



Performance Evaluation of Different Machine Learning Classification Algorithms for Disease Diagnosis

Munder Abdulatef Al-Hashem, Yarmouk University, Jordan

Ali Mohammad Alqudah, Yarmouk University, Jordan

 <https://orcid.org/0000-0002-5417-0043>

Qasem Qananwah, Yarmouk University, Jordan

 <https://orcid.org/0000-0002-4291-9643>

ABSTRACT

Knowledge extraction within a healthcare field is a very challenging task since we have many problems such as noise and imbalanced datasets. They are obtained from clinical studies where uncertainty and variability are popular. Lately, a wide number of machine learning algorithms are considered and evaluated to check their validity of being used in the medical field. Usually, the classification algorithms are compared against medical experts who are specialized in certain disease diagnoses and provide an effective methodological evaluation of classifiers by applying performance metrics. The performance metrics contain three criteria—accuracy, sensitivity, and specificity—forming the confusion matrix of each used algorithm. The authors have utilized eight different well-known machine learning algorithms to evaluate their performances in six different medical datasets. Based on the experimental results, the authors conclude that the XGBoost and K-nearest neighbor classifiers were the best overall among the used datasets and can be used for diagnosing various diseases.

KEYWORDS

Artificial Intelligence, Machine Learning, Medical Diseases, Performance

1. INTRODUCTION

Machine learning (ML) is one of the main fields of artificial intelligence (AI) in which different areas like statistics, computer science, and mathematics are integrated to propose intelligent systems that mimic human behaviors (D. Jain & Singh, 2020; Liu et al., 2019). This integration is used to build and extract a model by learning from the past experiences data (large and complex datasets) to predict feature data. Such a model may mimic the human behaviors analog to the human brain neurons and human decision-making rules (Harper, 2005; D. Jain & Singh, 2020).

Recently, ML are widely used in real-life applications and spread over various life sectors, one of these primary sectors is the healthcare sector (Uddin et al., 2019). The Healthcare sector is a vital and crucial sector that is related directly to human life (Harper, 2005). Using ML in this sector will enhance the health and welfare of the people and will help in decreasing the error rates in medical diagnosis and procedures. The ML is usually integrated with medical device systems to help them enhancing their diagnosis results and increasing stability and accuracy (Abbas et al., 2020).

DOI: 10.4018/IJEHMC.20211101.oa5

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

ML can provide helps in urgent situations in the medical field where quick and correct decision-making activities are needed to decide patient health. Moreover, ML can help clinicians to find and select the optimal operation methods especially in cases where it is hard to anticipate the outcomes of selected diagnosis on the patient health (Abbas et al., 2020; Uddin et al., 2019). Researchers paid their attention to designing and developing AI-based systems that will assist physicians to gain insights and anticipate different outcomes relevant to the area of interest (A. Jain et al., 2018).

Recent research on automatic features, decision, and knowledge extraction from healthcare data has been growing exponentially during the last two decades. However, developing a system that automatically extracts accurate and exhaustive information from health data is a complex and challenging task (Abbas et al., 2020; Liu et al., 2019). Healthcare datasets are available and sometimes have incomplete information or missing values. The main problem in the health datasets is the imbalanced class distribution between healthy and non-healthy which makes the system biased towards either of them which affects the systems accuracy, sensitivity, and precision of classification algorithms used for disease diagnosis and detection (Abbas et al., 2020; A. Jain et al., 2018; Uddin et al., 2019).

Many problems arise when using ML in healthcare for diagnosis (classification) purposes. The main issue is that there is no standard methodology or technique used for evaluating the performance of ML algorithms on healthcare datasets (Harper, 2005; Liu et al., 2019). Different researchers have used different methodologies to evaluate the performance of their systems such as training sets, testing sets, and cross-validations (CV) methods. The evaluation technique is usually reflected in the performance metrics values, especially in the case of biased dataset problems (D. Jain & Singh, 2020; Uddin et al., 2019).

The rest of the paper is organized as follows: Section 2 describes the related work; Section 3 explains the proposed methodologies and evaluation framework used for comparing multiple ML algorithms that were applied to different healthcare datasets. In section 4 the experimental results of proposed methodologies were reported and the issue of biased misclassification cost was addressed. Section 5 present the discussion of the results and the comparison between different utilized classifiers. Finally, the conclusion of our work is presented in Section 6.

2. LITERATURE REVIEW

In this work, six different diseases will be classified utilizing the available database. These diseases were addressed previously by many researchers. The related works related to these diseases is presented in the following review:

2.1 Chronic Kidney Disease (CKD)

The first disease considered in our study is CKD which was addressed by many researchers. Revathy et al (Revathy et al., 2019) had tried to propose a data mining framework and various classifiers on the CKD UCI dataset. Three ML algorithms such as decision tree (DT), support vector machines (SVM), and random forest (RF) were used to predict the early occurrence of CKD. The results showed that the accuracies were 94.16%,98.33%, 99.16% for the three algorithms respectively.

Devika et al (Devika et al., 2019) examines the performance of naive bayes (NB), K-nearest neighbor (KNN), and RF classifiers based on their accuracies, preciseness, and execution time for predicting CKD. The results showed that the RF classifier was the most accurate and had accuracy more than NB and KNN.

Tazin et al (Tazin et al., 2017) presented a method to predict the presence of CKD. The classifiers SVM, DT, NB, and KNN algorithms were used in analyzing the CKD obtained from a dataset collected by the UCI repository. The results were analyzed in terms of accuracy, root mean squared error (RMSE), mean absolute error (MAE), and receiver operating characteristic curve (ROC). The results showed that the DT algorithm was with the highest accuracy when it was implemented through

waikato environment for knowledge analysis (WEKA) data mining tool. It is a tool used to predict the presence of different attributes available in the dataset. The performance of the final results showed that the DT algorithm had an accuracy of 99% and the second-most accurate classifier was SVM with an accuracy of 97.75%.

