintzios, D., Argyris, A.A. A Clinical Score for Prediction of Elevated Aortic Stiffness: Derivation and Validation in 3943 Hypertensive Patients. Journal of Hypertension, 2019, 37, 339-346. https://doi.org/10.1097/HJH.0000000000001904
32. Yaxin, A., Yang, Y., Cao, B., Dong, H., Li, A., Zhao, W., Ke, J., Zhao, D. Increased Arterial Stiffness as a Predictor for Onset and Progression of Diabetic Retinopathy in Type 2 Diabetes Mellitus, Hindawi Journal of Diabetes Research Volume, 2021, 9124656. https://doi.org/10.1155/2021/9124656
33. Zhang, Z., Liu, J., Xi, J., Gong, Y., Zeng, L., Ma, P. Derivation and Validation of an Ensemble Model for the Prediction of Agitation in Mechanically Ventilated Patients Maintained under Light Sedation. Critical Care Medicine, 2021, 49, e279- e290. https://doi.org/10.1097/CCM.0000000000004821
34. Zhang, C., Hong, H. Aging Cardiovascular Continuum. Chinese Journal of Geriatrics, 2019, 38, 1180-1184. doi: 10.3760/cma.j.issn.0254-9026.2019.10.029