International Journal of Nanotechnology

ISSN online: 1741-8151 - ISSN print: 1475-7435
https://www.inderscience.com/ijnt

Diagnosing cardiovascular disease via intelligence in healthcare multimedia: a novel approach

Geeitha Senthilkumar, Fadi Al-Turjman, Rajagopal Kumar, Jothilakshmi Ramakrishnan

DOI: 10.1504/IJNT.2023.10056474

Article History:

Received:	20 January 2021
Last revised:	24 May 2021
Accepted:	02 June 2021
Published online:	31 May 2023

Diagnosing cardiovascular disease via intelligence in healthcare multimedia: a novel approach

Geeitha Senthilkumar*

Department of Information Technology,
M. Kumarasamy College of Engineering,
Karur, Thalavapalayam, 639113, Tamil Nadu, India
Email: geeitha.s@gmail.com
*Corresponding author

Fadi Al-Turjman

Faculty of Engineering,
Research Center for AI and IoT,
University of Kyrenia, Mersin 10, Turkey

and

Artificial Intelligence Engineering Department,
AI and Robotics Institute,
Near East University, Nicosia, Mersin 10, Turkey
Email: fadi.alturjman@neu.edu.tr

Rajagopal Kumar

Department of Electronics and Instrumentation Engineering,
National Institute of Technology,
Chumkedima, Dimapur, Nagaland-797103, India
Email: rajagopal.kumar4@gmail.com

Jothilakshmi Ramakrishnan

Department of Mathematics,
Mazharul Uloom College Ambur,
Ambur, 635802, Tamil Nadu, India
Email: jothilakshmiphd@gmail.com

Abstract: Diagnosing cardiovascular disease (CVD) in its early stages remains a challenge despite the existence of all medical technologies and devices that are being used. Besides the digitised form of collecting and organising data, prediction and diagnosis are two stumbling blocks in CVD. This study explores statistical machine learning models with a multimedia health care approach using AI to predict risk factors of heart diseases associated with type 2 diabetes mellitus (T2DM). This study investigates an efficacy of a mathematical model to perform attribute evaluation using information criteria-based selection in LASSO regression. The present study implements the deep learning algorithm using a multilayer perceptron (MLP) classifier with Gaussian process

classification (GPC) that provides probabilistic predictions in terms of linear and non-linear functions. The performance of the classifier is evaluated using precision, recall and accuracy metrics. The proposed classification model yields 93.59% accuracy of 10 cross-validations assorted with sigmoid function for better analysis.

Keywords: AI; artificial intelligence; CVD; cardiovascular disease; multimedia health care; feature selection; T2DM; type 2 diabetes mellitus.

Reference to this paper should be made as follows: Senthilkumar, G., Al-Turjman, F., Kumar, R. and Ramakrishnan, J. (2023) 'Diagnosing cardiovascular disease via intelligence in healthcare multimedia: a novel approach', *Int. J. Nanotechnol.*, Vol. 20, Nos. 1/2/3/4, pp.182–198.

Biographical notes: Geeitha Senthilkumar, Assistant Professor, has completed Master of Engineering in Computer Science and Engineering in Anna University. Her research expertise covers Medical data mining, machine learning, artificial intelligence, cloud computing, big data, fuzzy, soft computing and ontology. She has published 20 international journals and has presented 18 papers in national and international conferences in the above fields. She has got senior educator and researcher award from national foundation for entrepreneurship development award. She is currently working as Associate Professor, Information Technology Department in M. Kumarasamy College of Engineering, Karur, Thalavapalayam, Tamil Nadu.

Fadi Al-Turjman received his PhD in Computer Science from Queen's University, Canada, in 2011. He is the Associate Dean for research and the Founding Director of the International Research Center for AI and IoT at Near East University, Nicosia, Cyprus. He is the Head of Artificial Intelligence Engineering Dept., and a leading authority in the areas of smart/intelligent IoT systems, wireless, and mobile networks' architectures, protocols, deployments, and performance evaluation in Artificial Intelligence of Things (AIoT). His publication history spans over 350 SCI/E publications, in addition to numerous keynotes and plenary talks at flagship venues. He has authored and edited more than 40 books about cognition, security, and wireless sensor networks' deployments in smart IoT environments, which have been published by well-reputed publishers such as Taylor and Francis, Elsevier, IET, and Springer. He has received several recognitions and best papers' awards at top international conferences. He also received the prestigious Best Research Paper Award from Elsevier Computer Communications Journal for the period 2015–2018, in addition to the Top Researcher Award for 2018 at Antalya Bilim University, Turkey. He has led a number of international symposia and workshops in flagship communication society conferences. Currently, he serves as book series editor and the lead guest/associate editor for several top tier journals, including the IEEE Communications Surveys and Tutorials (IF 23.9) and the Elsevier Sustainable Cities and Society (IF 7.8), in addition to organizing international conferences and symposiums on the most up to date research topics in AI and IoT.

Rajagopal Kumar received the B-Tech degree in Electronics and Communication Engineering from Madras University, Tamil Nadu, India in 1991 and MTech degree in Instrumentation Engineering, Anna University, Tamil Nadu, India in 1994. He received PhD degree in Embedded Systems and Control Engineering from NIT, Trichy, 2007. He is currently working as Professor in Electronics and Instrumentation Engineering and Dean Academic of NIT Nagaland, India. He has 80 publications in international journals and conferences. His areas of interests include MEMS, sensors and actuators, wireless sensor networks (WSN), internet of things and embedded systems design.

Jothilakshmi Ramakrishnan obtained PhD from Thiruvalluvar University, Tamil Nadu, India in 2015. She is presently working as an Assistant Professor and Head, Department of Mathematics, Mazharul Uloom College, Ambur, Tamil Nadu, India. Her area of interest is difference equations, digital signal processing, artificial intelligence and internet of things. She has published more than 30 research articles in the International journals.