Vijayarani and Dhayanani (Vijayarani et al., 2015) proposed a method for predicting kidney diseases by using SVM and artificial neural network (ANN) on a dataset collected from several medical labs, centers, and hospitals. The method was concerned with a comparison for the performance of these two algorithms based on their accuracies and execution time. The results showed that the performance of ANN was better than the SVM algorithm.

Tekale et.al (Kotturu et al., 2019) reported different ML algorithms such as DT and SVM to predict CKD with its severity. The CKD dataset was obtained from the UCI respiratory database and it was reduced into 14 attributes instead of 25 attributes. The results showed that the SVM accuracy was 96.75%, while DT accuracy was 91.75%.

Abdulhamit et al (Abdulhamit Subas, Emina Alickovic, 2017) presented a study for predicting and identifying CKD through a real data set taken from the UCI machine learning repository. Various ML classifiers were applied to this dataset. The results showed that the RF classifier provided very high performances among ANN, KNN, SVM, and DT.

Imran et al (Imran et al., 2019) were applied three ML techniques namely LR, feedforward neural networks, and wide & deep learning to diagnose CKD targeting best performance. F1-score, precision, recall, area under the ROC curve (AUC), and loss score was analyzed. The results showed that the feedforward neural network had the best performance among all other techniques with 0.99 F1-score, 0.97 precision, 0.99 recall, and 0.99 AUC score for diagnosing CKD, while wide & deep learning found to be efficient for larger datasets.

Helmie et al (Wibawa et al., 2019) evaluated kernel-based extreme learning machine (ELM) to predict CKD. The method involved making a comparison of four kernels-based ELM performances, namely radial basis function (RBF-ELM), Linear-ELM, polynomial-ELM, and wavelet-ELM with the performance of standard ELM. The results showed that RBF-ELM had a higher sensitivity and specificity 99.38% and 100% respectively.

Anusorn et.al (Charleonnann et al., 2017) presented ML algorithms to diagnose CKD employing UCI machine learning repository clinical data. The results showed that SVM had the highest accuracy, sensitivity, and specificity among KNN, Logistic Regression (LR), and DT.

Gunarathne et al (W.H.S.D, 2017) focused on predicting the status of patients whether CKD or non-CKD. Different ML classification algorithms had been used to diagnose patients. These algorithms were applied on the CKD dataset downloaded from the UCI repository with 400 data records and 25 attributes which were reduced into 14 attributes. The results showed that the multiclass decision forest classifier was the best with an accuracy of 99.1%.

Merve and Aydin (Basar & Akan, 2018) applied Bagging, Adaboost, and Random Subspaces ensemble learning algorithms for diagnosing and predicting CKD exploiting CKD dataset from the UCI repository. The results showed that ensemble learning classifiers gave the perfectible classification performance with kappa and accuracy rather than individual classifiers.

Dilip (Sisodia & Verma, 2018) had implemented a method to predict CKD using three different individual classifiers namely, NB, minimal sequential optimization (SMO), DT (J48), and three ensemble classifiers namely, AdaBoost, bagging, RF. The CKD dataset was taken from the UCI repository and applied to the WEKA tool. The results were evaluated based on accuracy, precision, recall, F-measure, and ROC performance measures. The results suggested that DT (J48) from the individual learner and RF from the ensemble classifier were better compared to the other classifiers.

Sumana and Baisakhi (De, 2020) had developed a system called chronic kidney disease prediction system (CKDPS) and they built a Graphical User Interface (GUI) that can predict CKD among patients. Different ML algorithms such as KNN, LR, DT, RF, NB, SVM, MLP were applied to the dataset which was taken from Kaggle. Their performances were compared according to precision,

recall, and accuracy. The results showed that RF retained 100% precision, recall, and accuracy results. They also had made a comparison of the accuracy results for various machine learning algorithms from different previous related works where the same CKD dataset was used.

2.2 Breast Cancer Disease

The second disease which was also considered in our study is breast cancer disease. It has been addressed by many researchers e.g. Meriem et al (Eleyan, 2012) presented a comparison between two different machine learning classifiers: NB and KNN for classifying breast cancer and estimate their accuracy using cross-validation. The Results showed that KNN had the highest accuracy about 97.51% with the lowest error rate compared to the NB classifier.

David et al (Omondiage et al., 2019) presented three machine learning methods namely, SVM, ANN, and NB all besides making feature selection and extraction to identify the most significant features technique for diagnosing and predicting breast cancer disease. The Wisconsin diagnostic breast cancer (WDBC) Dataset was used, and it was going under dimensionality reduction and then applying the new reduced feature dataset to the three machine learning algorithms. The results showed that the SVM was the best compared to the other two classifiers with an accuracy of 98.82%, sensitivity of 98.41%, the specificity of 99.07%, and AUC of 0.9994.

Oyewola et al (Oyewola et al., 2017) presented a mammographic diagnostic method for the consequence prediction of breast cancer biopsy by using five machine learning algorithms: LR, RF, SVM, linear discriminant analysis (LDA), and Quadratic Discriminant Analysis (QDA). The results showed that SVM was better than others in diagnosing breast cancer with an accuracy of 95.8%, sensitivity of 98.4%, a specificity of 94.4%, ROC of 99.9%.

Agarap (Agarap, 2018) made a comparison of six different machine learning algorithms utilizing the WDBC dataset to predict breast cancer. gated recurrent unit (GRU-SVM), linear regression, multilayer perceptron (MLP), nearest neighbor (NN) search, softmax regression, and SVM were used. The results showed that MLP was the most reliable in diagnosing breast cancer with an accuracy of 99.04%

Idowu et al (Williams et al., 2015) focused on utilizing two data mining techniques namely, NB and DT (J48) to diagnose the risks of breast cancer in Nigerian patients. The results showed that the DT (J48) was a more efficient and effective classifier algorithm for predicting the risks of breast cancer due to its higher accuracy and lower error compared to NB.

