EAI Endorsed Transactions
on Pervasive Health and Technology Research Article **EAI.EU**

Support Vector Machine based eHealth Cloud System for Diabetes Classification

Chandrashekhar Azad[1,*], Ashok Kumar Mehta[1], Dindayal Mahto[2], Dharmveer Kumar Yadav[3]

[1]Department of Computer Applications, National Institute of Technology Jamshedpur, India
[2]School of Computing, SASTRA Deemed University, Thanjavur, India
[3]Department of Computer Science & Engineering, Katihar Engineering College, Katihar, India

Abstract

INTRODUCTION: Diabetes is a major health issue because it leaves people with physical disabilities. Therefore, methodologies with a reduced error rate must be used to diagnose this dangerous disease. Data Mining techniques such as Artificial Neural Network are common tools adopted for the classification of diabetes and one of the core components of the eHealth system. Data Mining techniques aim to provide reliable and timely diagnostic outcomes during the diagnosis of the disease.

OBJECTIVE: The objective of the research work is to propose a Support Vector Machine based eHealth Cloud System for Diabetes Classification. This work aims to improve the diagnostic accuracy of computer-assisted diagnostic systems.

METHOD: The proposed methodology implemented in two-phase, In the first phase system is trained using different Support Vector Machine (SVM) kernel functions and in the second phase effectiveness of the system is tested in terms of classification accuracy and error. Different SVMs have the ability to diagnose this disease. PIMA Indian Diabetes Dataset (PIDD) has been used in our experiments for training and testing. Kernel functions are usually used to refer to the kernel trick, a method of using a linear classifier to solve a non-linear problem.

RESULT: In this classification accuracy and classification error are used for performance evaluation. It is worth mentioning that the system giving remarkable accuracy of 77.50% in Coarse Gaussian SVM in 10-fold validation whereas fine Gaussian SVM gives 98.8% accuracy in No validation set.

CONCLUSION: This paper introduces the SVM eHealth Cloud System for Diabetes Classification. The system is trained using the PIDD. Such a system can be used as “Application-as-a-Service” in cloud computing. It is therefore believed that the system will enhance the process of clinical decision-making and also assist physicians concerning Diabetes Diagnosis. It is worth to mention that the SVM kernel-based system performed well in comparison to the different systems.

Keywords: Data Mining, eHealth, Medical Mining, PIDD, Support Vector Machine.

Received on 29 November 2019, accepted on 14 May 2020, published on 20 May 2020.

doi: 10.4108/eai.13-7-2018.164627

1. Introduction

People are susceptible to different health issues based on their lifestyle and work behaviours. Some health issues are avoided and reduce it with changing diet, way of living, and atmospheres. Lifestyle health issues describe a certain

[*]Corresponding author. Email: csazad.ca@nitjsr.ac.in

type of disease that is mainly based on people's daily activities and are the product of an unhealthy connection between people and their surroundings. The development of these lifestyle diseases is continuing, it takes ages to grow, and they do not readily lend them to remedy once they are encountered. Bad food practices, lack of exercise, incorrect posture, and an impaired biological clock like are factors that endorse to lifestyle diseases [1]. The resulting chronic diseases -cardiovascular disease, heart attack, heart disease, obesity, and respiratory problems which are long-term illnesses and slow progression will severely affect the earnings of individuals [2]. As per a study published in [3], it tells that lifestyle plays a key role in predisposing to lifestyle diseases, such as diabetes.

Lifestyle diseases, not limited to adults alone, have also begun to hit children. The shift in buying power and the advent of technology has changed the way our lives are now functioning. Less physical activity, more resource access and no time to spare, we were prey to some extremely rare diseases that our grandparents never heard of back in the 60s and 70s. While transmissible diseases such as malaria, cholera, polio can be managed with proper medical care, lifestyle diseases can be avoided if a healthy balanced way of life is followed.

1.1 Diabetes

Diabetes is a lifestyle disease fuelled by high blood sugar levels over an extended period. Symptoms of high blood sugar include excessive urination, enhanced thirst, and excessive hunger. Diabetes can cause many complications due to lack of treatment, complications may be cardiovascular disease, stroke, chronic kidney disease, foot ulcers, and eye damage, etc. [4]. Diabetes is caused either by the pancreas that does not produce enough insulin or by the cells of the body that do not respond properly to the insulin that is produced.

Types of Diabetes:

Three forms of diabetes are [5]:

(i) **Type 1 diabetes:** failure of the pancreas to produce enough insulin due to the loss of beta cells.
(ii) **Type 2 diabetes:** starts with insulin resistance, in which cells fail to respond properly to insulin. A lack of insulin may also occur as the disease progresses.
(iii) **Gestational diabetes:** develop high levels of blood sugar during pregnancy in women. This condition is known as gestational diabetes.

Symptoms of Diabetes:

i. Increased thirst
ii. excessive urination
iii. Extreme appetite
iv. weight loss
v. Fatigue
vi. Tiredness
vii. Blurry vision
viii. Frequent infections such as skin etc.

1.2 Medical Mining

In 2019, electronic health records are frequent in healthcare facilities. With increasing access to a large quantity of patient information, healthcare providers are now tended to focus on maximizing their organization's decision-making process with the help of data mining. Medical databases have collected large amounts of health information and medical conditions. Relationships and trends can provide new medical information within this data. Unfortunately, to discover this hidden knowledge, few methodologies have been developed and applied [6]. Data mining have the capability to find interesting fact from the large corpus data. The pattern extracted from data mining may play an important role in decision-making. In healthcare, data mining is effective in areas such as predictive medicine, relationship management, fraud detection, and the evaluation of the efficacy of some treatments[25, 26, 27].