1 Introduction

People with type 2 diabetic mellitus (T2DM) are often more prone to be diagnosed with a cardiovascular disease (CVD). According to the National Heart Association (NHA), it has been reported that nearly 65% of the patients with diabetes complaint die due to heart disease. CVD occurs when the heart is incapable of pumping the required amount of blood to other parts of the body [1]. The cardiovascular system comprises arteries, capillaries, and veins which work together to provide blood circulation throughout the entire body [2]. Specific abnormalities in the blood circulation from the heart may lead to CVDs. These may damage the heart and the blood vessels [3]. In recent years, there has been a remarkable increase in heart diseases among middle-aged people whose age spanning between 30–40 years old due to some significant factors [4] such as improper food habits, insufficient sleep, depression, unbalanced diet, family background, high blood pressure, high cholesterol, smoking and exhaustive nature of job [5]. According to the World Health Organization (WHO) report on CVD prevention and control, CVDs are considered the leading cause of mortality all over the world [6]. The symptoms of CVD include short breath, physical weakness, fatigue with relevant signs, some abnormalities in cardiac functioning, and others [7]. Over the years there have been a lot of medical data collected and provided by the health care industry that can be utilised to implement machine learning algorithms to be able to identify a CVD in its early stages [8]. Diagnosing CVD and imparting treatment is very difficult due to the lack of diagnosing apparatus and physician scarcity which in return are affecting the prediction and treatment of the concerned patients in the early stages of the disease [9]. Similarly, diabetes mellitus is one of the chronic conditions that prevail due to insufficient or ineffective use of the insulin produced by the body. When compared with patients without diabetes, individuals with M2TD possess a higher risk of CVD [10]. T2DMis considered as one of the important risk factors of CVD [11]. It has been estimated that nearly 415 million people across the world suffer from this disease and their age falls between 20 and 79 years old. This poses a critical risk factor in a human's life and can increase the mortality rate [12]. Many machine learning models have been developed and reported in various research works. Some of the machine learning models are support vector machine (SVM), logistic regression (LR), K-nearest neighbour (KNN), artificial neural networks (ANNs), and Adaboost (AB) [13, 14] have been utilised in predicting and diagnosing heart disease. Implementing these machine-learning techniques is useful in reducing the mortality ratio due to heart disease to a certain extent [15]. On the other hand, traditional based approaches are lagging in providing better results in predicting and diagnosing CVDs. Thus knowledge mining with a deep learning-based prediction model contributes a lot in predicting and monitoring the CVD patients from the time of admission [16]. In some hospitals, real-time facial disorders detection(RFDD) are made

available as public cameras to take images of the patients and are processed using WiFi services in the Internet of Things (IoT) environment to read the patients' facial disorders currently utilised by dermatologists expert [17].

In the proposed work, preprocessing is carried out with subspace clustering and RBF model to solve the data imputation. Principal component analysis (PCA) with the LASSO model is deployed for feature selection. Deep learning-based MLP with Gaussian process classifier (GPC) is implemented for classifying the target variables described in Section 3.

2 Literature review

Singh et al. [18] developed a model for an effective heart disease prediction system (EHDPS) using a neural network for risk level prediction of heart disease. The model has undertaken 15 attributes. The system also predicts the correlation between medical factors and the disease pattern. The model is deployed with the multilayer perceptron (MLP) with backpropagation for training the dataset.

Singh et al. deployed a new algorithm for feature selection where the features are selected based on the Correlation method. It is named Correlation-based feature selection (CFS) and is considered the best evaluator where only the subset of original features is selected. The method is embedded with the best first search method to reduce the dimensionality [19]. Chi-square statistical method was also utilised to better select the significant features [20].

Olaniyi and Oyedotun [21] deployed a three-step model on the ANN-based algorithm to predict the CVD and achieved 88.89% classification accuracy.

Jabbar et al. [22] proposed a novel system diagnosing heart disease by implementing the MLP classifier, an ANN with backpropagation technique and feature selection algorithm.

Samuel et al. [23] proposed a hybrid model with ANN and fuzzy AHP for a decision support-based medical system. This integrated model deployed a machine learning algorithm, namely ANN, the fuzzy analytical hierarchical methodology by achieving 91% of classification accuracy.

Anitha and Sridevi [24] proposed a framework for heart disease prediction by implementing machine learning techniques in the R tool. The techniques such as SVM, KNN and Naïve Bayes (NB). The experimental results reveal that the NB classifier produces more accuracy in predicting heart disease than the SVM and KNN.

Sathish Kumar and Padmapriya [25] proposed a prediction in terms of similarities of CVD by implementing the ID3 algorithm over television and mobile phone.

Rajesh et al. [26] proposed a Naïve Bayes model giving all possibilities that produced accurate results. The results of the NB algorithm would be accurate only when data is properly cleaned and well maintained.

Annepu and Gowtham [27] designed a framework using a machine learning model called RF for heart disease prediction. Cleveland datasets were used for training and testing the samples. Totally 303 instances with 9 features were exploited for analysis. Experimental results were obtained using the graphical user interface (GUI) under visual studio code. But the model could predict up to 50% of heart diseases.

Narain et al. [28] proposed a comparison amidst two methods in heart disease prediction. The author devised the Framingham risk score (FRS) and quantum neural

network (QNN) techniques. A heart disease dataset with 689 instances was trained, and the study was implemented in 5209 samples taken from the University of Washington. The QNN architecture was designed with one output node and 7 input nodes. A total of 8 hidden nodes were identified. Finally, the model proved that the QNN model achieved higher accuracy than the FRS model.

Gudadhe et al. [29] proposed a deep learning model MLP and SVM for heart disease classification and obtained an accuracy of 80.41%.

Jacob et al. [30] proposed an artificial muscle intelligence with deep learning (AMIDL) system that integrates the user intensions with help of artificial muscle movements in an effective method to enhance the performance. Electroencephalogram (EEG) sensors are inserted in the human body for analysing the movements with a microcontroller and transcutaneous electrical nerve stimulation (TENS) equipment.

Similarly, Kahramanli and Allahverdi [31] launched a hybrid technique that integrates neural networks and fuzzy and obtained a classification accuracy of 87.4%.

A genetic algorithm associated with a recurrent fuzzy neural network is implemented for diagnosing CVD producing an accuracy of 97.78% [32] and designed a hybrid model that integrates Naïve Bayesian classifier and particle swarm optimisation (PSO) for heart disease prediction. The model was trained with 270 observations with 14 attributes. The experimental results proved that the NB classifier with the PSO model for feature selection provided more accuracy [33].

Al-Turjman et al. [34] proposed a 5G-inspired Industrial Internet of Things (IIOT) in the health care domain. This enables the user to interact with different types of sensors through wireless medical sensor networks (WMSNs). And for providing security the author has devised a context-sensitive seamless identity provisioning (CSIP) framework for IIOT [35].