Asri et al (Asri et al., 2016) presented a comparison of performance for various machine learning algorithms: SVM, DT (C4.5), NB, and KNN on original WBC to diagnose the presence of women's breast cancer. The experimental results showed that the SVM classifier gave the highest accuracy 97.13% with the lowest error rate after executing the dataset into the WEKA data mining tool.

Delen et al (Delen et al., 2005) developed the prediction models for breast cancer survivability. ANN and DT are the two data mining algorithms along with the statistical method (logistic regression) that were used in the development process using a large dataset of more than 200,000 cases. The results displayed that DT (C5) is the best classifier in predicting, with 93.6% accuracy, while the ANN and DT had an accuracy of 91.2%, 89.2% respectively.

Aruna et al (Aruna et al., 2011) reported a performance comparison criterion of machine learning algorithms through classifying breast cancer on both binary and multi-class datasets namely WBC, WDBC, and breast tissue from the UCI repository. NB, SVM, Radial Basis Neural Networks (RBNN), DT (J48), and simple classification and regression trees (CART) are the data mining tools that were used. All the experiments were conducted in WEKA to discover the most perfectible classifier concerning the accuracy, precision, sensitivity, and specificity in diagnosing breast cancer.

Bhardwaj and Tiwari (Bhardwaj & Tiwari, 2015) proposed a genetically optimized neural network (GONN) algorithm to solve the problems related to breast cancer tumors. They were classified as benign or malignant through the WBCD database from the UCI machine learning repository. The

results showed that the GONN classifier gives an accuracy of 98.24%, 99.63%, and 100% for 50-50, 60-40, 70-30 training-testing partition respectively, and 100% for 10-fold cross-validation.

Abdel-Zaher and Eldeib (Abdel-Zaher & Eldeib, 2016) developed the computer-aided diagnosis (CAD) that uses a deep belief network unsupervised path followed by backpropagation supervised path for the detection and diagnosis of breast cancer. The technique was tested on the WBCD. The results showed that the proposed developed system reached an effective and efficient classification model for breast cancer detection.

Janghel et al (Pandey et al., 2020) focused on correct diagnosing of breast cancer from the UCI repository dataset, by employing thirteen machine learning models and comparing them to the various measures. The results showed that AdaBoost, LR, and KNN models were promising high accuracy of 98% in experimenting with all the models.

Muhammet (Ak, 2020) made a comparative analysis study using data visualization, different machine learning, and data mining techniques for detecting and diagnosing breast cancer from the breast cancer UCI repository dataset. The results showed that LR with all features included had the highest classification accuracy of 98.1%.

2.3 Heart Rate Disease

The third disease considered in this study was the heart rate (HR) disease. It has been addressed by many researchers, Kavitha and Kannan (Kavitha & Kannan, 2016) presented a study on removing irrelevant and redundant data to facilitate the prediction of heart diseases by using the UCI heart rate respiratory dataset. Principal component analysis (PCA) is used to reduce the dimension of high data to lower-dimensional data. This proposed study helps in improving the accuracy, efficiency, and speed of the process.

Sowmiya and Sumitra (Sowmiya & Sumitra, 2018) presented a comparison of several machine learning classification techniques for the prediction of heart disease. UCI respiratory heart disease dataset was used in this study. The analysis showed that the classification algorithms contributing to high accuracy and obtaining high effectiveness in diagnosing heart disease.

Golande and Kumar (Golande & Pavan Kumar, 2019) presented a study of different tools and algorithms that were used in identifying patients whether to be affected by heart diseases or not. DT, KNN, K-mean clustering, and Adaboost were the algorithms and classifiers used for prediction. The results showed that the accurateness of the construction can be made efficient by creating different combinations of data mining methods and also by tuning the parameters.

Latha and Jeeva (Latha & Jeeva, 2019) worked on increasing the accuracy of weak classifiers and the implementation of the algorithms with a medical dataset. The results showed an extreme increase in accuracy about 7% for weak classifiers, with the help of ensemble classification such as bagging and boosting. They were excellent as much as possible to demonstrate the performance in distinguishing the risks of heart diseases.

Pahwa and Kumar (Pahwa & Kumar, 2017) proposed a prediction algorithm of heart diseases by firstly conducting features selection to improve the performance of models then classifying them. SVM-RFE and gain ratio (GR) algorithms were applied to the dataset to make the features selection according to their weights. The experimental results showed that the proposed technique had improved the accuracy and reduced the computation time for both models RF and NB.

Krishnan and Geetha (Ansari et al., 2021) suggested a system for predicting the possibilities of heart disease in terms of percentage. Classifiers algorithms such as DT, NB were used to improve the accuracy level. The results showed that DT gives an accuracy of 91% while NB gives an accuracy of 87%.

Rajesh et al (Haribaabu et al., 2018) presented a study in predicting heart diseases by exhibiting the expectancy framework for heart infection by using learning vector quantization neural (LVQN) system calculation. Besides, NB and DT algorithms were also used. The results showed that the NB offered better accuracy if the dataset is small, while DT was better if the dataset is large.

Swain et al (Swain et al., 2020) presented a study to find the best classifier algorithm that helps the nonspecialized doctors or medical technicians in the way of diagnosing and predicting the risk of heart diseases. The study focused on using different machine learning algorithms like LR, SVM, KNN, NB, DT, and RF classifiers. The results showed that the LR was the best with high accuracy of 88.29% compared to the others.

Chen et al (Chen et al., 2011) developed a heart disease prediction system (HDPS) that can assist medical professionals in diagnosing and predicting the risks of heart disease based on the clinical data of patients. Also, an ANN algorithm was developed for classifying heart disease. Thirteen important clinical features were used in the classification, the results showed that the accuracy of ANN was nearly 80%.

