Diagnosis of chronic disease in a predictive model using machine learning algorithm

1st I. Preethi
Research Scholar, School of Computing sciences
VISTAS
Chennai, India
pkpijeevitha@gmail.com

2nd Dr.K. Dharmarajan
Associate Professor, Department of Information Technology
VISTAS
Chennai, India
dharmak07@gmail.com

***Abstract*— Today, digitization in healthcare industry takes the advantage on advancements in clinical healthcare services. The extensive growth in data for monitoring and analyzing the patients outcomes in predicting and diagnosis of chronic diseases lacks in traditional methods and are replaced by technologies to gather the most relevant insights from the medical data by using predictive analytics with very useful tool of machine learning. The importance of using machine learning algorithms in the model for diagnosis, shows its ability in high classification accuracy rate in reduced computational time. In this paper, a study of various machine learning techniques are used in classification of chronic diseases like heart, kidney, diabetes and cancer from multiple dataset by reducing the dimensionality using feature selection. Feature selection plays a significant role in machine learning by selecting the critical features for diagnosing chronic diseases. The performance of the classifiers are evaluated based on several metrics like classification accuracy, sensitivity, specificity, precision, F1- measure, AUC (the area under the receiver operating characteristic (ROC) curve) criterion, and processing time.**

Keywords*— *predictive analytics, diagnosis, feature selection, critical features, machine learning algorithms.

I. INTRODUCTION

The digitization of health care with evolving technology and analytics capabilities has given new insights to produce numerous tools and resources that improve healthcare services, by special software such as electronic health records (EHR). This big chunks of medical data comprising patient data, treatment history, medical records, and demographic information from a large database can benefit the healthcare practitioners in diagnosing the disease, predicting their causes along with symptoms helps in addressing new insights for better healthcare.

The electronic health records consists of various healthcare dataset that are produced from multiple sources has exploded the significance in diagnosing with machine learning techniques. In this paper, the study of various dataset are used in classification of chronic diseases. Though the chronic diseases are usually controlled by treatment, but not cured as it long-last. The chronic diseases like heart disease, cancer, diabetes, stroke, kidney disease and arthritis are the leading cause to death as they often need to manage their consequences in health problems. Diagnosis of disease is necessary at the early stage, through appropriate medical management and monitoring systematically in identifying new issues that relates to the diseases. These data's are maintained in the large database using electronic health records consisting of dataset that are maintained with different features or variables.

The different types of feature selection methods are applied in various classification problems. These features possess a high cost in prediction with classifiers as this can be avoided by reducing the attributes with various feature selection methods [10].This method is also helpful in identifying the critical factors that are related to disease for effective prediction. The feature selection is one of the pre-processing method, used in machine learning that helps the classifiers in predicting chronic disease effectively. The prediction and classification of diseases are trained and tested in the model by using different machine learning classification algorithms for high accuracy rate by optimizing the feature selection.

The optimization is a task of choosing the best objective function from various function by minimizing or maximizing function and comparing with different choices of availability in a function set. The exploratory algorithms, finds the best optimal solution by optimizing the parameters in classification of chronic diseases that outperforms in good result.

In this review paper, machine learning algorithms are used in prediction of chronic disease of different dataset. The datasets are used in the existing model are pre-processed, features are selected based on the importance and are used in the model for classification. The classification algorithms like Support Vector machine, Logistic Regression, Random Forest, Decision Tree, Artificial Neural Network are applied whether the patient is suffering from chronic disease or not. The results are evaluated based on accuracy and computational time.

The rest of the work is structured and as follows: Section 2 is a review of existing work in classifying chronic disease, Section 3 and Section 4 is a description of feature selection methods and SVM classifier, Section 5 is the evaluation of performance metrics, Section 6 is the discussion of the study using SVM and finally Section 7 is the conclusion of the study in chronic disease.

II. LITERATURE REVIEW

978-1-7281-7213-2/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: Auckland University of Technology. Downloaded on December 23,2020 at 11:37:12 UTC from IEEE Xplore. Restrictions apply.

This section represents a review of research papers in classifying the chronic disease by applying machine learning algorithms to solve the disease related issues in prediction.