Chakraborty and Abougreen [36] reported that AI technology can be used as a trustworthy for detecting COVID-19 by diagnosing using CT images and X-ray. Besides the author claimed that machine learning (ML) can be deployed in predicting and segmenting the COVID-19 features and also supports for drug discovery procedure reducing the clinical failures.

Despite many ML models launched and discovered for predicting the CVD, the study herein focused on the embedded model of combining the two classifiers for boosting the performance of feature optimisation with the statistical algorithm to proliferate the prediction accuracy. The proposed model solves the data imputation problem by implementing the subspace clustering with the RBF model and the prediction accuracy is improved by integrating the MLP classifier with the GPC model.

3 Methods and materials

Cleveland dataset and diabetes dataset are utilised in this research. These datasets are available online at the UCI repository [37]. The dataset comprises 76 attributes, among which 14 attributes of 303 patients are used. The patient's name and social security number have been removed for confidentiality. In the diabetes dataset, there are 390 instances with 19 attributes. Thus, the extracted dataset comprises the integrated features to classify CVD patients with T2DM. A description of the dataset that has been deployed in this research process is shown in Table 1.

Table 1 CVD feature description of Cleveland dataset

Feature code no	*Feature*	*Description*
#3	Age	Age in terms of years
#4	Sex	1 = male; 0 = female
#9	CP	Chest pain type (Val = 1 to 4)
#10	trestbps	Resting blood pressure (in mm Hg)
#12	chol	Serum Cholesterol in mg/dl
#16	fbs	Fasting blood sugar > 120 mg/dl (1 = *T*;0 = *F*)
#19	restecg	Resting electrocardiographic results
#32	thalach	Maximum Heart rate achieved
#38	exang	Exercise-induced angina (1 = Y; 0 = N)
#40	oldpeak	ST depression induced by exercise relative to rest
#41	slope	The slope of the peak exercise ST segment
#44	ca	Number of significant vessels (0–3) coloured by fluoroscopy
#51	thal	3 = Normal;6 = fixed defect;7 = reversible defect
#58	num	Diagnosis of heart disease (Angiographic disease status)

The proposed method aims at classifying CVD patients with diabetes. The dataset is pre-processed by subspace clustering and RBF model for solving missing data imputation by mean and median mode. After pre-processing, the CVD and diabetes features are selected by principal component attribute evaluator (PCA) [38]. The features are also resampled using information criterion-based selection under the LASSO regression model. MLP is then implemented with Gaussian processing classification (GPC) and stochastic gradient descent (SGD). GPC is forked out with probabilistic prediction. The proposed method is structured into five stages:

a data pre-processing

b feature selection

c statistical evaluation with cross-validation (CV)

d classification

e performance evaluation.

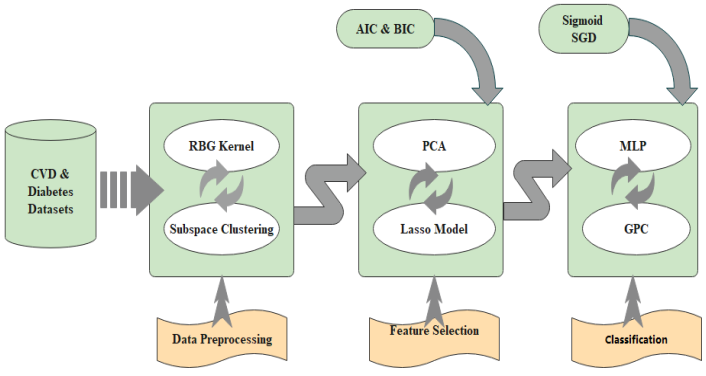
Figure 1 shows the architecture of the proposed model.

3.1 Data normalisation

Data imputation is carried out using RBF kernel for the diabetes dataset and subspace clustering for the Cleveland dataset to find valid clusters and to reduce the redundancy dimensions. There are 76 features where they are imputed to 14 relevant features. With RBF kernel, the dataset was normalised using Euclidean distance to reduce the feature space where the two samples y and y' represent feature vector such as the RBF kernel, which can be defined as:

$$k(y,y) = \exp\left[\left(\frac{||y-y'||^2}{2\sigma^2}\right)\right] \tag{2.1}$$

Figure 1 Architecture diagram of the proposed model (see online version for colours)



3.2 Feature selection

Proper feature selection is necessary to improve the performance of the classification. If feature selection was not performed, irrelevant features [39] may affect the performance of the classifier. An efficient attribute evaluator, namely principal component attribute (PCA) is implemented to evaluate the features by constructing a correlation matrix [40] ranking 10 attributes among the 14 features with an eigenvector. To estimate the sparse coefficients, a linear LASSO regression model is performed by implementing the Akaike information criterion (AIC) and Bayes information criterion (BIC) for computing optimal value with 10-fold cross-validation (CV) method. This method regularises the parameter α of the LASSO estimator. For computing, in this model two algorithms are mainly used, namely Gradient descent and Lars which produce approximately the same results but only differ in speed. Meanwhile, gradient descent is mainly used to solve the linear and non-linear equations and quadratic minimisation problem using linear least squares. It is solved by:

$$Ay - b = 0 \tag{2.2}$$

where A and b are Euclidean norms applied to find the local optimal step size α for every iteration with the ratio of maximum to minimum eigenvalues of A with y as samples. This minimising function is defined as:

$$f(x) = ||Ay - b||^2 \tag{2.3}$$

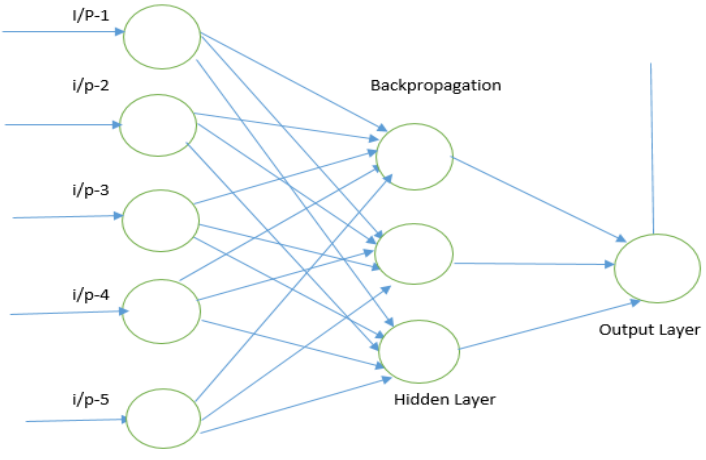
3.3 Classification

The proposed work is carried out with MLP using an SGD algorithm embedded with GPC, and it is compared to three classifiers: logistic regression (LR), random forest (RF) and Adaboost (AB). Their performance is estimated with the performance metrics: precision, recall and accuracy.