2.4 Parkinson Disease

The fourth disease considered in this study was Parkinson's disease. It was addressed by many researchers, Resul (Das, 2010) compared various types of classification methods such as Neural Networks, DMneural, Regression, and Decision Tree that used to diagnose Parkinson's diseases effectively. The results showed that the neural network was the best classifier in detection with an accuracy of about 92.9%.

Sood Khandnor (B et al., 2019) showcased a comparison study of assorted datasets, algorithms for classifying Parkinson's disease. They presented various machine learning algorithms for Parkinson's disease classifications based on the UCI Spiral dataset of Parkinson's disease. The results showed that machine learning algorithms such as RF, DT, and KNN had accuracies of 92.1%, 90.6%, and 89% respectively while utilizing deep learning techniques such as CNN with VGG16 and CNN with MobileNet gave accuracies of 63.15%, 84.62% respectively.

Senturk (Karapinar Senturk, 2020) was diagnosed with Parkinson's disease based on machine learning techniques. The dataset was passed through feature selection using Feature Importance and recursive feature elimination (RFE) methods. Classification utilizing Regression Trees, ANN, and SVM were applied to classify Parkinson's patient's disease. The results showed that SVM with RFE had better accuracy at 93.84% rather than other techniques. Mandal and Sairam (Mandal & Sairam, 2014) presented a study for predicting Parkinson's disease to prevent the delay and misdiagnosis of patients. New machine-learning methods were employed such as sparse multinomial logistic regression, rotation forest ensemble with SVM, PCA, ANN, and boosting methods. The results showed that linear logistic regression and sparse multinomial logistic regression had the highest accuracy, sensitivity, and specificity with 100%, 98.3%, and 99.6% respectively.

Gupta et al (Gupta et al., 2018) presented a method to fulfill the optimal subset of features for Parkinson's disease dataset based on the cuttlefish algorithm. The results showed that the proposed method had employed fewer numbers of features and achieved a higher accuracy of 92.19% in comparison to the average accuracy of 87.11% and 84.48% obtained by the other machine learning algorithms like KNN and DT respectively.

Wroge et al (Wroge et al., 2019) proposed a study on exploring the efficiency of using supervised classification algorithms, such as deep neural networks to diagnose patients with Parkinson's disease accurately. The results showed that the highest accuracy of 85% provided by machine learning techniques overrun the average accuracy of clinical non-experts diagnosis of 73.8%

Rus et al (Rusz et al., 2011) presented a study on detecting Parkinson's disease by using vocal to determine whether the speech and voice disorders are presenting from the early stages of Parkinson's disease. The study was conducted on Forty-six Czech native speakers. The methodology involved different speech tasks such as reading, reading the text rhythmically, etc. The results showed that the SVM machine classification algorithm reached an accuracy of 91%. Also, Goel et al (Ella & Deepak, 2018) addressed the features extraction from multiple voice recordings and the handwriting patterns while the Parkinson's patients were drawing meanders and spirals. SVM, KNN, DT, and RF were the four machine learning classifiers used for diagnosing Parkinson's disease. The results showed that

the highest accuracy obtained was 99.67% utilizing LDA with DT algorithms with sensitivity and specificity of 100%. The accuracy of hand meander pattern features using PCA with RF was 87.36% and the specificity was 94.53%, and lastly, the accuracy of hand spiral pattern features using PCA with SVM was 84.73% and the specificity was 96.09%.

2.5 Indian Liver Disease

The fifth disease considered in this study was Indian Liver disease. It was addressed by many researchers, Rahman et al (Rahman et al., 2019) evaluated the performance of several machine learning algorithms to easily detect and diagnosis chronic liver disease. LR, KNN, DT, SVM, NB, and RF classifiers were used in this study. The results showed that the accuracy for the LR, RF, DT, SVM, KNN, and NB were 75%, 74%, 69%, 64%, 62%, and 53% respectively, where the LR exhibited the highest accuracy.

Joloudari et al (Joloudari et al., 2019) conducted a study on selecting worthy features by comparing various data mining methods to predict liver disease. Their approach was depending on extraction, loading, transformation, and analysis (ELTA) for better diagnosing. RF, Multi-Layer Perceptron (MLP) neural network, Bayesian networks, SVM, and particle swarm optimization (PSO)-SVM were the data mining tools used. The results showed that the average accuracies were 87.35%, 78.91%, 66.78%, 76.51%, and 95.17% for RF, MLP neural network, Bayesian networks, SVM, and PSO-SVM models respectively, where the PSO-SVM model had the highest performance.

Jacob et al (Jacob et al., 2018) displayed a study on how to identify liver patients from healthy individuals by using machine learning classification algorithms through a GUI to help the medicinal community easily diagnose liver disease patients. SVM, LR, KNN, and ANN were used as classifiers on a real dataset taken from UCI respiratory. The results showed that the accuracies were 73.23%, 72.05%, 75.04%, and 92.8 for LR, KNN, SVM, and ANN respectively, the ANN classifier had the best performance in diagnosing liver disease.

Abdar et al (Abdar et al., 2018) presented a method for early detecting liver disease through a Multilayer Perceptron Neural Network (MLPNN) algorithm based on various DT algorithms such as See5 (C5.0), chi-square automatic interaction detector (CHAID) algorithms, and classification and regression tree (CART) with boosting technique. The algorithms were employed on the ILPD which was taken from the UCI repository dataset. The study was divided into two experiments; in the first experiment, the results showed that the B-C5.0 method leads to giving better performance than B-CHAID and B-CART methods. In the second experiment, the results showed that the hybridization of B-C5.0 and MLPNN methods, namely MLPNNB-C5.0, suggested the highest rates of detecting liver disease as compared to the other algorithms.