(Almansour, Njoud Abdullah et al., 2019) [1] applies the machine learning algorithms for effective diagnosis of chronic kidney disease at the early stage by comparing two classifiers ANN & SVM by optimizing the parameters in a random exhaustive search method. The dataset form UCI repository of 400 instances are pre-processed and features are selected using correlation coefficient. The performance of the classifiers are compared with various best features(2,3,6,12 & all) and computational time involved in training the model. The best features of 12 are chosen for the prediction of kidney disease using ANN and SVM as the ANN outperforms SVM with high accuracy.

(Abdelaziz, Ahmed et al., 2019) [3] in this study, a hybrid machine learning model is used with two techniques LR & NN. The dataset are pre-processed by detecting incomplete records and normalized. The critical features that affects CKD are found by applying the feature weight using Linear regression algorithm. The features are reduced to 13 out of 24. The NN uses 13 critical features for diagnosis of CKD. The two class module with the critical features uses the score module to represents the probability of CKD, whose value is closer to zero and Not CKD, whose value is closer to 1. The performance of the model NN is evaluated using confusion matrix to predict the precision, recall, accuracy & F1-measure.

(Harimoorthy, Karthikeyan et al., 2020) [5] uses the SVM-Radial basis kernel method in UCI repository dataset of kidney disease, Heart disease & diabetics for predicting multi disease. The improved SVM-radial basis kernel function optimizes the cost parameters and gamma value for variance reduction. Before optimizing the parameters the data are pre-processed using mean and feature selection is done using chi square test for diabetes, heart and chronic kidney disease. The accuracy of the model is predicted by comparing with SVM-Linear, polynomial, RF, DT. The performance of the machine learning algorithm are predicted based on the accuracy, misclassification rate, precision, sensitivity and specificity. The accuracy for SVM-radial basis kernel technique outperforms other classifiers.

(Zhang, Jianfeng et al., 2017) [9] in this study the diabetes diagnosis is based on tongue image using SVM. The proposed model, GA-SVM is used to optimize the parameter of diabetes prediction based on tongue image features as input. The dataset is collected from Shuguang Hospital & TCM Hospital as it consists of 827 data, among 296 are diabetes and 531 are non-diabetes. The data is imbalanced and it is been balanced by applying SMOTE oversampling algorithm. The dimensionality of the features are reduced using PCA algorithm and the dataset are trained using SVM classifier by optimizing the parameter with GA algorithm in a 10-cross fold validation. The proposed method is compared with the other algorithm like KNN,NB,BP-NN using raw data, normalized data and normalized data with PCA. The result of normalized data with PCA as high sensitivity and specificity by other two and also has the low training time of 465.52 sec.

(Krishnani, Divya et al., 2019) [13] the machine learning algorithms of RF, DT, KNN are used as an experimental analysis for predicting the heart disease at the early stage. The data are collected from the Framingham Heart study consisting of 4240 records out of 644 are heart disease and 3596 are not suffering from heart disease. The class imbalanced dataset are balanced by oversampling method. The ROC uses 10-fold to find the accuracy of the classifier with mean-AUC. The mean AUC of RF results to 0.99 closer to 1, as the model predicts the better accuracy, were 0.92 for DT and 0.91 for KNN. The result of RF outperforms DT & KNN.

(Sneha, N and Tarun Gangil., 2019) [15] The dataset of diabetes from UCI repository is used for the prediction using machine learning algorithms like DT, RF, NB, SVM & KNN. The proposed method uses the optimal feature for the prediction. It is chosen by the method of co-relating the one attribute with all attribute and finding the difference. If the attribute produces a large difference, the attribute has low significance. The co-relation value of 11 attributes are chosen as optimal attributes & 4 attributes are removed. The result is compared with the five classifiers of all features. The proposed method with optimal features of improved SVM has high accuracy compared to others.

(Yao, Lijun et al., 2019) [17] in this study, ML classifier like XGboost, DT, RF, SVM & LR(logistic) are used to predict antiepileptic drug treatment of patient with newly diagnosed epilepsy. The outcome of AED treatment is predict using ML based on two model. In the first model, the classifiers are trained to classify the remission & never remission of 287 patients and the most significant feature is no. of seizures before treatment(>3). The second model, uses the classifiers classify the early remission & late remission of 207 remission patients and the most significant feature is multiple seizures types. The dataset collected from Second Affiliated hospital of Zhejiang university, china are pre-processed by using SMOTE oversampling and further 5-cross validation is applied. The feature selection for the two model is based on chi-square(x^2) statistics and ranked 5 features among 14 features and also calculated odds ratio, confidence interval. In both the model XGboost outperforms the best result among all classifier.