Medical mining is a broad area of research in which mining methods are used to solve diagnostic and treatment problems as well as to understand the progression of the disease. Medical mining includes learning from hospital records (for diagnostic and treatment decision support), learning from data related to health care, and learning from epidemiological data. In medical mining, classification strategies are widely used to classify data into different classes according to the history of different patients in a domain. Data mining has many applications in various domains like network security, medical mining, cloud computing, etc. [19-21].

2. Related Work

In [7] authors have presented an adaptive evolutionary RBF network algorithm to improve RBF network accuracy. The quality is authenticated using accuracy and is evaluated on three datasets from the UCI database. The results show which approach is an effective means of solving medical disease diagnosis multi-objective RBF network.

In [8] authors have presented a method for classifying diabetes using genetic programming. In this research, numerous approaches used to assess the efficacy of the characteristics of diabetes, to enable the selection of features. The superiority of the method is demonstrated by comparing the outcome with other approaches.

In [9] authors have presented the use of SVM to diagnose diabetes. In this an added interpretation module that turns the SVM's "black box" model into a comprehensible SVM diagnostic decision representation.

In [10] LDA and ANFIS were used to diagnose diabetes. LDA is used to isolate the distinguishing variables healthy and diabetes data, whereas ANFIS is used to classify the results of LDA. The approaches used to deliver the preceding findings with good accuracy.

[11] presented a GA and fuzzy logic-based system for diabetes identification. The proposed system uses a fuzzy-based classification scheme to demonstrate improved classification accuracy.

[12] Presented FCS-ANTMINER, achieved by combining ACO with Fuzzy logic. The outcomes are based on 10-fold cross-validation. The system attained high accuracy against several approaches, which strongly recommends that combining ACO and Fuzzy Logic assists us to accurately identify diabetes.

The goal of [13] is to use ACO to obtain a set of diabetes disease diagnostic guidelines. The program is tested using the PIDD. The outcome demonstrates that the system can detect diabetes with remarkable accuracy and competitiveness.

In this study [14], a new method for classifying medical database information is presented. To determine the applicability of the projected method, real-time problems were examined. The system's performance is good.

3. Methods

3.1 Support Vector Machine (SVM)

SVM is a type of supervised algorithm for Data Mining that provides data analysis for classification and regression analysis. Although regression can be used, SVM is mostly used for classification purposes. The data is plotted in the n-dimensional space, every feature's value is also the unique coordinate's value. Through the learning process, we find the perfect hyperplane that separates the data instances.

3.2 SVM's ideology

SVM is based on the idea of having a hyperplane that better divides features into different domains.

3.3 How does SVM work?

SVM Determine the correct hyperplane

from three hyperplanes A, B, and C identify the correct hyperplane that has the best margin and separates classes. B’s Hyperplane has done this very well in this scenario and is shown in Figure 1.

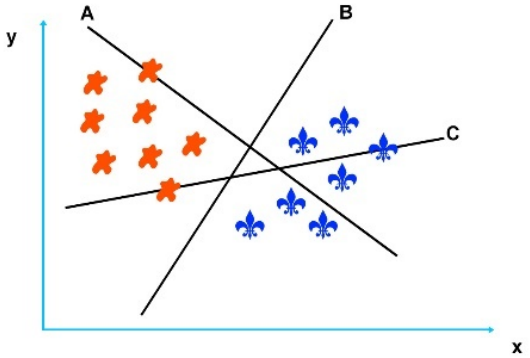


Figure 1. SVM with the different hyperplane

Identify the right hyperplane: Given three hyperplane and all groups are finely separated and is shown in Figure 2.

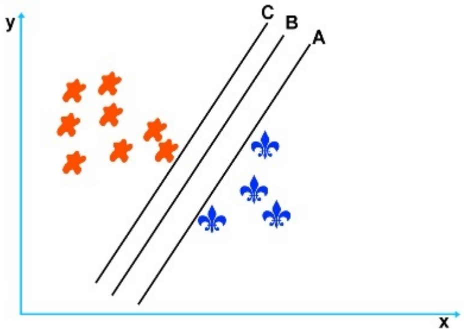


Figure 2. Linearly Separable hyperplanes

In such a situation hyperplane that Maximizes the distances amongst the adjacent data point and this will assist us to find the right hyperplane, the distance is referred to as Margin and is shown in Figure 3.

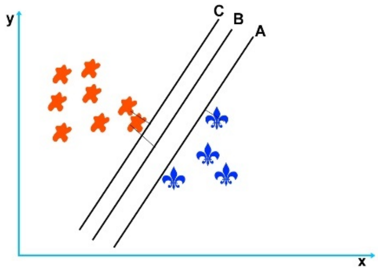


Figure 3. SVM with margins

Here hyperplane C’s margin is better in comparison with hyperplane A and B. Therefore, right hyper-plane is C. Robustness is another important reason to choose the hyper-plane with a higher margin. If we choose a low-margin hyper-plane then there is a high likelihood of misclassification. In Figure 2 and Figure 3, x and y represent the field or attribute of the 2-dimensional dataset.

3.4 Advantages and Disadvantages of Support Vector Machine

Advantages of SVM

- Guaranteed Optimality: Due to Convex Optimization 's existence, the solution will always be a global minimum, not a local minimum.
- Easily access it from either Python or MATLAB for implementation.
- SVM can be used for both linear & non-linear separability. linearly separable data poses a hard margin, even though the non-linearly separable poses the soft margin.
- SVMs comply with semi-supervised models of learning. It can be used as well as unlabeled in areas where the data is labeled.
- SVM can do the function mapping using simple dot product with the help of Kernel Trick.

Disadvantages of SVM

- SVM can't handle categorical data. This leads to sequential information loss and thus leads to worse results.
- Kernel selection may be the greatest limitation of the vector support machine.

3.5 Kernel Function

It takes low-dimensional input space and transforms into a higher-dimension space, in other words transforming the non-separable situation into separable, these functions are called kernels. SVM Kernel projection from 2D to 3D is shown in Figure 4.