Likewise, Abdar et al (Abdar et al., 2017) presented a study for predicting liver disease and compared it to other studies. The dataset used was taken from the UCI liver disease dataset and analyzed using two techniques namely, Boosted C5.0 and CHAID algorithms. The results showed high performance of accuracy for detecting liver disease. Another author, Durai et al (Durai, n.d.) proposed a method for analyzing a dataset to predict liver disease based on classification models. The classifiers' algorithms were applied to the liver disease dataset from the UCI repository. The results showed that the J48 algorithm had the highest accuracy with 95.04% when feature selection techniques were employed. Predicting liver diseases was also reported by Banu (Mehtaj Banu, 2019) who addressed an overview of special sorts of machine learning algorithms that had been used for predicting liver disease. The results of this overview showed that various distributions of predicting liver disease reached a short level of approval or testing for the diagnosing, where only about 15 to 25% of machine learning techniques could be used for expecting liver disease.

Kumar and Thakur (P. Kumar & Thakur, 2020) presented a method that aims to improve the performance of the imponderable's liver disease data classification. variable neighbor weighted fuzzy KNN approach (Variable-NWFKNN) was proposed to be used instead of using the Fuzzy KNN classifier because it's not working well on the imbalanced dataset due to its neighbor equally. This

method was employed on three real imbalanced liver disease datasets such as BUPA, ILPD from UCI, and MPRLPD. The results showed that the Variable-NWFKNN gives an accuracy of 73.91% utilizing BUPA Dataset, 77.59% for ILPD Dataset, and 87.01% for MPRLPD Dataset, where these accuracies were improved through Tomek Links Random Under-Sampling (TL RUS) method and the results obtained after improving the accuracy reached to 78.46% for BUPA Dataset, 78.46% for ILPD Dataset, and 95.79% for MPRLPD Dataset.

Srivenkatesh ("Performance Evolution of Different Machine Learning Algorithms for Prediction of Liver Disease," 2019) applied five different sorts of classification models namely NB, LR, SVM, KNN, and RF for the inspection of the liver malady and easy diagnosis. The results showed that the LR classifier had the best performance among other classifiers with Mean Absolute Error (MAE) of 0.23, Root Mean Square Error (RMSE) of 0.48, Relative Absolute Error (RAE) of 60.02%, Root Relative Squared Error (RRSR) of 38.83%, and the accuracy of 76.27%.

Chen Wu et al (Wu et al., 2019) developed a model utilizing machine learning to predict fatty liver disease (FLD) that may assist the medical community in diagnosing this kind of disease. RF, NB, ANN, and LR classifiers were used to develop and predict FLD. The results showed that the accuracies were 87.48%, 82.65%, 81.85%, and 76.96% for RF, NB, ANN, and LR respectively, where RF was the best with the highest accuracy.

El-Shafeiy et al (Saa & Al-emran, 2018) used machine learning algorithms for large datasets to predict liver disease and therapeutic discovery. The used classifiers were SVM, Boosted C5.0, NB, and data mining techniques where the most significant features were employed with the models to predict liver disease. The results showed an increase in accuracy, sensitivity, and specificity about 7.5% after the features selection process.

2.6 Epileptic Seizures Disease

The sixth disease considered in this study was Epileptic seizure disease. It was addressed by many researchers. Hussain (Hussain, 2018) made a deep analysis of different strategies for feature extraction based on time and frequency domain characteristics, wavelet-based entropy, nonlinear, and few statistical features. The features were used to perfectly diagnose and detection of epileptic seizures of individuals by using machine learning classifiers with more tuning parameters to enhancement the results. The results showed that the Kernel SVM with KNN with city block distance metric had the highest accuracy of 99.5% which was higher than the default parameters of the classifiers

Usman et al (Muhammad Usman et al., 2020) presented a technique for seizure prediction by using deep learning techniques. Their method can preprocess the EEG signals measured on the scalp, extract the main features automatically, and then apply CNN and SVM classifications. This classification was employed on 24 subjects and the performance results achieved were 92.7% and 90.8% for sensitivity and specificity, respectively.

Also, Usman et al (Usman et al., 2017) proposed a model that can provide reliable methods for both preprocessing and extracting features to predict the epileptic seizure disease in sufficient time before the starting of the seizure. empirical mode decomposition (EMD) was used for preprocessing and for features extraction in both time and frequency domain. The results showed a higher true positive rate reached 92.23% compared to the traditional methods with a maximum prediction time of 33 minutes.

Yang Si (Si, 2020) addressed a quick review for the signal processing of electroencephalograph (EEG) signals measured during epilepsy. Their algorithm was focusing on the applications of automatic detection for the epileptic seizure and the possible future work direction of machine learning techniques used in the analysis of EEG signals leading to the epileptic seizure detection.

Kumar and Kolekar (A. Kumar & Kolekar, 2014) presented a study that employs discrete wavelet transform for the analysis of the EEG signals. The EEG signals were decomposed into five different brainwaves (according to frequency bands) namely delta, theta, alpha, beta, and gamma waves. Only theta, alpha, and beta waves were carrying the information of the seizure. Statistical features were

extracted from each brainwave and fed into the SVM classifier. The results showed excellent accuracy, sensitivity, and specificity reached 98% compared to the existing methods.

Rasheed et al (Rasheed et al., 2020) provided a comprehensive review of the modern machine learning techniques for the early detection of seizures using EEG signals. The review involved many aspects like identifying and illustrating the challenges, gaps, difficulties in the current researches, future work, and the suggested techniques that may improve the accuracies.

Osman and Alzahrani (Osman & Alzahrani, 2018) proposed a method for automatic detection and diagnosing epilepsy seizure based on a self-organization map (SOM). Their technique employed a radial basis function (RBF) neural networks approach. This hybrid technique was tested on the UCI epilepsy dataset and aims to enhance the diagnosis of epilepsy without failing in the misdiagnosis. The results showed that the SOM-RBF method had accuracy with 10-fold cross-validation about 97.47% compared to the modern classification algorithms for predicting epileptic seizure in terms of the evaluation factor.