3.4 Multilayer perceptron (MLP)

Total samples of 303 patients with 10 extracted features were taken as training samples with 2 target values. MLP is a deep learning model that solves non-linear problems and can be applied for complex paradigms due to its supervised learning with a backpropagation algorithm. Figure 2 depicts the MLP architecture with many input layers with one or more hidden layers with output layers. For each layer, an independent weight is assigned with the corresponding hidden layer.

Figure 2 The architecture of MLP with backpropagation model (see online version for colours)



Let $(sx_1, sy_1), (sx_2, sy_2) \ldots (sx_n, sy_n)$ where $sx_i \in \mathbb{R}^m$ (where R is a regularisation function and R^m regularisation for the model parameter and $sy_i \in \{0,1\}$) be used for classification. The linear scoring function with model parameters a_1, a_2 is given by:

$$f(sx) = (a_1.a_2)^T x + b \tag{2.4}$$

where a_1, $a_2 \in \mathbb{R}^m$, $b \in \mathbb{R}$. The model is trained using SGD to solve the optimisation problem. This algorithm iterates over training samples, and updates the model parameter for every sample according to the update rule presented as below:

$$w = w - \eta\left(\alpha \frac{\partial R(w)}{\partial w} + \frac{\partial L}{\partial w}\right) \tag{2.5}$$

where w is a partial loss, L is loss function in-network, α is parameter and η is a learning rate that controls the step size in the parameter space search. With 303 samples and 14 features, one hidden layer and one neuron MLP, the updated function becomes as given below:

$$F(x) = w_2 g\left(w_1^n sx_n + a_1\right) + a_2 \tag{2.6}$$

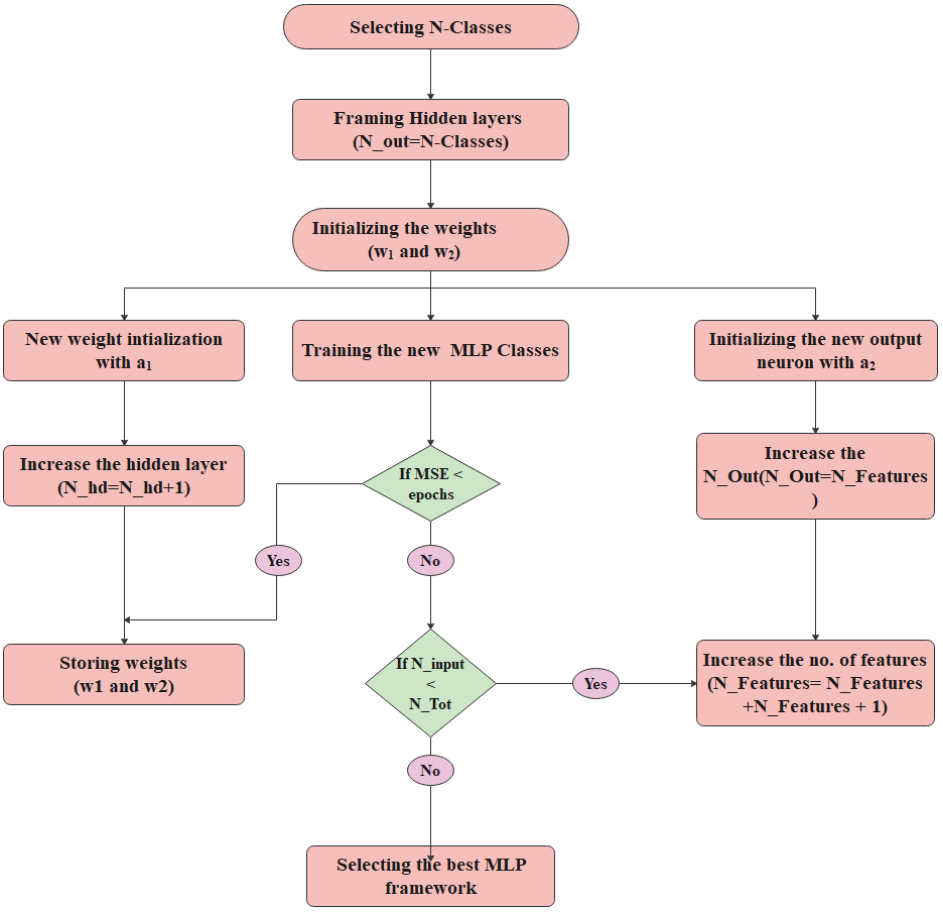
where $w_1^n \in \mathbb{R}^m$ and $w_2, a_1, a_2 \in \mathbb{R}$ are model parameters. Here w_1 and w_2 denote weights of the input layer and hidden layer respectively, where a_1, a_2 indicates the bias that is added to the hidden layer and output layer [41] correspondingly. Figure 3 propounds the

training classes using the MLP algorithm by assigning weight values. R is an activation function. The gradient Loss function concerning weights is computed and reduced from w. From equation (2.5), the regularised training error can be minimised by:

$$E(w,b) = \frac{1}{n}\sum_{i=1}^{n} L(sy_i, f(sx_i)) + R(w) \tag{2.7}$$

L is a loss function that fits the model and R regularisation function that reduces the model complexity.

Figure 3 Deep learning algorithm using MLP architecture (see online version for colours)



3.5 *Gaussian process classifier (GPC)*

GPC is mainly used for probabilistic classification that performs binary classification based on training and prediction. It provides a GP prior with latent function F, to obtain a probabilistic classification. This function is so-called a nuisance function to remove the irrelevant values during prediction. In our proposed model, two datasets are integrated to

predict CVD patients whose risk score of T2DM is high. The remaining features are eliminated using this latent function *F*.