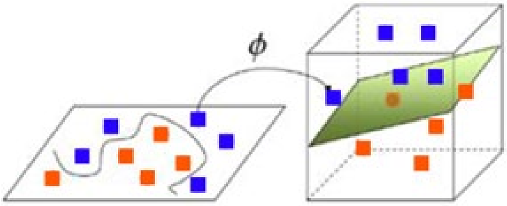
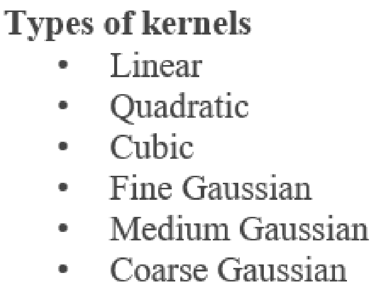


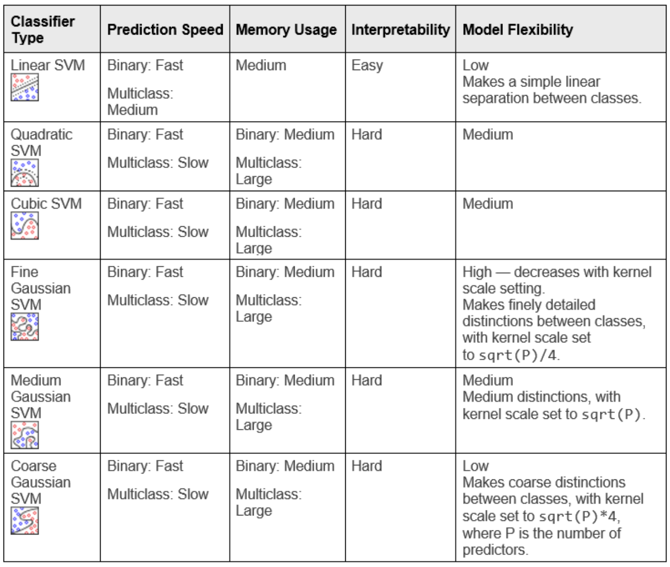
Figure 4. SVM Kernel projection from 2D to 3D



Types of kernels

- Linear
- Quadratic
- Cubic
- Fine Gaussian
- Medium Gaussian
- Coarse Gaussian

Table 1. Comparison of Kernel functions [18]



Classifier Type	Prediction Speed	Memory Usage	Interpretability	Model Flexibility
Linear SVM	Binary: Fast Multiclass: Medium	Medium	Easy	Low Makes a simple linear separation between classes.
Quadratic SVM	Binary: Fast Multiclass: Slow	Binary: Medium Multiclass: Large	Hard	Medium
Cubic SVM	Binary: Fast Multiclass: Slow	Binary: Medium Multiclass: Large	Hard	Medium
Fine Gaussian SVM	Binary: Fast Multiclass: Slow	Binary: Medium Multiclass: Large	Hard	High — decreases with kernel scale setting. Makes finely detailed distinctions between classes, with kernel scale set to `sqrt(P)/4`.
Medium Gaussian SVM	Binary: Fast Multiclass: Slow	Binary: Medium Multiclass: Large	Hard	Medium Medium distinctions, with kernel scale set to `sqrt(P)`.
Coarse Gaussian SVM	Binary: Fast Multiclass: Slow	Binary: Medium Multiclass: Large	Hard	Low Makes coarse distinctions between classes, with kernel scale set to `sqrt(P)*4`, where P is the number of predictors.

Linear Kernel

The simplest kernel is the linear kernel. It is dot product < a, b > plus a constant c.

$$k(a, b) = a^t b + c$$

Polynomial kernel

The polynomial kernel for degree-d is:

$$k(a, b) = (a^t b + c)^d$$

Where a and b are vectors in the state space and d is the degree, c ≥ 0 is a constraint that trade-off the consequence of higher-dimensional vs. lower-dimensional terms in the polynomial. While c = 0, the kernel labeled as homogeneous. Polynomial kernels: linear having degree 1, quadratic having degree 2, and cubic having degree 3.

Gaussian Kernel

Used when the data is not previously known. The formula is the following:

$$k(a, b) = \exp\left(\frac{||a - b||^2}{2\sigma^2}\right)$$

Furthermore, it could also be used as

$$k(a,b) = \exp\left(-\gamma||a-b||^2\right)$$

The variable σ plays a major role in the kernel's efficiency. The exp will act linearly if overestimated, and higher-dimensional estimation will begin to fail its non-linearity. For Fine Gaussian SVM, the scale is$\boldsymbol{sqrt(P)/4}$. For Medium Gaussian SVM, the scale is $\boldsymbol{sqrt(P)}$. The coarse Gaussian SVM scale is $4\boldsymbol{sqrt(P)}$.

Kernel scale mode

When setting the kernel scale mode to Auto, the system will select the scale value using a heuristic procedure. Subsampling is used in the heuristic procedure. Set a random number of seed using rng before training the classifier to reproduce results.

Using a heuristic method, the program should pick the scale value while setting the kernel scale mode to auto. In the heuristic method, sub-sampling is used. Set a random number of seeds using rng to replicate outcomes before supervising the classifier.

4. Dataset

PIDD data originates from Diabetes Diseases. The aim is to determine a person has diabetes or not based on medical diagnosis. The collection of these instances from a wider list has been done with several constraints. In particular, all patients here are women who are at least 21 years old [21].

The dataset includes data from 768 women with 8 characteristics and 1 class attribute, in particular:

i. Pregnancies
ii. Glucose
iii. blood pressure
iv. SkinThickness
v. Insulin
vi. BMI
vii. DiabetesPedigreeFunction
viii. Age
ix. Outcome

Figure 5 shows the Distribution of PIDD Attributes and Figure 10 shows the visualization of data set with each attribute.