Siddiqui et al (Siddiqui et al., 2020) have presented an overview of the broad diversity of machine learning techniques over the last few years with the divided the classifiers into black-box' and 'non-black-box' and presenting ideas in details about seizure detection, classification, and the direction of researchers in the future for detecting epileptic seizure. A small result of this review shows that the decision forest is a perfect classifier in diagnosing epileptic seizures.

This research aims to provide a comprehensive comparison between different ML algorithms evaluated on six different healthcare datasets collected from the UCI machine learning repository. The datasets include chorionic kidney disease, breast cancer, heart diseases, Parkinson's disease, Indian liver disease, and epileptic seizure disease. Eight different ML algorithms are used to detect and diagnose these diseases. The utilized ML algorithms are logistic regression, k-nearest neighbor, support vector machine, naive bays, decision tree, random forest, XGBoost, and AdaBoost. The algorithms are evaluated using training and testing methodology with two separated percentages (70% - 30% and 80% - 20%) and then, the performance metrics are calculated using the confusion matrix of each classifier.

3. MATERIAL AND METHODS

In this section, we will discuss the datasets utilized in the study and our proposed methodology for comparing different classifiers' performances among different medical disease datasets. Figure 1 shows the block diagram for the flow of the proposed study.

3.1. Materials

Details about each dataset used in this research will be provided. The details contain several attributes, number of instances, and number of classes (labels) in the classes. Also, we will show details about the attributes collected from the study group from which the dataset was acquired such as age, health record, ... etc.

3.1.1. Chronic Kidney Disease Dataset

Chronic Kidney Diseases Dataset details, Table 1 shows a short description of the dataset to be fed to the classifiers while Table 2 shows the full details about the recorded (collected) attributes from the study group.

Figure 1. Proposed study block diagram



3.2. BREAST CANCER DETECTION DATASET

Breast Cancer Detection Dataset details, Table 3 shows a short description of the dataset to be fed to the classifiers while Table 4 shows the full details about the recorded (collected) attributes from the study group.

3.3. Heart Rate Prediction Dataset

Heart Rate Prediction Dataset details, Table 5 shows a short description of the dataset to be fed to the classifiers while Table 6 shows the full details about the recorded (collected) attributes from the study group.

3.4. Parkinson Disease Dataset

Parkinson Disease Dataset details, Table 7 shows a short description of the dataset to be fed to the classifiers while Table 8 shows the full details about the recorded (collected) attributes from the study group.

3.5. Indian Liver Disease Dataset

In this section we will discuss the Indian Liver Patients Dataset details, Table 9 shows a short description of the dataset to be fed to the classifiers while Table 10 shows the full details about the recorded (collected) attributes from the study group.

3.6. Epileptic Seizure Dataset

Epileptic Seizure Dataset details, Table 11 shows a short description of the dataset to be fed to the classifiers while Table 12 shows the full details about the recorded (collected) attributes from the study group.

Table 1. Data description

| Data used | Total attributes | Total instances | Total classes |
|------------------------|------------------|-----------------|---------------|
| Chronic Kidney Disease | 25 | 400 | 2 |

Table 2. Dataset attributes with their values

| Attributes | Values |
|-------------------------------------|----------------------------|
| Age | 2 – 90 |
| Blood pressure (mm/Hg) | 50 – 180 |
| Specific gravity | 1.005 – 1.025 |
| Albumin | 0 – 5 |
| Sugar | 0 – 5 |
| Red blood cells | 1: Normal, 0: Abnormal |
| Pus cell | 1: Normal, 0: Abnormal |
| Pus cell clumps | 1: Present, 0: Not present |
| Bacteria | 1: Present, 0: Not present |
| Blood glucose random (mgs/dl) | 22 – 490 |
| Blood urea (mgs/dl) | 1.5 – 391 |
| Serum creatinine (mgs/dl) | 0.4 – 76 |
| Sodium (mEq/L) | 4.5 – 163 |
| Potassium (mEq/L) | 2.5 – 47 |
| Hemoglobin (gms) | 3.1 – 17.8 |
| Packed cell volume | 9 – 54 |
| White blood cell count (cells/cumm) | 2200 – 26400 |
| Red blood cell count (millions/cmm) | 2.1 – 8 |
| Hypertension | 1: Yes, 0: No |
| Diabetes mellitus | 1: Yes, 0: No |
| Coronary artery disease | 1: Yes, 0: No |
| Appetite | 1: Yes, 0: No |
| Pedal edema | 1: Good, 0: Poor |
| Anemia | 1: Yes, 0: No |
| Class | 1: CKD, 0: Not CKD |

Table 3. Data description

| Data used | Total attributes | Total instances | Total classes |
|-------------------------|------------------|-----------------|---------------|
| Breast Cancer Detection | 32 | 569 | 2 |

3.7 Methodology

The proposed study includes the use of eight ML classification algorithms to predict and estimate the performance of six different medical diseases in humans. Two consecutive phases will be employed to diagnose the six different diseases. The phases are shown in Figure 2.

The First Phase: consist of the following procedure

Table 4. Dataset attributes information

| Attributes |
|---|
| Radius (mean of distances from the center to points on the perimeter) |
| Texture (standard deviation of gray-scale values) |
| Perimeter |
| Area |
| Smoothness (local variation in radius lengths) |
| Compactness ($\text{perimeter}^2 / \text{area} - 1.0$) |
| Concavity (severity of concave portions of the contour) |
| Concave points (number of concave portions of the contour) |
| Symmetry |
| Fractal dimension ("coastline approximation" - 1) |
| Diagnosis (M = malignant, B = benign) |

1. **Data Preprocessing:** After collecting the datasets for the six various medical diseases, most of the datasets contain some attributes with nominal data and numerical data, and other datasets have attributes with null values. The role-play of data preprocessing is a data mining technique that can transform the raw data into an obvious and understandable format.
 - a. **Data Transformation:** All of the nominal data are transformed into numerical digital data. For example, the normal is converted into 1 and the abnormal is converted into 0, etc.
 - b. **Missing Data Handling:** The null values of the attributes are replaced using the KNN imputation technique which works on data that identifies the neighboring points through the measure of the distance between them then, the missing values can be estimated using completed values of neighboring observations.
2. **Class Targeting:** To classify the patient whether to be normal or abnormal according to the values of the available features.
3. **Training & Testing Datasets Division:** The whole datasets are divided into two ratios; 70% - 30 and 80% - 20% training and testing sets. Where the training sets are used to train the models and the testing sets are used to test their performance.