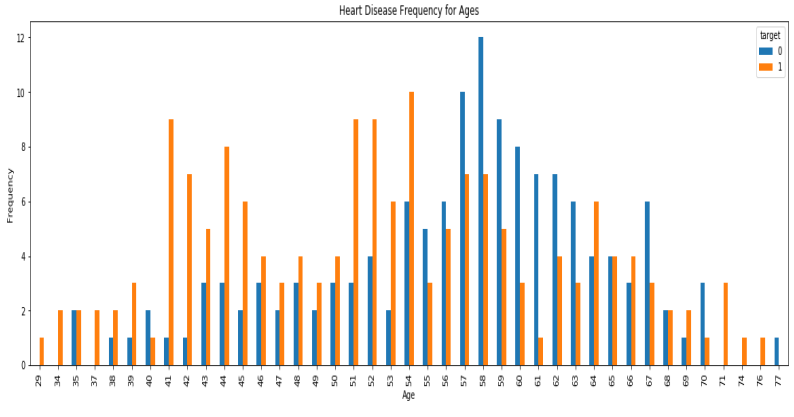
4 Results and discussions

This section discusses the classification models and the outcome from the various classifiers with feature selection models. Figure 4 portrays CVD with age frequency and the target class. For classifying the CVD associated with T2DM, the diabetes risk score (DRS) is computed using five variables: age, blood pressure (BP), insulin, body mass index (BMI), and cholesterol. The risk score is integrated with the CVD features, and the weight is calculated with the product of the threshold level of 7 nodes using the sigmoid function and DRS as depicted in Table 2. The experimental results are well-formed using Python [42, 43]

Table 2 Weight computation using gradient descent

Feature code no.	*Feature*	*GD Weight = T. level × DRS*
#9	cp	1.46
#19	restecg	2.60
#12	chol	3.58
#44	ca	3.63
#38	exang	4.62
#16	FBS	5.96
#10	trestbps	5.21

Figure 4 Cardiovascular disease with AGE frequency vs. target class (see online version for colours)



Among 76 features, 14 features are taken for feature analysis. Figure 5(a) and (b) represent the cholesterol and cardiographic results concerning the changes in the heart rate.

Figure 5(a) Cholesterol level vs. maximum heart rate achieved (see online version for colours)

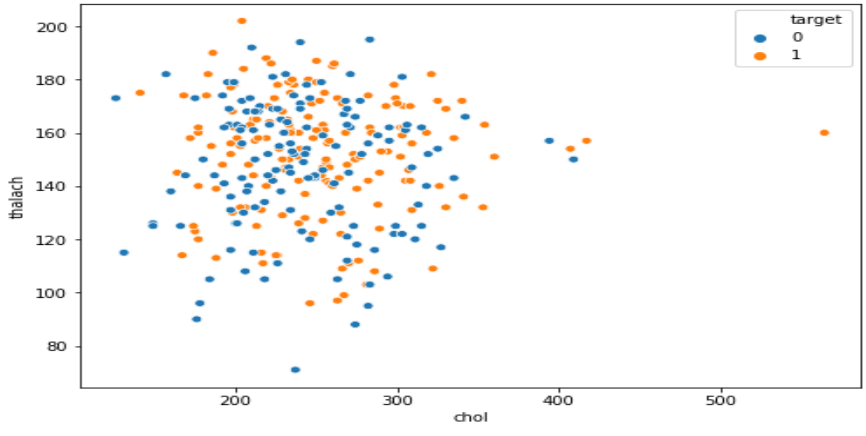
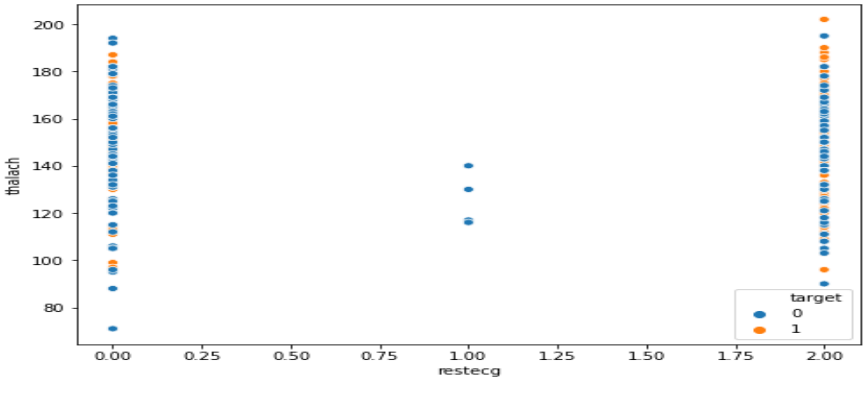


Figure 5(b) Resting cardio graphical results vs. maximum heart rate achieved (see online version for colours)



LASSO model for feature selection [44] along with PCA evaluator is utilised to select 10 relevant attributes using the rank function and computing correlation matrix with eigenvalues. Information criterion-based selection is carried out in the feature selection stage by estimating AIC and BIC estimators with 10 cross-validations and it is presented in Figure 6. With a training time of 0.01 s, the relevant features are selected when the parameter α which falls closer to 1.

During pre-processing, the RBF kernel is implemented to impute missing data. The same RBF kernel is then used for training and testing the CVD data samples. The results initially and after optimisation are shown in Figure 7. The length scale is approximately fixed between $0.25 < x < 0.5$. The initial kernel length scale is closer to approximately 0.25 while it is closer to approximately 0.5 for the optimised kernel state. And the probability prediction was also found to be around 1 exhibited by the RBF on the graph.

Figure 6 Alpha – parameter estimation using AIC and BIC information criterion (see online version for colours)

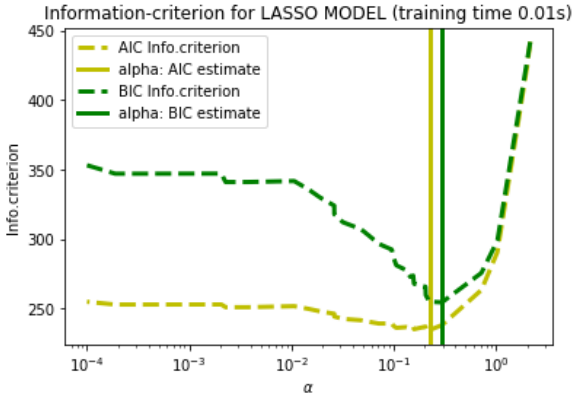
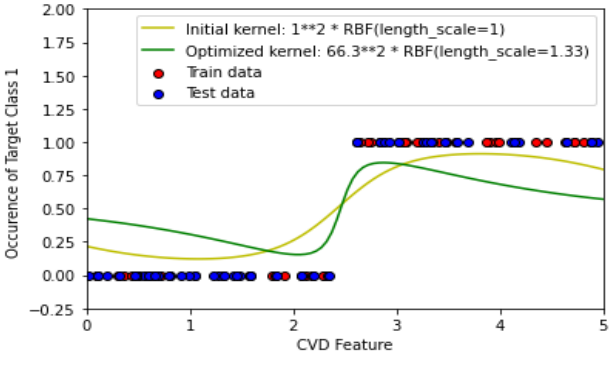


