DETECTING DISEASES OF VARIANT NATURE IN HUMANS BY ENHANCED ALGORITHM USING SVM

Chitralekha Dwivedi[1], Priyanshika[2], Srishti Kamble[3], Saurav Nikum[4], Mrudula Avhad[5]

Associate Professor, Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pune, India[1]

Student, Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pune, India[2,3,4,5]

Abstract: The healthcare organisation creates a massive amount of patient data, which may be analysed in a variety of ways. As a result, with the assistance of a machine learning, we developed a prediction system that can identify many diseases at the same time. We have focused on various diseases: heart disease, liver disease, diabetes, etc however many more diseases may be included in the future. The user must enter numerous illness parameters, and the system will determine whether the person has the diseases or not. Support vector machines with adaptivity were utilised to identify numerous illnesses. The goal was to offer an adaptive SVM-based diagnostic technique that was automated, rapid, and versatile. To improve outcomes, the bias value in traditional SVM was changed. The suggested classifier produced 'if-then' rules. Using the recommended technique, several diseases were detected, as well as increased categorization rates. The key emphasis of future research should be the development of more effective ways for changing the bias value in classical SVM.

Keywords: Diseases Prediction System, Supervised Machine Learning, Classification, Prediction, Support Vector Machine, Health Care Analysis.

I. INTRODUCTION

In the era of digital advancements, data has become a crucial resource across various industries, including healthcare. The healthcare sector generates a vast amount of patient-related data, making it essential to develop robust architectures for illness prediction. While existing models often focus on analysing individual diseases, there is a lack of systems that can simultaneously analyse multiple illnesses. To address this gap, we have concentrated our efforts on providing users with fast and accurate illness predictions based on the symptoms they provide. Our system specifically focuses on the prediction of Diabetes, Heart Disease, Breast Cancer, Liver Disease, and Kidney Disease assessments, with the potential for adding more ailments in the future. By enabling the prediction of multiple diseases simultaneously, users no longer need to visit multiple websites for illness forecasting. Machine learning plays a vital role in our system, as we utilize it to implement various disease analyses. When accessing our API, users are required to specify the name of the ailment along with its corresponding parameters. Our model will then invoke the appropriate algorithms and return the predicted state of the patient. For accurate illness outcomes, we employ the Support Vector Machine (SVM) approach within our machine learning framework. One notable aspect of our system analysis is its comprehensive consideration of all factors contributing to the development of the diseases under investigation. This holistic approach allows for more effective and precise detection. To ensure the model's availability and efficiency, we save its behavior as a Python pickle file, facilitating seamless integration into our architecture. By providing users with a unified platform for multi-illness prediction, we aim to streamline the process and enhance the overall user experience. Our innovative approach not only saves time but also improves the accuracy of predictions. With the continuous advancements in machine learning and the potential for incorporating additional diseases, our system holds great promise in revolutionizing the field of illness prediction in healthcare. Through our unique architecture and commitment to staying at the forefront of technological advancements, we are dedicated to providing healthcare professionals and patients with a reliable and efficient tool for early detection and improved healthcare outcomes. Join us on this transformative journey as we leverage the power of data and machine learning to revolutionize illness prediction in the healthcare sector.

II. LITERATURE SURVEY

Khurana, Sarthak et al[1] has employed Decision tree, Random forest, and Nave Bayes algorithms to forecast a disease based on systems and to enable synchronised and well-versed medical systems that provide optimum patient satisfaction. Meghna Kamboj[2] used LR,NB,KNN,SVM,DT, and RF algorithms for prediction of heart disease with adequate data processing and implementation of ML algorithm with varied parameters, and among all Machine Learning algorithms,