The Second Phase: consist of the following procedure

1. **Applying Machine Learning Algorithms:** The main idea of the algorithms is to build ML models that could extradite input data. Then, statistically analyzing them to predict the output and to classify the disease as normal and abnormal with more accuracy. In this paper, eight different ML algorithms such as LR, KNN, SVM, NB, DT, RF, XGB, and ADB are applied to the medical datasets to predict and diagnose the presence of certain diseases.
2. **Training & Testing Datasets:** Each ML algorithm is trained using the training set, the learning algorithm finds patterns in the training data that map the input data attributes to the target which is

Table 5. Data description

| Data used | Total attributes | Total instances | Total classes |
|-----------------------|------------------|-----------------|---------------|
| Heart Rate Prediction | 14 | 303 | 2 |

Table 6. Dataset attributes description with their values

| Attributes | Values and Description |
|--|--|
| Age | 29 – 77 |
| Sex | 1 = male; 0 = female |
| Chest pain type | Value 1: typical angina Value 2: atypical angina Value 3: non-anginal pain |
| Resting blood pressure | 94 – 200 (mm/Hg) |
| Serum cholesterol | 126 – 564 (mg/dl) |
| Fasting blood sugar | 1: true, 0: false |
| Resting ECG | Value 0: normal Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria |
| Max heart rate achieved | 71 – 202 |
| Exercise induced angina | 1 = yes; 0 = no |
| ST depression induced by exercise relative to rest | 0 – 6.2 |
| The slope of the peak exercise ST segment | Value 1: upsloping Value 2: flat Value 3: downsloping |
| Number of major vessels (0-3) colored by fluoroscopy | 0 – 4 |
| Thalassemia | 3 = normal; 6 = fixed defect; 7 = reversable defect |
| Target [diagnosis of heart disease (angiographic disease status)] | Value 0: < 50% diameter narrowing Value 1: > 50% diameter narrowing |

Table 7. Data description

| Data used | Total attributes | Total instances | Total classes |
|-------------------|------------------|-----------------|---------------|
| Parkinson Disease | 23 | 195 | 2 |

the answer wanted to be predicted, then generating an ML model that can capture these patterns. The testing set is then applied to the ML models to examine their performances.

3. **Comparing the Results:** Accuracy, sensitivity, and specificity results, obtained from each ML model are compared.
4. **Best Model Selection:** After comparing the results from the eight different models, the most suitable model with the best accuracy, sensitivity, and specificity result is selected to predict and diagnose a patient's disease.

The outcomes of each disease will be presented in the results section with the difference between the two ratios 70% - 30% and 80% - 20% of training and testing separately.

Table 8. Dataset attributes with their description

| Abbreviation | Description |
|--|---|
| MDVP:Fo(Hz) | Average vocal fundamental frequency |
| MDVP:Fhi(Hz) | Maximum vocal fundamental frequency |
| MDVP:Flo(Hz) | Minimum vocal fundamental frequency |
| MDVP: Jitter (%), MDVP: Jitter (Abs), MDVP: RAP, MDVP: PPQ, Jitter: DDP | Several measures of variation in fundamental frequency |
| MDVP: Shimmer, MDVP: Shimmer (dB), Shimmer: APQ3, Shimmer: APQ5, MDVP: APQ, Shimmer: DDA | Several measures of variation in amplitude |
| NHR, HNR | Two measures of ratio of noise to tonal components in the voice |
| RPDE, D2 | Two nonlinear dynamical complexity measures |
| DFA | The signal fractal scaling exponent |
| Spread1, spread2, PPE | Three nonlinear measures of fundamental frequency variation |
| Status | Health status of the subject: 1: Parkinson's, 0: healthy |

Table 9. Data description

| Data used | Total attributes | Total instances | Total classes |
|------------------------------|------------------|-----------------|---------------|
| Indian Liver Disease Dataset | 11 | 583 | 2 |

Table 10. Dataset attributes with their values

| Attributes | Values |
|----------------------------|------------------------|
| Age | 4 – 90 |
| Gender | 0: Female, 1: Male |
| Total Bilirubin | 0.4 – 75 |
| Direct Bilirubin | 0.1 – 19.7 |
| Alkaline Phosphatase | 63 – 2110 |
| Alamine Aminotransferase | 10 – 2000 |
| Aspartate Aminotransferase | 10 – 4929 |
| Total Proteins | 2.7 – 9.6 |
| Albumin | 0.9 – 9.9 |
| Albumin and Globulin Ratio | 0.3 – 2.8 |
| Class | 1: Normal, 2: Abnormal |