(Zhao, Dandan et al., 2019) [18] The colorectal cancer uses the significant feature selection by Logistic Regression. The significant features are firmicutes, Bacteroidetes, BMI, Age, combined factor. These features are trained using SVM classifier by comparing with all kernel types and also compared with the other classifiers of LR+NB, LR+RF, LR+KNN, LR+ANN. The performance of the classifier is evaluated using confusion matrix by calculating sensitivity, specificity, accuracy, Matthew's co-relation matrix. The LR+SVM classifier with RBF kernel outperforms both in 5- fold cross validation and 10-fold cross validation.

(Divya, Jain et al., 2019) [20] a hybrid feature reduction method of ReliefF & PCA algorithm with adaptive SVM classifier is used in predicting nine disease dataset of high & low dimensional dataset. The proposed model is divided into two phases. In the first phase, the features are removed whose value is below the threshold value by ReliefF then it passes the list of features to PCA algorithm it extracts the feature subset by assigning the score to each feature. Second phase, adaptive SVM classifier is used were the parameters are tuned in Grid search

Authorized licensed use limited to: Auckland University of Technology. Downloaded on December 23,2020 at 11:37:12 UTC from IEEE Xplore. Restrictions apply.

method(cost, gamma) based on MAE, accuracy and computational time. The results are compared with standard SVM classifier & proposed model without feature reduction. Adaptive SVM classifier outperforms well with the execution time of 68 sec and reduced 61.6% of features in a multiclass dataset.

III. FEATURE SELECTION

The selection of features plays a significant role in the pre-processing before the model is trained for the prediction. Feature selection is widely used in dataset to reduce the cost of prediction, removing the irrelevant and redundant features that may lead to overfitting as this affects the prediction of disease. Moreover this technique provides better accuracy and reduced training time. The existence of various feature selection methods comes under three categories namely filter, wrapper and embedded methods. But in recent times, a new method of hybrid feature selection is also used in the classification dataset. The feature selection methods are applied in many research works and is presented in TABLE I.

A. Filter method

The filter approach in feature selection proposes a new subset of features in search space based on ranking without depending the machine learning algorithms. The performance of selected features reduces the computational time when it is evaluated by applying the classifiers to the model. This method works very fast for better result in low computational complexity for high dimensional dataset [10]. (Fig. 1)

Correlation coefficient filter method is used in the proposed method to find the optimal feature for the diabetes prediction. This method co-relates the one attribute with all attribute and finding the large difference and this produce that the attribute has low significance. The co-relation value of 11 attributes are chosen as optimal attributes & 4 attributes are removed. The result of the proposed method with optimal features improved SVM & NB with accuracy [15].

Liyun Yao et al. [17] applied Chi-square(X^2) method for feature selection in two models to predict antiepileptic drug treatment of patient with newly diagnosed epilepsy. The outcome of AED treatment is predict using ML based on two model. The feature selection ranked 5 features among 14 features. In both the model XGboost outperforms the best result among all classifier of DT, RF, SVM and LR with the better result using feature selection.

B. Wrapper Method

The wrapper approach produces a score to subset of features with the trained model. The performance of the wrapper method for every new feature subset is evaluated by the machine learning algorithms. Though this process takes in a high computational time for processing each and every subset but shows better accuracy in performance by finding out the best features in detecting their dependencies[10](Fig. 1).

The author analysed the usage of a genetic algorithm with no feature selection [6] in a hybrid model of GA-SVM to find optimal parameters using the classifier SVM as search strategy with RBF kernel. The fitness function in GA orients searching strategy to find the best solution in large space. The fitness evaluation is performed based on the conjunction of SVM with GA to optimize the feature weights is being tested randomly by using 10-cross fold validation in training model. The classical SVM show low accuracy with other classifiers but when conjunctive with GA-SVM, it has the high accuracy in both diabetes and heart disease.