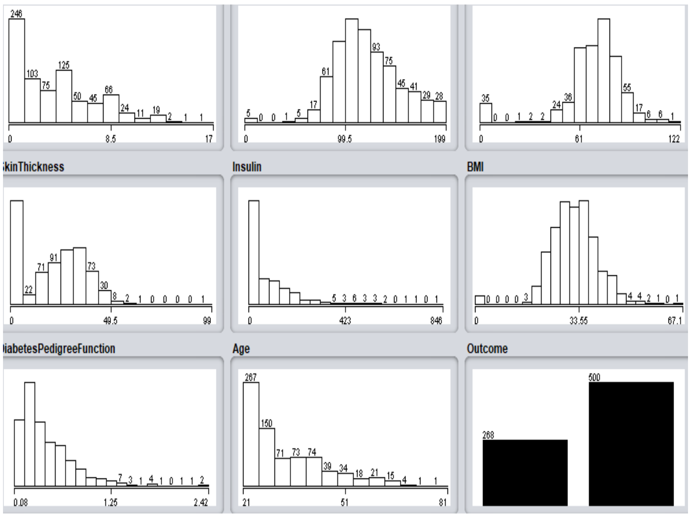


Figure 5. Distribution of PIDD Attributes

Table 2. Dataset Statistics

Attribute	Minimum	Maximum	Mean	Std. Dev.
Pregnancies	0	017	3.845	03.370
Glucose	0	199	120.895	31.973
BloodPressure	0	122	69.105	19.356
SkinThickness	0	099	20.536	15.952
Insulin	0	846	79.799	115.244
BMI	0	67.1	31.993	7.8840
DiabetesPedigreeFunctio	0.078	2.42	0.472	0.3310
Age	21	081	33.241	11.760

5. Proposed System and Results

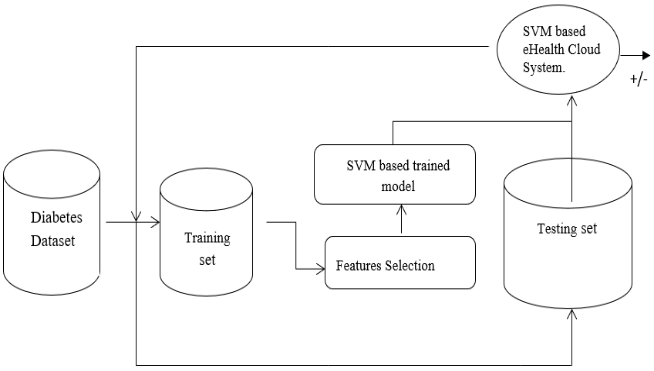


Figure 6. SVM eHealth System

Figure 6 shows the architecture of the proposed expert system which is based on Different kernel functions. The input to the system is diabetes dataset and the final output is whether the person is diabetic (+) or not.

- **Diabetes Dataset:** Input given to the model. Here PIDD is to the system.
- **Training Set:** The part of the PIDD is used to train the prediction model.
- **Testing Set:** The part of the PIDD is used to test the effectiveness of the prediction model.
- **Feature Selection:** Attributes are considered for training the model. In this study, we considered all the features of the PIDD.
- **SVM based Trained Model:** This is the final SVM based eHealth for Diabetes Classification. It may be used in the cloud environment Application-as-a-Services.

SVM eHealth cloud System

As healthcare services cost increases & healthcare specialists are becoming rare and tough to discover, so nowadays it is necessary for healthcare institutions to adopt an IT & data mining-based system. IT & data mining enables health institutions to modernize many of their procedures and more efficiently and cost-effectively deliver services. The new technological developments like cloud computing deliver a cost-effective network & offer an enabler for IT services. This can also be done on the "e-Health Cloud" pay-as-you-use platform to support the healthcare trade deal as per demands while reducing their expenses. Cloud computing is the technology of using the remote server system to store, handle the data, rather than in the local system. The trained SVM eHealth system can be used in the form “Application-as-a-Services”[23, 24] as shown in Figure 7.

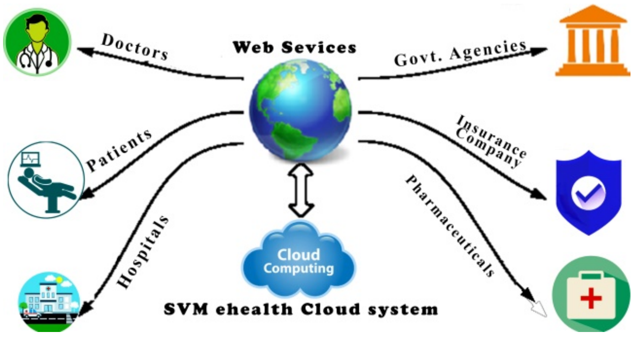


Figure 7. SVM eHealth Cloud System

Steps:

Step 0: Start

Step 1: Load the PIDD

Step 2: SVM eHealth cloud System Model Development using the Training set

Step 3: Testing Model using the validation set

Step 4: Performance analysis

Step 5: Selection of best Model

Step 6: Stop

5.1 Hold Out Validation

Table 3. Hold out validation in 80-20 and 70-30

Hold out validation	80-20		70-30	
Classifier Type	**Accuracy**	**Error**	**Accuracy**	**Error**
Linear SVM	74.50	25.50	77.40	22.60
Quadratic SVM	77.10	22.90	76.40	23.60
Cubic SVM	71.20	28.80	69.60	30.40
Fine Gaussian SVM	64.70	35.30	64.30	35.70
Medium Gaussian SVM	75.80	24.20	76.50	23.50
Coarse Gaussian SVM	74.50	25.50	77.00	23.00