the KNN has the greatest accuracy of 87%. Ware, Miss.Sangya et al[3] have examined multiple Machine Learning models using performance metrics and to predict heart related disorders with the greatest accuracy of 89.34% using SVM. Shirsath et al[4] has introduced a CNN-MDRP technique that integrates structured and unstructured data and shown that CNN-MDRP is more accurate than earlier prediction algorithms. Marimuthu et al[5] has employed several types of DM and ML approaches to forecast the occurrence of heart disease and apply the suggested system to the region that was required. Battineni et al[6] has the study models linked with chronic illness diagnosis were evaluated in the work These models are quite useful in the categorization and diagnosis of CD. Ardabili et al[7] has used MLP and ANBEIS and offered a comparison analysis of ML and soft models to forecast the COVID-19 epidemic, as well as preliminary benchmarking to illustrate the potential of ML in the future. Shrestha et al[8] has developed a heart disease prediction system that achieves 88.163% accuracy, among other things. Magar et al[9] has proposed a robust model for predicting heart illness and discovered that the Logistic Regression approach was the most efficient, with an accuracy of 82.89%, followed by DT and NB, both with 80.40%, and SVM, with 81.75%. Alotaibi and Fahd. Saleh [10] "Implementation of Machine Learning Model to Predict Heart Failure Disease" investigated, recommended, and implemented a Machine Learning model in which the Rapid Miner tool was employed, which computed a higher degree of accuracy than Matlab and Weka. Godse et al[11] has proposes a multiple illness prediction system that delivers disease prediction medicine and medication consulting. Anitha et al[12] has employed KNN, Nave Bayes, and SVM algorithms and compared their performance using a heart disease dataset, with Nave Bayes achieving the greatest accuracy of 86.6%. Bindhika et al[13] has provided a method for heart disease prediction using machine learning techniques, and the findings demonstrated high accuracy criteria for delivering better projected outcomes. Pingale et al[14]has uses KNN, Nave Bayes, Logistic Regression, and Decision Tree algorithms to create a disease prediction system that can predict the disease based on symptoms and was implemented using the grails framework. Chauhan Raj H. et al[15] has uses the Nave Bayes, Decision Tree, and Random Forest algorithms to develop a disease prediction system with higher accuracy, as well as motivational ideas and visuals. . Ananey-Obiri et al[16] in contrast, some researchers have eliminated missing values. Based on DT, LR, and Gaussian NB algorithms, the features are reduced from 13 to 4 using feature selection method and reported an accuracy of 82.75%. Mohan et al. [17] has used a hybrid random forest (RF) with the linear model and improved the accuracy of 297 records and 13 characteristics of the Cleveland dataset for heart disease prediction. Kodati et al. [18] tested several types of classifiers using Orange and Weka data-mining tools to predict heart disease with 297 records and 13 features. In addition, the feature selection method plays an important role in improving the accuracy of the model. Shah et al. [19] utilized probabilistic principal component analysis (PCA). R. Perumal et al. [20] has develop LR and support vector machine (SVM) models with similar accuracy levels (87% and 85%, respectively). To train the ML classifiers, they used a dataset of 303 data instances and standardized and reduced features using PCA.

III. PROPOSED SYSTEM

In this study, we used structured and unstructured data from the healthcare domains to assess illness risk. The use of a latent component model to the reconstruction of missing data in medical records collected from internet sources. Using statistical data, we might potentially examine the primary chronic illnesses in a certain location and population. In the case of unstructured text files, we employ the Support vector method to choose features automatically.

A. Inputs (Patient Symptoms): When designing the algorithm, we anticipated that the customer would have a clear idea of the symptoms he is experiencing. The generated prediction takes into account numerous manifestations, in the middle of which the customer might let the signals his preparation as the input.

B. Data pre-processing: Data pre-processing is the mining of data techniques that transform the raw information or then encrypts the information to produce a structure so that it may be efficiently decoded with the use of computation.

C. Model: The complete system is built to forecast illnesses using the support vector machine technique, so that the predictive analysis can be investigated quickly and efficiently for the input.

D. Output(diseases): While building a framework with the preparation set and using validated calculations, standard datasets are formed, and whenever the client indications are provided as a contribution as an input of the algorithm, and the side effects are composed in accordance with the standard dataset created, thus creating arrangements, and anticipating the high probability infection.
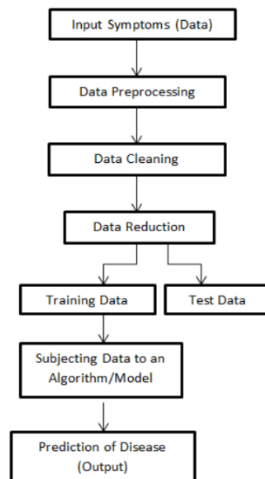
Fig 1: Methodology Flowchart

IV. IMPLEMENTATION

1. ALGORITHM AND RELATED MATHEMATICS

SVM (Support Vector Machine) Algorithm:

SVM is a supervised approach. This method is used in both classification and regression research. This approach uses coordinates to plot data in n-dimensional spaces, which might be linear or nonlinear. Because we are working with linearly separable data, we use the linear SVM classifier in our research. Classification is accomplished by locating the best hyperplane, which acts as the border that divides classes into categories. The line is a hyperplane in two dimensions. The line is adequate to separate the classes in two-dimensional space.