Liming shen et al. [7] a new swarm intelligence technique FOA-SVM is used to maximize the capability of addressing the parameter set in SVM. The FOA-SVM involves two methods, the first method is inner parameter optimization that the parameters are adjusted by the FOA techniques using 5-Cross validation techniques. The second method is outer classification optimization that the optimal parameters are trained by the SVM classifier using 10-Cross validation. The model is evaluated by comparing optimization algorithm like PSO-SVM, GA-SVM, BFO-SVM & Grid -SVM. The performance of FOA-SVM outperforms with the accuracy of 73.34% in accuracy and time.

Miray Akgul et al. [12] The heart disease is predicted by a hybrid approach of ANN-GA as the data is collected from UCI repository. The model uses the ANN-GA for classification by optimizing the parameters of ANN and feature selection using GA respectively. The accuracy is compared with NB, KNN, ANN, C4.5 algorithm using the confusion matrix. The ANN-GA outperforms with 95.82% of accuracy with feature selection method.

A comparative analysis of filter method and wrapper method was done by the author [14]. The approach used in GA-SVM is to optimize the parameter using SVM and the objective function is to select the significant features using the best fitness value for the heart disease prediction. The wrapper method GA, outperforms other filter feature selection method (Relief, Info Gain, Chi, Filtered, one attribute, consistency, Gain ratio, Filtered subset, CFS) and are compared with different classifiers like SVM, MLP, J48 & K-NN as GA-SVM outperforms SVM classifier with accuracy. The GA-SVM selects 7 features out of 13 to evaluate the significance in the prediction of heart disease.

C. Embedded Method

The embedded method benefits the approach of filter and wrapper as the feature subset are chosen based on the learning algorithm to train the model by deriving the important features for prediction. This method produces effective feature dependencies among other features, that optimizes the classifier to perform better accuracy than filter method with least computational cost than wrapper method[10](Fig. 1).

The author utilizes three base learners Lasso, Elastic Net and Ridge in a proposed model of EPL [19] that are fused together through majority voting in bootstrap m samples and final classification done on the basis of high vote from more than two or three models that predicts the same value. In simulated dataset, p=400 predictors are chosen with informative and non-informative features & sample size are 50,100 & 200 is used. The results of all the examples are compared based on the feature selection that are tested with the misclassification error rate. This proposed model outperforms in accuracy for high-dimensional data such as microarray dataset, when the sample size is small, the ratio of irrelevant features are high and also

Authorized licensed use limited to: Auckland University of Technology. Downloaded on December 23,2020 at 11:37:12 UTC from IEEE Xplore. Restrictions apply.

when the features are highly co-related. EPL resolves the problem of multicollinearity and non-informative features.

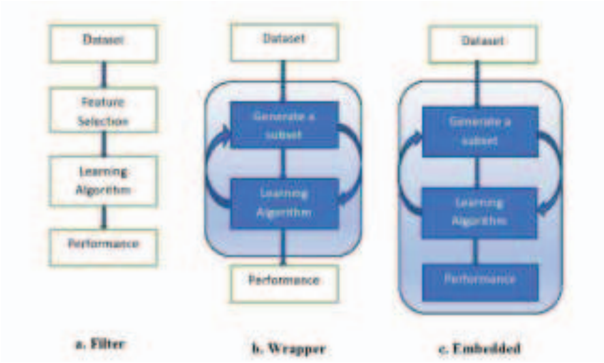

Fig. 1 Feature Selection methods

IV. SUPPORT VECTOR MACHINE

This section describes about the supervised machine learning algorithm support vector machine (SVM) is broadly used by many researchers in the review above mentioned for the classification of chronic diseases.

Support Vector machine (SVM) is mainly used to classify the two distinct groups in a supervised classification dataset. The two distinct groups or classes of their data points are separated by a hyperplane. The linear hyperplane is used for the low dimensional dataset (linear). Moreover many number of hyperplanes are drawn to reach the optimal value using margin. The margin is the distance between the data points and hyperplane. The objective of the hyperplane is to find the best data points of the class by maximizing the margin for a linear dataset (Fig. 2).