Table 4. Configuration and complexity in 80-20 & 70-30

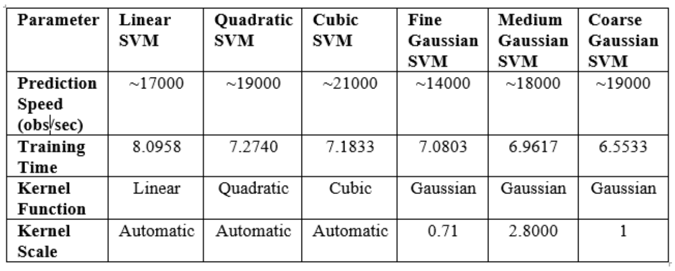
Parameter	Linear SVM	Quadratic SVM	Cubic SVM	Fine Gaussian SVM	Medium Gaussian SVM	Coarse Gaussian SVM
80-20						
Prediction Speed (obs/sec)	~40000	~57000	~53000	~37000	~48000	~45000
Training Time(Sec)	1.2899	0.66973	1.423	0.56618	0.47671	0.40794
Kernel Function	Linear	Quadratic	Cubic	Gaussian	Gaussian	Gaussian
Kernel Scale	Automatic	Automatic	Automatic	0.71	2.8000	1
70-30						
Prediction Speed (obs/sec)	~65000	~77000	~81000	~34000	~36000	~42000
Training Time(Sec)	1.1967	0.68881	1.4065	0.0.59546	0.52496	0.45898
Kernel Function	Linear	Quadratic	Cubic	Gaussian	Gaussian	Gaussian
Kernel Scale	Automatic	Automatic	Automatic	0.71	2.8000	1

5.2 No Validation: **No protection against overfitting.**

Table 5. No Validation

Classifier Type	Accuracy	Error
Linear SVM	77.3	22.70
Quadratic SVM	80.5	19.50
Cubic SVM	87.5	12.50
Fine Gaussian SVM	98.8	01.20
Medium Gaussian SVM	82.7	17.30
Coarse Gaussian SVM	78.4	21.60

Table 6. Configuration &complexity in No Validation



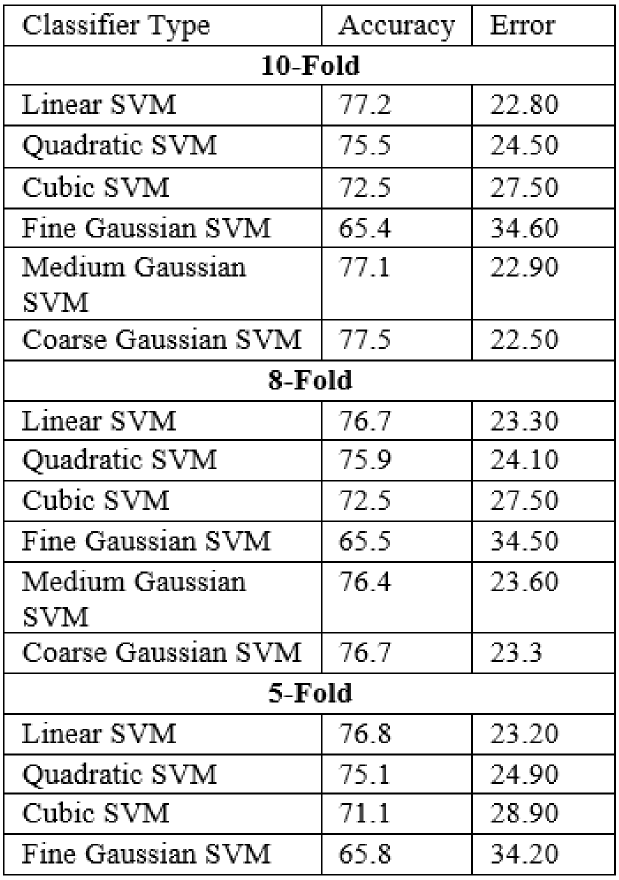
Parameter	Linear SVM	Quadratic SVM	Cubic SVM	Fine Gaussian SVM	Medium Gaussian SVM	Coarse Gaussian SVM
Prediction Speed (obs/sec)	~17000	~19000	~21000	~14000	~18000	~19000
Training Time	8.0958	7.2740	7.1833	7.0803	6.9617	6.5533
Kernel Function	Linear	Quadratic	Cubic	Gaussian	Gaussian	Gaussian
Kernel Scale	Automatic	Automatic	Automatic	0.71	2.8000	1

5.3 K-Fold Validation: **Defend in contrast to overfitting by dividing the data into the different folds and evaluating accuracy on each fold.**

The process is as follows:

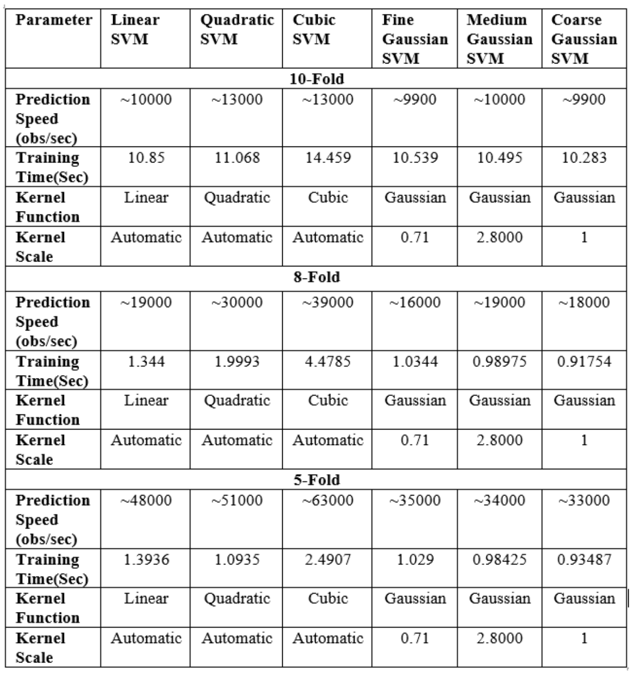
- Randomly shuffle the set of data.
- Divide the data set into K folds.
- repeat K-times:
    - Take 1-fold as a testing set.
    - Take the remaining fold as a training set
    - Train a model on the training set and assess it on the testing set.
    - Retain the performance score and discard the model
- Summarize the model's performance using the list of model's performance scores.