For example, consider the equation S0+(S1*U1) +(S2*U2) =0.

The coefficients are B0 and B1, and the line's intercept is B2. K1 and K2 are input points. This line is used in categorization. The equation yields a number larger than zero over the line, and the resulting data value belongs to category "0." The data point is classified as "1" if it is below the line and the value returned by the equation is less than zero. A point with a value close to zero is difficult to categorise. The margin is defined as the distance between the line and the nearest data point. The ideal line can divide the classes if it has the greatest advantage. This line represents the maximum margin hyperplane. The perpendicular distance between the line and the nearest highlight is utilised to calculate this margin. The data values at these points are known as support vectors, and they are critical for both characterising the line and developing the classifier. Support vectors are used to define and support hyperplanes.
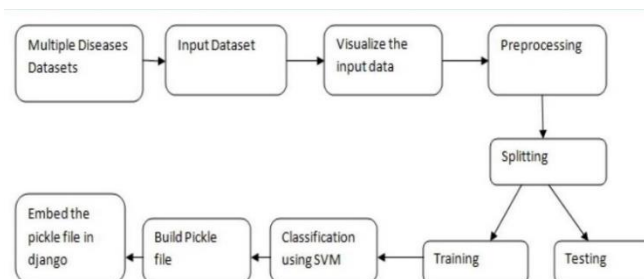

Fig 2: Block Diagram

Pseudo code of SVM:

The formula for SVM can be broken down into the following steps:

1) Given a collection of training data points (x1, y1), (x2, y2),..., (xn, yn), where xi represents the feature vector and yi represents the class label (-1 or 1), do the following algorithm.

2) The SVM method determines a hyperplane that maximises the difference between the two classes.

3) The hyperplane is defined by the equation w.x + b = 0, where w is the hyperplane's normal vector and b is the bias factor.

4) The margin is defined as the distance between the hyperplane and the nearest data points in each class.

5) The optimisation problem for locating the hyperplane may be written as follows:

6) Reduce (1/2) * ||w||2 for every I if yi(w.xi + b) >= 1. This is a linear constraint quadratic optimisation issue.

7) After locating the hyperplane, fresh data points may be categorised by computing w.x + b and examining the sign of the result.

By utilising kernel functions, the SVM algorithm may be expanded to accommodate non-linearly separable data. Transform the data into a higher-dimensional space where the classes may be separated by a linear hyperplane.

Inputs:
- Training dataset: X_train, y_train (feature vectors and corresponding labels)
- Testing dataset: X_test (feature vectors)

Procedure:

1. Initialize SVM parameters:
   - Set the number of iterations: max_iter
   - Set the learning rate: learning_rate
   - Initialize the weights: w = [0, 0, ..., 0]
   - Initialize the bias term: b = 0

2. Define the SVM training function:
   - For each iteration i in range(max_iter):
   - For each training example (X, y) in zip(X_train, y_train):
   - Compute the activation value: a = dot_product(w, X) + b
   - Compute the loss: loss = max(0, 1 - y * a)
   - Update the weights and bias:
     - If loss > 0:
       - Update the weights: w = w - learning_rate * (2 * regularization_parameter * w - y * X)
       - Update the bias: b = b - learning_rate * (-y)
   - End for

3. Train the SVM model:
   - Call the SVM training function with the training dataset.

4. Define the SVM prediction function:
   - For each test example X in X_test:
   - Compute the activation value: a = dot_product(w, X) + b
   - Assign the predicted label:
     - If a >= 0:
       - Predicted label = 1
     - Else:
       - Predicted label = -1

5. Predict the labels for the test dataset:
   - Call the SVM prediction function with the testing dataset.

6. Evaluate the performance of the SVM model:
   - Compare the predicted labels with the true labels and compute evaluation metrics such as accuracy, precision, recall, and F1-score.