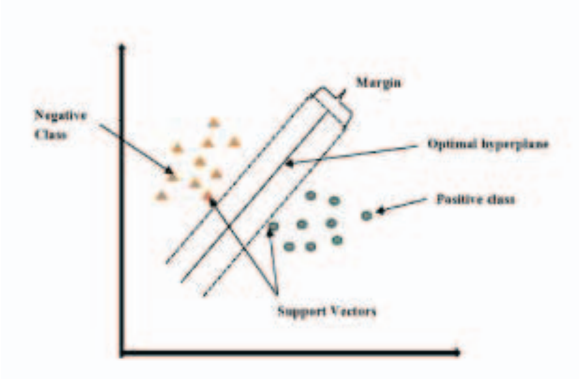

Fig. 2 Schematic of linear SVM

The support vector machines also effectively performs for a high dimensional dataset (non- linear), were the input features are greater than the number of instances. The data points in the non-linear dataset are classified in a three dimensional space by kernel parameters of Radial basis function (RBF) and polynomial. The parameters like C penalty for finding the misclassification error rate and gamma value to fit the training dataset for prediction are tuned to find the optimal solution in non-linear hyperplane. The SVM algorithm is found to have better performance in accuracy, utilization of memory and computational time.

TABLE I. ACCURACY OF DIFFERENT CLASSIFIERS WITH FEATURE SELECTION METHODS IN A DIFFERENT DATASET

S. no	Disease	Feature Selection	Dataset	Classifier	Accuracy
1.	Chronic Kidney disease	Correlation Coefficient	UCI repository	ANN,SVM	ANN – 99.75%
2.	Chronic Kidney disease	-	UCI repository	PNN,SVM,RBF,MLP	PNN – 96.7%
3.	Chronic Kidney disease	-	UCI repository	LR & NN	NN – 97.8%
4.	Heart	Chi-square	UCI repository	SVM(linear, poly, rbf, sigmoid), RF,DT	SVM(rbf) – 89.9%
	Chronic Kidney disease	Chi-square	UCI repository	SVM(linear, poly, rbf, sigmoid), RF,DT	SVM(rbf) – 98.3%
	Diabetes	Chi-square	UCI repository	SVM(linear, poly, rbf, sigmoid), RF,DT	SVM(rbf) – 98.7%
5.	Diabetes	Genetic Algorithm-wrapper	Pima Indian Diabetes	GA-SVM(rbf), C4.5,RF, NB, Boosting	GA-SVM(rbf) – 81.51%
	Heart	Genetic Algorithm-wrapper	Statlog Project	GA-SVM(rbf), C4.5,RF, NB, Boosting	GA-SVM(rbf) – 89.63%
6.	Diabetes	Fruit Fly algorithm – wrapper	-	FOA-SVM, GA-SVM, PSO-SVM, BFO-SVM,Grid-SVM	FOA-SVM-72.34%
7.	Chronic Kidney disease	Classifier subset evaluator with greedy stepwise search engine	UCI repository	SVM	SVM - 98%
		Wrapper subset evaluator with Best First search		SVM	SVM - 98.25%
		Correlation feature subset evaluator with greedy stepwise search engine		SVM	SVM – 98.75%

Authorized licensed use limited to: Auckland University of Technology. Downloaded on December 23,2020 at 11:37:12 UTC from IEEE Xplore. Restrictions apply.

S. no	Disease	Feature Selection	Dataset	Classifier	Accuracy
		Filter subset evaluator with Best First search		SVM	SVM – 98.5%
9.	Heart	Genetic algorithm-wrapper	UCI repository	ANN-GA, NB, KNN, C4.5, ANN	ANN-GA- 95.82%
10.	Heart	-	Framingham heart study	RF, DT, KNN	RF- 96.8%
11.	Heart	Genetic algorithm – wrapper	UCI repository	GA-SVM, SVM, MLP, J48, KNN	GA-SVM – 88.34%
12.	Diabetes	Correlation coefficient	UCI repository	SVM, KNN, NB, DT, RF	NB – 82.30%
13.	Colorectal cancer	-	NCBI	LR+SVM, LR+NB, LR+RF, LR+KNN, LR+ANN	LR+SVM – 91.2%
14.	Colon	ReliefF & PCA	Microarray	SVM(rbf)	SVM(rbf) - 71.43%
15.	Heart	ReliefF & PCA	UCI repository	SVM(rbf)	SVM(rbf) - 86.67%
	Chronic Kidney disease		UCI repository		SVM(rbf) - 97.48%
	Diabetes		Pima Indian Diabetes		SVM(rbf) - 73.91%