Table 7. K-Fold validation



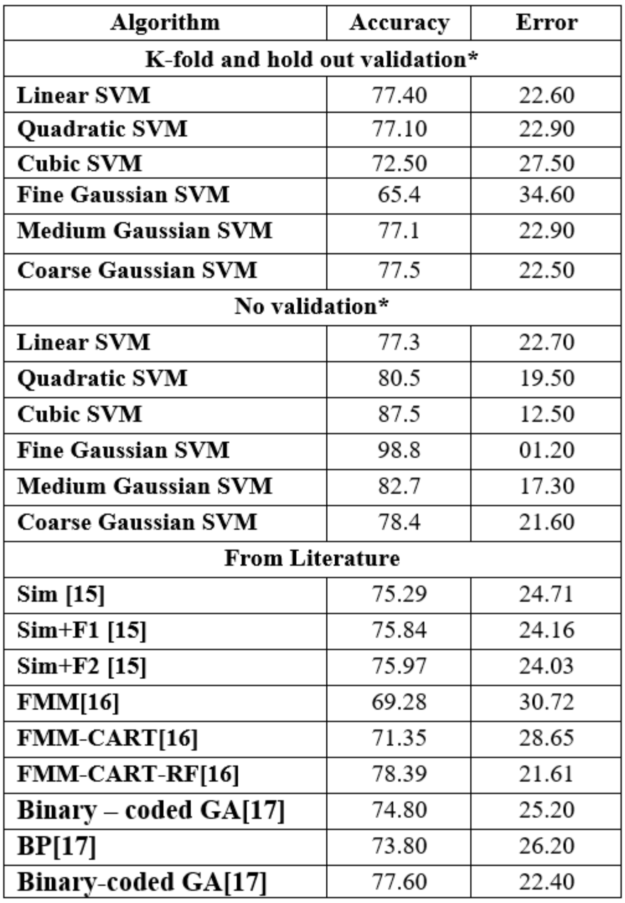
Classifier Type	Accuracy	Error
10-Fold		
Linear SVM	77.2	22.80
Quadratic SVM	75.5	24.50
Cubic SVM	72.5	27.50
Fine Gaussian SVM	65.4	34.60
Medium Gaussian SVM	77.1	22.90
Coarse Gaussian SVM	77.5	22.50
8-Fold		
Linear SVM	76.7	23.30
Quadratic SVM	75.9	24.10
Cubic SVM	72.5	27.50
Fine Gaussian SVM	65.5	34.50
Medium Gaussian SVM	76.4	23.60
Coarse Gaussian SVM	76.7	23.3
5-Fold		
Linear SVM	76.8	23.20
Quadratic SVM	75.1	24.90
Cubic SVM	71.1	28.90
Fine Gaussian SVM	65.8	34.20

Table 8. Configuration &complexity in K-fold



Parameter	Linear SVM	Quadratic SVM	Cubic SVM	Fine Gaussian SVM	Medium Gaussian SVM	Coarse Gaussian SVM
10-Fold						
Prediction Speed (obs/sec)	~10000	~13000	~13000	~9900	~10000	~9900
Training Time(Sec)	10.85	11.068	14.459	10.539	10.495	10.283
Kernel Function	Linear	Quadratic	Cubic	Gaussian	Gaussian	Gaussian
Kernel Scale	Automatic	Automatic	Automatic	0.71	2.8000	1
8-Fold						
Prediction Speed (obs/sec)	~19000	~30000	~39000	~16000	~19000	~18000
Training Time(Sec)	1.344	1.9993	4.4785	1.0344	0.98975	0.91754
Kernel Function	Linear	Quadratic	Cubic	Gaussian	Gaussian	Gaussian
Kernel Scale	Automatic	Automatic	Automatic	0.71	2.8000	1
5-Fold						
Prediction Speed (obs/sec)	~48000	~51000	~63000	~35000	~34000	~33000
Training Time(Sec)	1.3936	1.0935	2.4907	1.029	0.98425	0.93487
Kernel Function	Linear	Quadratic	Cubic	Gaussian	Gaussian	Gaussian
Kernel Scale	Automatic	Automatic	Automatic	0.71	2.8000	1

Table 9. Performance Comparison



Algorithm	Accuracy	Error
K-fold and hold out validation*		
Linear SVM	77.40	22.60
Quadratic SVM	77.10	22.90
Cubic SVM	72.50	27.50
Fine Gaussian SVM	65.4	34.60
Medium Gaussian SVM	77.1	22.90
Coarse Gaussian SVM	77.5	22.50
No validation*		
Linear SVM	77.3	22.70
Quadratic SVM	80.5	19.50
Cubic SVM	87.5	12.50
Fine Gaussian SVM	98.8	01.20
Medium Gaussian SVM	82.7	17.30
Coarse Gaussian SVM	78.4	21.60
From Literature		
Sim [15]	75.29	24.71
Sim+F1 [15]	75.84	24.16
Sim+F2 [15]	75.97	24.03
FMM[16]	69.28	30.72
FMM-CART[16]	71.35	28.65
FMM-CART-RF[16]	78.39	21.61
Binary – coded GA[17]	74.80	25.20
BP[17]	73.80	26.20
Binary-coded GA[17]	77.60	22.40

*Our Trained System

5.4 Discussion

Our SVM Kernel-based expert system for Diabetes classification is simulated in MATLAB, Windows 10, 8 GB RAM, Intel Core i7-8700 CPU @3.20 GHz processor. The proposed system is trained using different SVM Kernel functions using PIDD. The system is trained with hold out validation, no validation, and different k-fold validation methods and performance and recorded in table 3-9. Table 1 shows the comparison of different kernel functions, table 2 shows the statistics of the dataset. Table 4,6, and 8 shows the configuration and complexity like prediction speed, Training Time, Kernel scale, Kernel function. Table 3 shows the performance of various kernel functions with hold out validation (80-20 and 70-30). Table 5 shows the performance of various kernel functions with No validation. Table 7 shows the performance of various kernel functions with K-fold cross-validation (K=10,8 5). In no validation set the best performance is observed on fine gaussian SVM and the values of various performance measures such as accuracy and error are 98.8 % and 01.20 respectively.