7. Output the evaluation results and conclusions.

2. USER INTERFACE

Many assessments of the present processes in the health care business only considered one sickness at a time. Several systems, for example, are used to investigate diabetes, analyse diabetic retinopathy, and predict heart disease. Most methods focus on a single condition. When analysing patient health information, an organisation must employ a number of models. The existing system's strategy is only suitable for examining particular illnesses. Using the many illnesses prediction approach, a user may analyse various diseases on a single webpage. The user does not need to visit many sites to ascertain whether or not they have a specific condition. In the many diseases prediction system, the user must select the name of the specific disease, provide any necessary information, and then simply click the submit button. When you invoke the proper machine learning model, it will anticipate the outcome and present it on the screen.

Home   Heart   Diabetes   Breast Cancer   Kidney   Liver

Multiple Disease Predictor

Information about the Diseases:

Diabetes

Diabetes mellitus refers to a group of diseases that affect how your body uses blood sugar (glucose). Glucose is vital to your health because it's an important source of energy for the cells that make up your muscles and tissues. It's also your brain's main source of fuel. The underlying cause of diabetes varies by type. But, no matter what type of diabetes you have, it can lead to excess sugar in your blood. Too much sugar in your blood can lead to serious health problems.

Symptoms

- Increased thirst
- Frequent urination
- Extreme hunger
- Unexplained weight loss
- Blurred vision

Home   Heart   Diabetes   Breast Cancer   Kidney   Liver

Heart/ Cardiovascular Disease

Cardiovascular disease (heart disease) refers to a group of diseases that affect the heart and blood vessels of your body. These diseases can affect one or many parts of your heart and /or blood vessels. A person may be symptomatic (physically experience the disease) or be asymptomatic (not feel anything at all).

Symptoms

- Pounding or racing heart (palpitations)
- Chest pain
- Sweating
- Lightheadedness
- Shortness of breath

Breast Cancer

Breast cancer is a type of cancer that starts in the breast. It can start in one or both breasts. After skin cancer, breast cancer is the most common cancer diagnosed in women in the United States. Breast cancer can occur in both men and women, but it's far more common in women.

Symptoms

- New lump in the breast or underarm (armpit)
- Thickening or swelling of part of the breast.
- Irritation or dimpling of breast skin
- Redness or flaky skin in the nipple area or the breast

Home   Heart   Diabetes   Breast Cancer   Kidney   Liver

Chronic Kidney Disease

Chronic kidney disease (CKD) means your kidneys are damaged and can't filter blood the way they should. The disease is called "chronic" because the damage to your kidneys happens slowly over a long period of time. This damage can cause wastes to build up in your body. CKD can also cause other health problems.

Symptoms

- Nausea
- Vomiting
- Fatigue and weakness
- Muscle twitches and cramps
- Loss of appetite

Liver Disease

The term "liver disease" refers to any of several conditions that can affect and damage your liver. Over time, liver disease can cause cirrhosis (scarring). As more scar tissue replaces healthy liver tissue, the liver can no longer function properly. Left untreated, liver disease can lead to liver failure and liver cancer.

Symptoms

- Abdominal (belly) pain (especially on the right side)
- Bruising easily
- Changes in the color of your urine or stool

Home Heart Diabetes Breast Cancer Kidney Liver

Diabetes Predictor

No. of Pregnancies

Glucose

Blood Pressure (in mmHg)

Skin Thickness (in mm)

Insulin (in µU/ml)

BMI

Diabetes Pedigree Function

Age (in years)

Predict

Home Heart Diabetes Breast Cancer Kidney Liver

Diabetes Predictor

1

85

66

29

0

26.6

0.351

31

Predict

Home Heart Diabetes Breast Cancer Kidney Liver

POSITIVE. The patient might have the disease.

Back to Home

V. RESULTS

SVM algorithms gave the best levels of accuracy, they were used in the system's diabetes illness prediction model. When the patient enters the disease-specific parameter, the system determines whether or not the patient has the condition in issue. The parameters will indicate the required value range, and if the value is outside of that range, invalid, or empty, a warning message will appear urging the user to provide a valid value.