Figure 7 RBF kernel for training and testing data samples (see online version for colours)

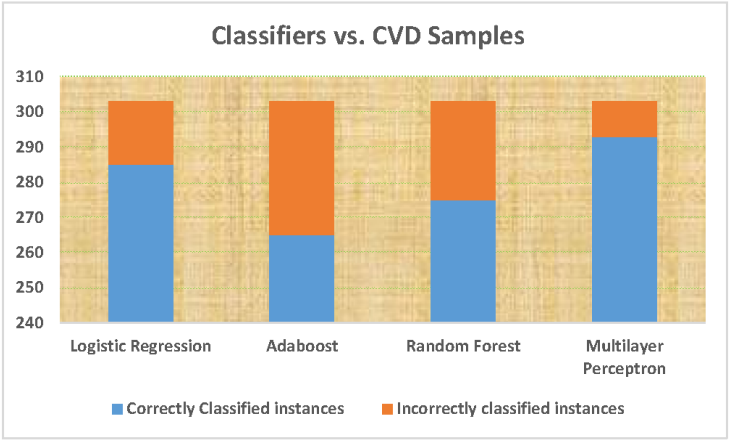


The proposed system classifies the training and testing samples with MLP classifier using SGD algorithm and proves that the proposed classifier correctly identifies 293 instances negating only 10 instances. The other models such as Logistic Regression, Adaboost, Random Forest classify 285, 265, 275 instances respectively. GPC classification supports probabilistic prediction in the hidden layer neurons; see Table 3. The corresponding graphical representation of the classifiers is shown in the Figure 8.

Table 3 ML models for predicting target labels

ML models	*Correctly classified instances*	*Incorrectly classified instances*
Logistic regression	285	18
Adaboost	265	38
Random forest	275	28
Multilayer perceptron	293	10

Figure 8 Performance of ML models in CVD classifications (see online version for colours)

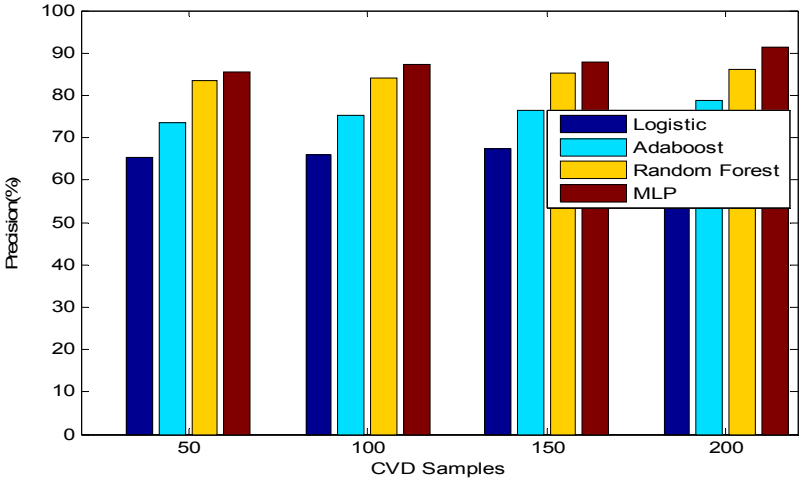


In the proposed model, three performance metrics have been used to estimate the efficiency of the classifier and they are presented in Table 4. In all three metrics, the proposed model MLP shows the highest score when compared with other ML models. The proposed model achieved a 92.65%, 89.34% and 95.59% for precision, recall and accuracy metrics respectively.

Table 4 Performance comparison of four models

Models	*Precision*	*Recall*	*Accuracy*
Logistic regression	68.34%	78.01%	85.05%
Adaboost	75.45%	79.25%	89.21%
Random forest	84.56%	87.15%	93.45%
Multilayer perceptron	89.65%	89.34%	93.59%

Figure 9 Precision vs. ML models (see online version for colours)



The performance of the classifiers is measured by splitting the samples into increments of 50 samples such as 50, 100, 150 and 200 samples. Figure 9 illustrates the higher performance of the proposed MLP model in terms of precision measurement. The proposed model has resulted in an 89.65% precision, whereas other models have resulted in lower precision values. This shows that the other ML models have a higher chance of incorrectly classifying patients with CVD than the proposed model.

Similarly, the recall and accuracy percentages have also been illustrated in Figures 10 and 11 respectively. Figure 10 shows how the proposed model achieves a higher recall percentage when compared to the rest of the ML models including Logistic, Adaboost, and Random forest. The proposed MLP model reached 89.34% recall.

Figure 10 Recall vs. ML models (see online version for colours)

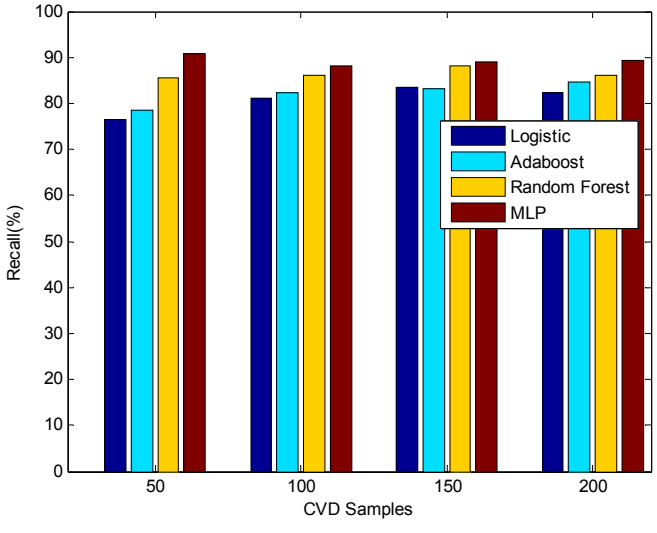


Figure 11 Accuracy vs. ML models (see online version for colours)