In Holdout validation set the best performance is observed on Liner SVM and the values of various performance measures such as accuracy and error are 77.40% and 22.60% respectively.

In K-Fold validation best performance is observed on Coarse Gaussian SVM in 10 fold validation and the values of various performance measures such as accuracy and error are 77.50% and 22.50% respectively.

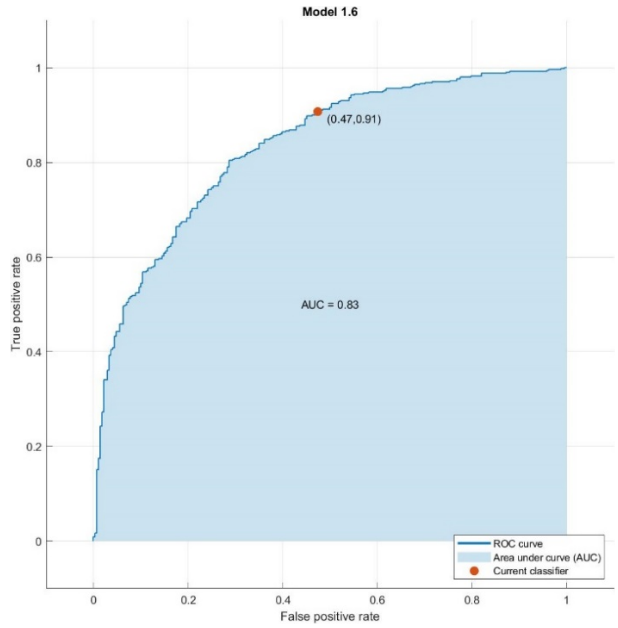


Figure 8. ROC of Coarse gaussian SVM

Figure 8. shows the ROC of Coarse Gaussian SVM, it is observed that the AUC is 0.83 in 10-fold validation.

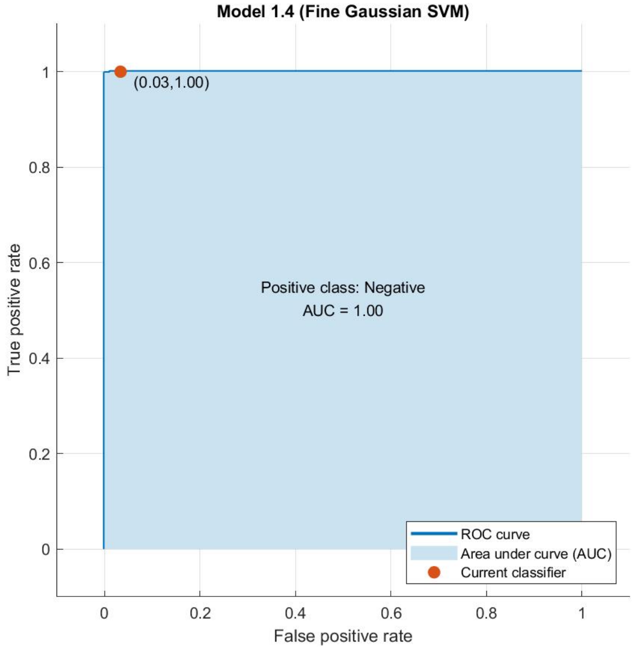


Figure 9. ROC of Fine gaussian SVM

Figure 9. shows the ROC of Fine Gaussian SVM, it is observed that the AUC is 1.00 in no validation, in this case, accuracy is very high but the system may have a problem of overfitting.

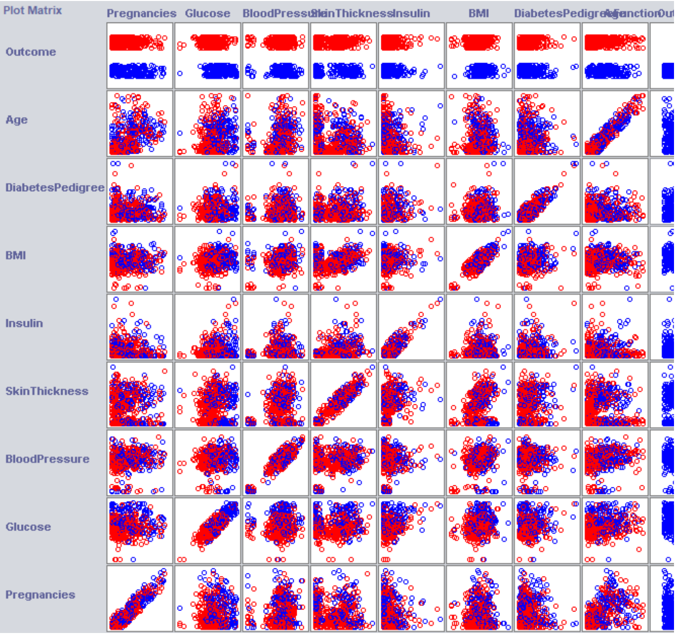


Figure 10. Visualization of Dataset

6. Conclusion

ICT advances lead to the use of data mining techniques in different fields, including medical sciences. By using data mining techniques, we can design and implement complex medical processes. In this, we proposed the SVM eHealth Cloud system for diabetes prediction. The system is trained using the PIDD. Such a system can be used as "Application-as-a-Service" in cloud computing, also, it is useful for various medical science fields such as diagnosis and assisting surgeons, doctors, etc. It is worth to mention that the system giving remarkable accuracy 77.50% by coarse gaussian SVM in 10-fold validation and fine gaussian SVM gives 98.8% accuracy in No validation set. In the future, we will use some optimization techniques like GA, PSO, ACO, etc. along with the machine learning algorithm. It is worth to mention that such a machine learning-based eHealth system may be used for the early prediction of diabetes. Such a system enables health institutions to modernize many of their procedures and more efficiently and cost-effectively deliver services.