Model	Accuracy
Diabetes Model	98.25%
Breast Cancer Model	98.25%
Heart Disease Model	85.25%
Kidney Disease Model	99%
Liver Disease Model	78%

Table 1: displays the degree of accuracy attained with svm for each disease.
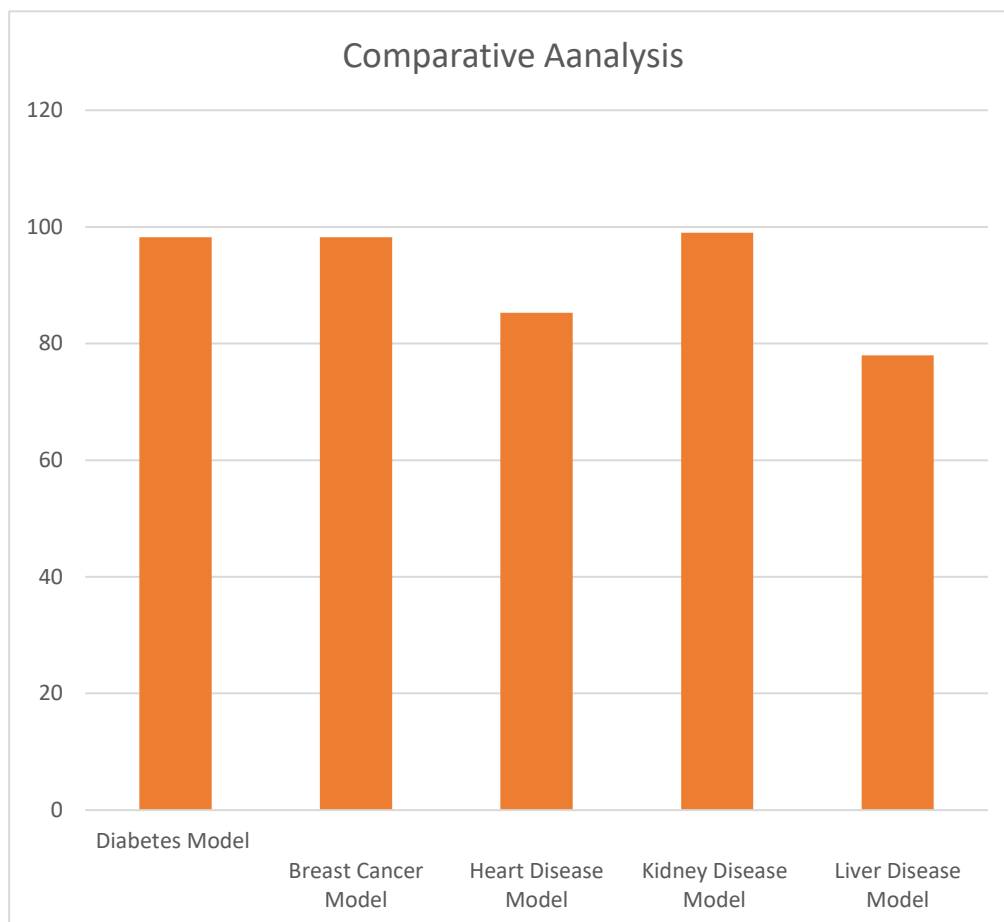

Fig 1: comparative analysis of each disease

VI. CONCLUSIONS

The trained models are deployed with the help of the lightweight Django API framework, which brings diabetes, heart disease, breast cancer, and liver disease together on a single platform. A classification method is used to train the model. The fundamental purpose of this research was to create a system capable of reliably predicting numerous illnesses. Because of this project, the user does not have to travel across many websites, which saves time. Early illness detection can help you live longer and avoid financial hardship. In order to maximise accuracy, we used the svm machine learning method to achieve this aim. In the future, we may broaden our study to include new diseases learnt by machine learning models, as well as diseases that employ deep learning models.

VII. REFERENCES

[1] Khurana, Sarthak., Jain, Atishay., Kataria, Shikhar., Bhasin, Kunal., Arora, Sunny., & Gupta, Dr. Akhilesh. Das. "Disease Prediction System. International Research Journal of Engineering and Technology ", IRJET, Vol 6 Issue 5, pp. 5178-5184, 2019.

[2] Kamboj Megha "Heart Disease Prediction with Machine Learning Approaches", IJSR, Vol 9 Issue 7 ,pp. 1454-1458,2020.

[3]Ware,Miss.Sangya . , Rakesh,Mrs.Shanu. K.,&Choudhary,Mr.Bharat "Heart Attack Prediction By Using Machine Learning Techniques", IJRTE , Vol 8 Issue 5 ,pp.1577- 1580,2020.

[4] Shirsath ,Shraddha.Subhash .,& Patil , Prof. Shubhangi "Disease Prediction Using Machine Learning over Big Data ",IJIRSET, Vol 7 Issue 6,pp. 6752-6757,2018.