V. MEASURE OF PERFORMANCE METRICS

This section describes the performance of the experimental analysis by confusion matrix as many existing work is evaluated by utilizing the metrics such as accuracy, sensitivity, specificity, precision, recall, True positive, False positive, F1- measure for comparing with their related work. The elements in metrics are described as follows

True Positive (TP) – The no. of instances with the disease is classified correctly as the output is positive .

True Negative (TN) – The no. of instances without disease is classified correctly as the output is negative.

False Positive (FP) – The no. of instances without the disease is wrongly classified as the output is positive.

False Negative (FN) – The no. of instances with disease is wrongly classified as the output is negative. The metrics calculation are described as follows with Eq.

Accuracy- calculates the performance of the classifier based on the dataset.

$$\text{Accuracy}=TP+TN/TP+TN+FP+FN \tag{1}$$

Sensitivity or Recall – The target class is positive and classified as positive correctly with high value.

$$\text{Sensitivity}=\text{Recall}=TP/TP+FN \tag{2}$$

$$\text{Specificity}=TN/TN+FP \tag{3}$$

$$\text{Precision}=TP/TP+FP \tag{4}$$

$$\text{F-Measure}=2*\text{Recall} * \text{Precision}/\text{Recall}+\text{Precision} \tag{5}$$

Receiver operating characteristic curve (ROC) – It is a graphical tool that distinguish the two classes in a two dimensional graph with the probability of different threshold value.

VI. DISCUSSION

In this section, the study of various chronic disease papers are discussed based on the feature selection method that are used in learning algorithm for classification in the prediction model. The result of the model focuses on the accuracy and runtime of the learning algorithm. The feature selection method applied in the model makes a difference in the result in every paper. The filter method like correlation coefficient, chi square are used in the dataset of chronic kidney disease using SVM classifier that outperforms the best result in accuracy and computational time compared to other machine learning algorithms[8,15,17].

In wrapper method, the learning algorithm is effectively used in selecting the feature subset. The genetic algorithm and swarm intelligence algorithm like fruit fly algorithm are used as the feature selection method with the adaptive learning classifier of SVM. Low dimensional datasets are used for the classification as it results less computational time for feature selection.

The feature dependencies are highly correlated with learning algorithm this lacks in filter method[6,7,14]. Moreover accuracy is highly predictable than filter method. The embedded method uses the learning algorithm to optimize its performance based on right feature selection, adaptive learning algorithm SVM is used in comparison of ensembled logistic regression model in high dimensional dataset. The SVM outperformed the ensembled logistic regression model were the features are high correlated and sample size is high with more no. of non- informative features[20].

The major contribution in the existing works involves the usage of support vector learning algorithm in a dataset of linear or non-linear. The SVM can transform the data in a new dimensional space using kernel techniques for a non-separable data to reach the optimal hyperplane to distinguish the target class. The classifier performance outperforms with feature selection method with the enhanced accuracy.

VII. CONCLUSION

Chronic disease are leading cause to death due to the unawareness of disease at the early stage. The diagnosis of the disease is necessary as it can only be treated with proper medications but not cured. This paper presents a review of predicting the chronic disease with different feature selection method and classification algorithm using various dataset (see in TABLE I). The classifiers that are used in predicting the chronic disease in the existing work are Support Vector Machine, Logistic Regression, Random Forest, Decision Tree, k-nearest neighbors. The performance of the prediction model in the review papers implies an evaluation matrix for calculating the accuracy, sensitivity, specificity, precision, recall, F1-measure and computational time. Among all the classifiers

Authorized licensed use limited to: Auckland University of Technology. Downloaded on December 23,2020 at 11:37:12 UTC from IEEE Xplore. Restrictions apply.

involved in the prediction of chronic diseases, Support vector machine algorithm is mostly used in many research works with the selective feature selection method in both high dimensional and low dimensional dataset as it is handles the multicollinearity effectively in large samples.