References

[1] Sharma M, Majumdar PK. Occupational lifestyle diseases: An emerging issue. Indian journal of occupational and environmental medicine. 2009 Dec;13(3):109.

[2] Sobra J, Ceska R. Diseases of civilization from the aspect of the evolution of the human diet. Casopis learn ceskych. 1992 Apr;131(7):193-7.

[3] Key TJ, Allen NE, Spencer EA. The effect of diet on risk of cancer. Lancet. 2002;360:861–8.

[4] Diabetes,https://www.who.int/en/news-room/fact-sheets/detail/diabetes, Accessed on 27/11/2019.

[5] Tripathy BB, Chandalia HB, Das AK. RSSDI textbook of diabetes mellitus. JP Medical Ltd; 2012 Jan 15.

[6] Prather JC, Lobach DF, Goodwin LK, Hales JW, Hage ML, Hammond WE. Medical data mining: knowledge discovery in a clinical data warehouse. InProceedings of the AMIA annual fall symposium 1997 (p. 101). American Medical Informatics Association.

[7] Qasem SN, Shamsuddin SM. Radial basis function network based on time variant multi-objective particle swarm optimization for medical disease diagnosis. Applied Soft Computing. 2011 Jan 1;11(1):1427-38.

[8] Aslam MW, Zhu Z, Nandi AK. Feature generation using genetic programming with comparative partner selection for diabetes classification. Expert Systems with Applications. 2013 Oct 1;40(13):5402-12.

[9] Barakat N, Bradley AP, Barakat MN. Intelligible support vector machines for diagnosis of diabetes mellitus. IEEE transactions on information technology in biomedicine. 2010 Jan 12;14(4):1114-20.

[10] Dogantekin E, Dogantekin A, Avci D, Avci L. An intelligent diagnosis system for diabetes on linear discriminant analysis and adaptive network based fuzzy inference system: LDA-ANFIS. Digital Signal Processing. 2010 Jul 1;20(4):1248-55.

[11] Ephzibah EP. Cost effective approach on feature selection using genetic algorithms and fuzzy logic for diabetes diagnosis. arXiv preprint arXiv:1103.0087. 2011 Mar 1.

[12] Ganji MF, Abadeh MS. A fuzzy classification system based on Ant Colony Optimization for diabetes disease diagnosis. Expert Systems with Applications. 2011 Nov 1;38(12):14650-9.

[13] Ganji MF, Abadeh MS. Using fuzzy ant colony optimization for diagnosis of diabetes disease. In2010 18th Iranian Conference on Electrical Engineering 2010 May 11 (pp. 501-505). IEEE.

[14] Kahramanli H, Allahverdi N. Design of a hybrid system for the diabetes and heart diseases. Expert systems with applications. 2008 Jul 1;35(1-2):82-9.

[15] Luukka P. Feature selection using fuzzy entropy measures with similarity classifier. Expert Systems with Applications. 2011 Apr 1;38(4):4600-7.

[16] Seera M, Lim CP. A hybrid intelligent system for medical data classification. Expert Systems with Applications. 2014 Apr 1;41(5):2239-49.

[17] Örkcü HH, Bal H. Comparing performances of backpropagation and genetic algorithms in the data classification. Expert systems with applications. 2011 Apr 1;38(4):3703-9.

[18] https://in.mathworks.com/help/stats/choose-a-classifier.html

[19] Azad C, Jain S, Jha VK. Design and Analysis of Data Mining Based Prediction Model forParkinson's disease. Issues.;1(1):181-9.

[20] Azad C., Jha VK. Fuzzy min–max neural network and particle swarm optimization based intrusion detection system. Microsystem Technologies. 2017 Apr 1;23(4):907-18.

[21] PIDD: PIMA Indian Diabetes Dataset, https://www.kaggle.com/uciml/pima-indians-diabetes-database

[22] Sisodia D, Sisodia DS. Prediction of diabetes using classification algorithms. Procedia computer science. 2018 Jan 1;132:1578-85.

[23] Jangiti S, VS SS. Scalable and direct vector bin-packing heuristic based on residual resource ratios for virtual machine placement in cloud data centers. Computers & Electrical Engineering. 2018 May 1;68:44-61.

[24] Jangiti S, Ram ES, Sriram VS. Aggregated Rank in First-Fit-Decreasing for Green Cloud Computing. In Cognitive Informatics and Soft Computing 2019 (pp. 545-555). Springer, Singapore.

[25] Azad C, V. K. Jha. Fuzzy min–max neural network and particle swarm optimization based intrusion detection system. Microsystem Technologies. 2017 Apr 1;23(4):907-18.

[26] Azad C., V. K. Jha. Data mining in intrusion detection: a comparative study of methods, types and data sets. International Journal of Information Technology and Computer Science (IJITCS). 2013 Jul 1;5(8):75-90.

[27] Azad C., V. K. Jha. A novel fuzzy min-max neural network and genetic algorithm-based intrusion detection system. InProceedings of the Second International Conference on Computer and Communication Technologies 2016 (pp. 429-439). Springer, New Delhi.