[5]Marimuthu , M. , Abinaya, M. ,Hariesh,K.S., Madhan,K.,& Pavithra, Kumar. V "A Review of Heart Disease Prediction Using Machine Learning and Data Analytics Approach" , IJCA, Vol 181 Issue18 , pp. 20-25, 2018.

[6] Battineni ,Gopi. , Sagaro,Getu.Gamo. ,Chinatalapudi, Nalini. ,&Amenta,Francesco. "Application Of Machine Learning Predictive Models in the Chronic Disease", MDPI,Vol 10 Issue 21, pp. 1- 11 , 2020.

[7]Ardabili ,Sina. F.,Mosavi ,Amir.,Khamosi, Pedram. , Ferdinand ,Filip. ,Varkonyi-Koczy. Annamaria.R. Reuter, Uwe. ,Rabczuk ,Timon. , & Atkinson,Peter M. "COVID -19 Outbreak Prediction with Machine Learning.Journal of Algorithms", STCR , Vol 13 Issue 249 , 1-36 , 2020

[8]Shrestha,Ranjit.,& Chatterjee,Jyotir. Moy.".Heart Disease Prediction System Using Machine Learning " LBEF Research Journal of Science Technology and Management , Vol 1 Issue 2, pp. 115-132 , 2019.

[9] Magar ,Rishabh. , Memane,Rohan.,Raut ,Sura. , &Rupnar,Prof. V.S. "Heart Disease Prediction Using Machine Learning ", JETIR , Vol 7 Issue 6 ,pp. 2081-2085 , 2020.

[10] Alotaibi, Fahd. Saleh."Implementation of Machine Learning Model to Predict Heart Failure Disease", International Journal of Advanced Computer Science and Application , Vol 10 issue 6 , pp. 261-268, 2019.

[11] Godse, Rudra A.,Gunjal,Smita S., JagtapKaran A .,Mahamuni ,Neha S., &Wankhade, Prof. Suchita. "Multiple Disease Prediction Using Different Machine Learning Algorithms Comparatively",.IJARCCE, Vol 8 Issue 12,pp. 50-52 , 2020.

[12] Anitha ,Dr.S.,& Sridevi,Dr.N. "Heart Disease Prediction Using Data Mining Techniques.", JAC,Vol 13 Issue 2, pp. 48-55,2019.

[13] Bindhika,Galla Siva Sai., Meghana,Munaga., ReddyManchuriSathvika. , &Rajalakshmi. "Heart Disease Prediction Using Machine Learning Techniques ",IRJET, Vol 7 Issue 4 ,pp. 5272-5276, 2020.

[14] Pingale,Kedar., Surwase, Sushant., Kulkarni,Vaibhav.,Sarage ,Saurabh., &Karve, Prof. Abhijeet "Disease Prediction using Machine Learning" ,IRJET, Vol 6 Issue12 , pp. 2810-2813, 2019.

[15] Chauhan Raj H., NaikDaksh N. ,Halpati,Rinal A., Patel,Sagarkumar J. , &PrajapatiMr. A.D. "Disease Prediction using Machine Learning." ,IRJET, Vol 7 Issue 5,pp. 2000-2002, 2020.

[16] Ananey-Obiri, D.; Sarku, E. "Predicting the Presence of Heart Diseases Using Comparative Data Mining and Machine Learning Algorithms", IJCA , Vol 176 Issue 11,pp. 17–21,2020.

[17] Mohan, S.; Thirumalai, C.; Srivastava, G. "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", IEEE , Vol 7, pp. 81542–81554,2019.

[18] Kodati, S.; Vivekanandam, R. "Analysis of Heart Disease Using in Data Mining Tools Orange and Weka", Glob. J. Comput. Sci. Technol. C, Vol 18 Issue 1, pp. 17–21,2018.

[19] Shah, S.M.S.; Batool, S.; Khan, I.; Ashraf, M.U.; Abbas, S.H.; Hussain, S.A. "Feature Extraction through Parallel Probabilistic Principal Component Analysis for Heart Disease Diagnosis" ,Phys. A Stat. Mech., Vol 482, pp. 796–807,2017.

[20] Perumal, R. "Early Prediction of Coronary Heart Disease from Cleveland Dataset Using Machine Learning Techniques ",Int. J. Adv. Sci. Technol., Vol 29, pp. 4225–4234,2020.