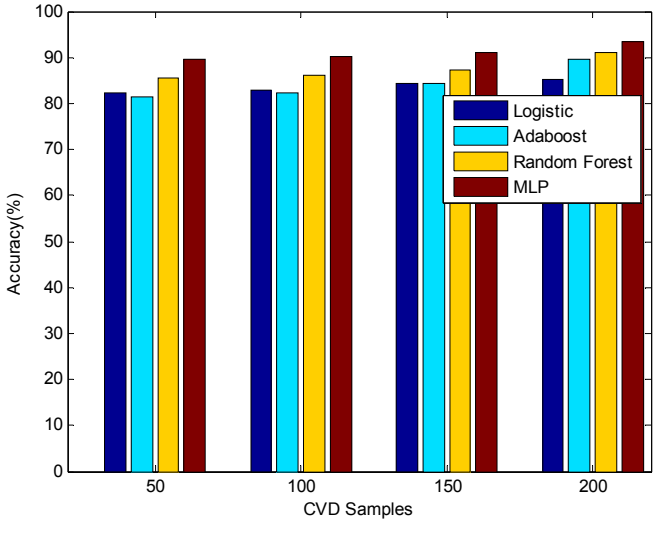


Figure 11 shows enhanced improvement in terms of accuracy metric of 93.59% when compared to all existing competing models proving that the proposed model is efficient in classifying and predicting CVD patients integrated with T2DM dataset.

5 Conclusion

In this paper, the integration model of MLP with Gaussian process classification (GPC) was proposed for predicting CVD with T2DM. The model was implemented on Cleveland CVD and Diabetes dataset. Two feature selection models were launched for attribute evaluators. The feature selection was then effectively carried out using AIC and BIC information criterion-based selection algorithms for obtaining an optimised outcome. The RBF kernel model was utilised to improve the classifier performance and optimise the kernel. GPCs were embedded with MLP to enhance the probabilistic prediction of CVD patients. Four classifiers were compared and estimated using three performance metrics. The proposed MLP model was found to have 89.65%, 89.34% and 93.59% of precision, recall and accuracy, respectively. These results showcase the superior performance of the MLP model in predicting patients with CVD when compared to other ML models including Logistic, Adaboost and Random forest. Despite the deep learning model, the analysed training data is not sufficient. Many datasets may be combined or large datasets can be taken for future analysis with this deep learning model.

References

1 Haq, U.I.A., Li, J.P., Memon, M.H., Nazir, S. and Sun, R. (2018) 'A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms', *Mob. Inf. Syst.*, Hindawi, Vol. 2018, Article ID 3860146, pp.1–21.

2 Hussein, M.U. (2017) *Physics and the Cardiovascular System*, 29 October, Retrieved from ResearchGate: https://www.researchgate.net

3 WHO (2011) *Global Atlas on Cardiovascular Disease Prevention and Control*, WHO Library Cataloguing, Geneva.

4 D'Souza, A. (2015) 'Heart disease prediction using data mining techniques', *Int. J. Res. Eng. Sci.*, Vol. 3, No. 3, pp.74–77.

5 Benjamin, H., David, F. and AntonyBelcy, S. (2018) 'Heart disease prediction using data mining techniques', *ICTACT J. Soft Comput.*, Vol. 9, No. 01, pp.1817–1823, doi: 10.21917/ijsc.2018.0253.

6 Nagendra, K.V. and Ussenaiah, M. (2018) 'A study on various data mining techniques used for heart diseases', *Int. J. Recent Sci. Res.*, pp.24350–24354.

7 Durairaj, M. and Ramasamy, N. (2016) 'A comparison of the perceptive approaches for reprocessing the data set for predicting fertility success rate', *Int. J. Optim. Control: Theor.*, Vol. 9, pp.256–260.

8 Dinesh, K.G., Arumugaraj, K., Santhosh, K.D., Mareeswari, V. 'Prediction of cardiovascular disease using machine learning algorithms', *2018 International Conference on Current Trends Towards Converging Technologies* (*ICCTCT*), Coimbatore, pp.1–7, doi: 10.1109/Icctct.2018.8550857.

9 Ghwanmeh, S., Mohammad, A. and Al-Ibrahim, A. (2013) 'Innovative artificial neural networks-based decision support system for heart diseases diagnosis', *J. Intell. Learn. Syst. Appl.*, Vol. 5, No. 3, pp.176–183.

10 Martín-Timón, I., Sevillano-Collantes, C., Segura-Galindo, A. and Del Cañizo-Gómez, F.J. (2014) 'Type 2 diabetes and cardiovascular disease: have all risk factors the same strength?', *World J. Diabetes*, Vol. 5, No. 4, pp.444–470, https://doi.org/10.4239/wjd.v5.i4.444

11 Dokken, B.B. (2008) 'The pathophysiology of cardiovascular disease and diabetes: beyond blood pressure and lipids', *Diabetes Spectr.*, Vol. 21, No. 3, pp.160–165.

12 Shan, P-F., Li, Q., Khamaisi, M. and Qiang, G-f (2017) 'Type 2 diabetes mellitus and macrovascular complications', *Int. J. Endocrinol.*, Vol. 2017, Article ID: 4301461, pp.1–2.

13 Nazir, S., Shahzad, S., Mahfooz, S. and Nazir, M. (2015) 'Fuzzy logic based decision support system for component security evaluation', *Int. Arab. J. Inf. Tech.*, Vol. 15, pp.1–9.

14 Nazir, S., Shahzad, S. and SeptemRiza, L. (2017) 'Birthmark-based software classification using rough sets', *Arab J. Sci. Eng.*, Vol. 42, No. 2, pp.859–871.

15 Methaila, A., Kansal, P., Arya, H. and Kumar, P. (2014) 'Early heart disease prediction using data mining techniques', *Proceedings of Computer Science and Information Technology* (*CCSIT-2014*), Sydney, NSW, Australia, Vol. 24, pp.53–59.

16 Kim, Y.J., Saqlian, M. and Lee, J.Y. (2019) 'Deep learning-based prediction model of occurrences of major adverse cardiac events during 1-year follow-up after hospital discharge in patients with AMI using knowledge mining', *Pers Ubiquit Comput*, https://doi.org/10.1007/s00779-019-01248-7

17 Al-Turjman, F.M. (2016) 'Towards smart ehealth in the ultra large-scale internet of things era', *2016 23rd Iranian Conference on Biomedical Engineering and 2016 1st International Iranian Conference on Biomedical Engineering (ICBME)*, Tehran, Iran, pp.102–105, doi: 10.1109/Icbme.2016.7890938.

18 Singh, P., Singh, S. and Pandi-Jain, G.S. (2018) 'Effective heart disease prediction system using data mining techniques', *Int. J. Nanomedicine*, Vol. 13, pp.121–124.