REFERENCE

[1] Almansour, Njoud Abdullah, et al. "Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study." Computers in biology and medicine 109 (2019): 101-111.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Rady, El-Houssainy A., and Ayman S. Anwar. "Prediction of kidney disease stages using data mining algorithms." Informatics in Medicine Unlocked (2019): 100178.

[3] Abdelaziz, Ahmed, et al. "A machine learning model for predicting of chronic kidney disease based internet of things and cloud computing in smart cities." Security in Smart Cities: Models, Applications, and Challenges. Springer, Cham, 2019. 93-114.

[4] Al Imran, Abdullah, Md Nur Amin, and Fatema Tuj Johora. "Classification of Chronic Kidney Disease using Logistic Regression, Feedforward Neural Network and Wide & Deep Learning." *2018 International Conference on Innovation in Engineering and Technology (ICIET)*. IEEE, 2018.

[5] Harimoorthy, Karthikeyan, and Menakadevi Thangavelu. "Multi-disease prediction model using improved SVM-radial bias technique in healthcare monitoring system." *Journal of Ambient Intelligence and Humanized Computing* (2020): 1-9.

[6] Phan, Anh Viet, Minh Le Nguyen, and Lam Thu Bui. "Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems." *Applied Intelligence* 46.2 (2017): 455-469.

[7] Shen, Liming, et al. "Evolving support vector machines using fruit fly optimization for medical data classification." *Knowledge-Based Systems* 96 (2016): 61-75.

[8] Polat, Huseyin, Homay Danaei Mehr, and Aydin Cetin. "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods." Journal of medical systems 41.4 (2017): 55.

[9] Zhang, Jianfeng, et al. "Diagnostic method of diabetes based on support vector machine and tongue images." BioMed research international 2017 (2017).

[10] Jain, Divya, and Vijendra Singh. "Feature selection and classification systems for chronic disease prediction: A review." *Egyptian Informatics Journal* 19.3 (2018): 179-189.

[11] Jain, Divya, and Vijendra Singh. "An efficient hybrid feature selection model for dimensionality reduction." *Procedia Computer Science* 132 (2018): 333-341.

[12] Akgül, Miray, Özlen Erkal Sönmez, and Tuncay Özcan. "Diagnosis of Heart Disease Using an Intelligent Method: A Hybrid ANN–GA Approach." *International Conference on Intelligent and Fuzzy Systems*. Springer, Cham, 2019.

[13] Krishnani, Divya, et al. "Prediction of Coronary Heart Disease using Supervised Machine Learning Algorithms." *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*. IEEE, 2019.

[14] Gokulnath, Chandra Babu, and S. P. Shantharajah. "An optimized feature selection based on genetic approach and support vector machine for heart disease." *Cluster Computing* 22.6 (2019): 14777-14787.

[15] Sneha, N., and Tarun Gangil. "Analysis of diabetes mellitus for early prediction using optimal features selection." *Journal of Big Data* 6.1 (2019): 13.

[16] Boukenze, Basma, Abdelkrim Haqiq, and Hajar Mousannif. "Predicting Chronic Kidney Failure Disease Using Data Mining Techniques." *International Symposium on Ubiquitous Networking*. Springer, Singapore, 2016.

[17] Yao, Lijun, et al. "Prediction of antiepileptic drug treatment outcomes of patients with newly diagnosed epilepsy by machine learning." *Epilepsy & Behavior* 96 (2019): 92-97.

[18] Zhao, Dandan, et al. "A reliable method for colorectal cancer prediction based on feature selection and support vector machine." Medical & biological engineering & computing 57.4 (2019): 901-912.

[19] Ijaz, Musarrat, Zahid Asghar, and Asma Gul. "Ensemble of penalized logistic models for classification of high-dimensional data." *Communications in Statistics-Simulation and Computation* (2019): 1-17.

[20] Jain, Divya, and Vijendra Singh. "A two-phase hybrid approach using feature selection and Adaptive SVM for chronic disease classification." *International Journal of Computers and Applications* (2019): 1-13.

Authorized licensed use limited to: Auckland University of Technology. Downloaded on December 23,2020 at 11:37:12 UTC from IEEE Xplore. Restrictions apply.