19 Xu, S., Zhang, Z., Wang, D., Hu, J., Duan, X. and Zhu, T. (2017) 'Cardiovascular risk prediction method based on CFS subset evaluation and random forest classification framework', *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, pp.228–232, doi: 10.1109/ICBDA.2017.8078813.

20 Singh, M., Martins, L.M., Joanis, P. and Mago, V.K. (2016) 'Building a cardiovascular disease predictive model using structural equation model & fuzzy cognitive map', *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp.1377–1382, doi: 10.1109/FUZZ-IEEE.2016.7737850.

21 Olaniyi, E., Oyedotun, O. and Khashman, A. (2015) 'Heart diseases diagnosis using neural networks arbitration', *Int. J. Intell. Syst. Appl.*, Vol. 7, pp.75–82, 10.5815/ijisa.2015.12.08.

22 Jabbar, M.A., Deekshatulu, B.L. and Chandra, P. (2013) 'Classification of heart disease using artificial neural network and feature subset selection', *Glob. J. Comput. Sci. Technol.*

23 Samuel, O.W., Asogbon, G.M., Sangaiah, A.K., Fang, P. and Li, G. (2017) 'An integrated decision support system based on ANN and fuzzy_AHP for heart failure risk prediction', *Expert Syst. Appl.*, Vol. 68, pp.163–172.

24 Anitha, S. and Sridevi, N. (2019) 'Heart disease prediction using data mining techniques', *J. Appl. Anal. Comput.*, Vol. 8, No. 2, pp.48–55.

25 Lakshmanan, S.K. and Arumugam, P. (2012) 'Prediction for common disease using ID3 Algorithm in mobile phone and television', *Int. J. Comput. Appl.*, Vol. 50, pp.30–33, 10.5120/7762-0830.

26 Rajesh, N., Maneesha, T., Hafeez, S. and Krishna, H. (2018) 'Prediction of heart disease using machine learning algorithms', *Int. J. Eng. Technol.*, Vol. 7, No. 2.32, pp.363–366.

27 Annepu, D. and Gowtham, G. (2019) 'Cardiovascular disease prediction using machine learning techniques', *Int. Res. J. Eng. Technol.*, Vol. 6, No. 4, pp.3963–3971.

28 Narain, R., Saxena, S. and Goyal, A.K. (2016) 'Cardiovascular risk prediction: a comparative study of framingham and quantum neural network-based approach', *Dovepress Journal: Patient Preference and Adherence*, Vol. 10, pp.1259–1270.

29 Gudadhe, M., Wankhade, K. and Dongre, S. (2010) 'Decision support system for heart disease based on support vector machine and artificial neural network', *Proceedings of International Conference on Computer and Communication Technology (ICCCT),* September, Allahabad, India, pp.741–745.

30 Jacob, S., Menon, V.G., Al-Turjman, F.V.P.G. and Mostarda, L (2019) 'Artificial muscle intelligence system with deep learning for post-stroke assistance and rehabilitation', *IEEE Access*, Vol. 7, pp.133463–133473, doi: 10.1109/ACCESS.2019.2941491.

31 Kahramanli, H. and Allahverdi, N. (2008) 'Design of a hybrid system for the diabetes and heart diseases', *Expert Syst. Appl.*, Vol. 35, Nos. 1–2, pp.82–89.

32 Samuel, O.W., Asogbon, G.M., Sangaiah, A.K., Fang, P. and Li, G. (2017) 'An integrated decision support system based on ANN and fuzzy_AHP for heart failure risk prediction', *Expert Syst. Appl.*, Vol. 68, pp.163–172.

33 Ilhan, A. and Uyar, K. (2017) 'Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks', *9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW*, August, Budapest, Hungary, pp.24–25.

34 Al-Turjman, F. and Alturjman, S. (2018) 'Context-sensitive access in industrial internet of things (IIoT) healthcare applications', *IEEE Trans. Industr. Inform.*, pp.1–1, doi: 10.1109/TII.2018.2808190.

35 Velusamy, N., Al-Turjman, F., Kumar, R. and Ramakrishnan, J. (2021) 'A framework for task allocation in IoT-oriented industrial manufacturing systems', *Comput. Netw.*, Vol. 190, p.107971.

36 Chakraborty, C. and Abougreen, A.N. (2021) 'Intelligent internet of things and advanced machine learning techniques for COVID-19', *EAI Endorsed Trans. Pervasive Health Technol.*, Vol. 7, No. 26, p.e1.

37 Kanchan B.D. and Kishor, M.M. (2016) 'Study of machine learning algorithms for special disease prediction using principal of component analysis', *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, pp.5–10, doi: 10.1109/ICGTSPICC.2016.7955260.

38 Kavitha, R. and Kannan, E. (2016) 'An efficient framework for heart disease classification using feature extraction and feature selection technique in data mining', *2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)*, pp.1–5, doi: 10.1109/ICETETS.2016.7603000.

39 Geeitha, S. and Thangamani, M. (2021) 'Integrating HSICBFO and FWSMOTE algorithm-prediction through risk factors in cervical cancer', *J. Ambient Intell. Humaniz. Comput.*, Vol. 12, No. 3, pp.3213–3225.

40 Gopal, V.N., Al-Turjman, F., Kumar, R., Anand L. and Rajesh, M. (2021) 'Feature selection and classification in breast cancer prediction using IoT and machine learning', *Measurement*, p.109442.

41 Nandagopal, V., Geeitha, S., Kumar, K.V. and Anbarasi, J. (2019) 'Feasible analysis of gene expression–a computational based classification for breast cancer', *Measurement*, Vol. 140, pp.120–125.

42 Kishor, A., Chakraborty, C. and Jeberson, W. (2021) 'Intelligent healthcare data segregation using fog computing with internet of things and machine learning', *Int. J. Eng. Syst. Model. Simul.*, doi: 10.1504/IJESMS.2021.10036745.

43 Garg, L., Chukwu, E., Nasser, N., Chakraborty, C. and Garg, G. (2020) 'Anonymity preserving IoT-based COVID-19 and other infectious disease contact tracing model', *IEEE Access*, Vol. 8, pp.159402–159414.

44 Geeitha, S. and Thangamani, M. (2018) 'Incorporating EBO-HSIC with SVM for gene selection associated with cervical cancer classification', *J. Med. Syst.*, Vol. 42, No. 11, pp.1–10.