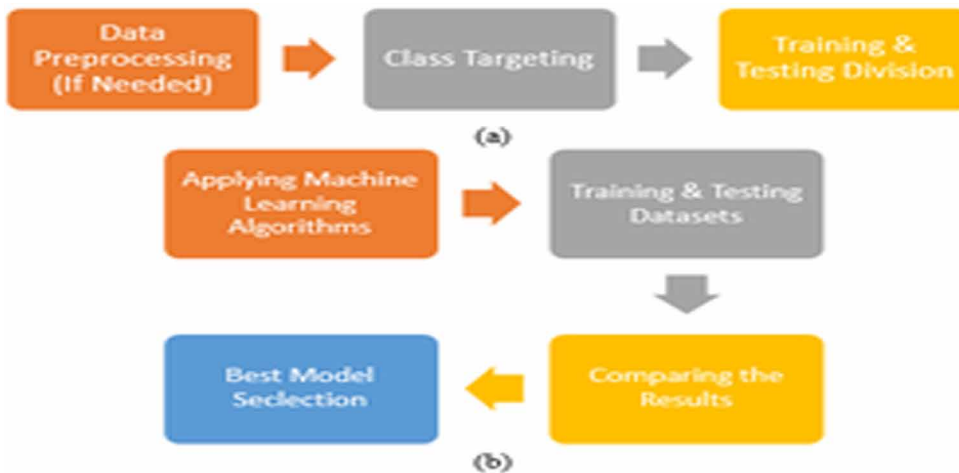
Table 11. Data description

| Data used | Total attributes | Total instances | Total classes |
|---------------------------|------------------|-----------------|---------------|
| Epileptic Seizure Dataset | 179 | 11500 | 2 |

Table 12. Dataset attributes with their values

| Attributes | Values |
|--|-------------------|
| Recording Duration | 1 Sec |
| The dimension of Explanatory variables | 178 Variables |
| Record Location | Above Brain Tumor |

Figure 2. (a) First Phase Block Diagram, (b) Second Phase Block Diagram



4. RESULTS

This section will show the details about applying our methodology to the mentioned datasets. Therefore, Table 13 shows details about the dataset division for training and testing phases using two scenarios 70% - 30% and 80% - 20%.

For chronic kidney disease, Figure 3 shows the results for dataset splitting ratios of 70% - 30% training and testing accuracy, sensitivity, and specificity obtained from eight different ML algorithms while Figure 4 shows the results for dataset splitting ratios of 80% - 20% training and testing accuracy, sensitivity, and specificity and 80%-20% obtained from eight different ML algorithms.

For breast cancer disease, Figure 5 shows the results for dataset splitting ratios of 70% - 30% training and testing accuracy, sensitivity, and specificity obtained from eight different ML algorithms while Figure 6 shows the results for dataset splitting ratios of 80% - 20% training and testing accuracy, sensitivity, and specificity and 80%-20% obtained from eight different ML algorithms.

For heart rate disease, Figure 7 shows the results for dataset splitting ratios of 70% - 30% training and testing accuracy, sensitivity, and specificity obtained from eight different ML algorithms while

Table 13. Training and testing division dataset

| Dataset | Training set 70% | Testing set 30% | Training set 80% | Testing set 20% |
|----------------------------|------------------|-----------------|------------------|-----------------|
| Chronic Kidney Disease | 280 | 120 | 320 | 80 |
| Breast Cancer Disease | 398 | 171 | 455 | 114 |
| Heart Rate Disease | 212 | 91 | 242 | 61 |
| Parkinson Disease | 136 | 59 | 156 | 39 |
| Indian Liver Disease | 408 | 175 | 466 | 117 |
| Epileptic Seizures Disease | 8050 | 3450 | 9200 | 2300 |

Figure 3. Performance evaluation of 70% - 30% training and testing for chronic kidney disease

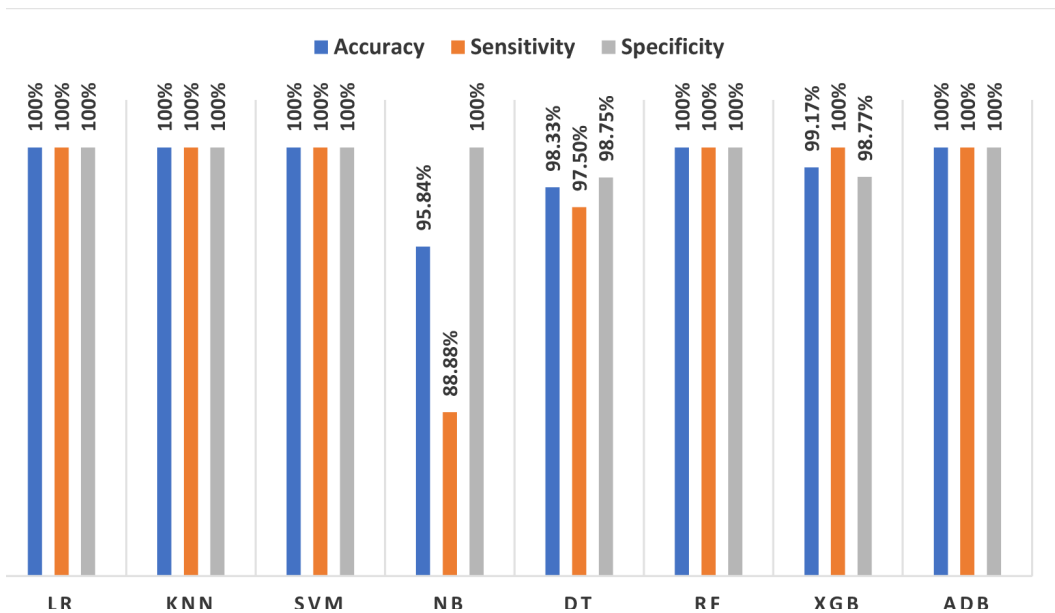


Figure 8 shows the results for dataset splitting ratios of 80% - 20% training and testing accuracy, sensitivity, and specificity and 80%-20% obtained from eight different ML algorithms.

For Parkinson disease, Figure 9 shows the results for dataset splitting ratios of 70% - 30% training and testing accuracy, sensitivity, and specificity obtained from eight different ML algorithms while Figure 10 shows the results for dataset splitting ratios of 80% - 20% training and testing accuracy, sensitivity, and specificity and 80%-20% obtained from eight different ML algorithms.

For Indian liver disease, Figure 11 shows the results for dataset splitting ratios of 70% - 30% training and testing accuracy, sensitivity, and specificity obtained from eight different ML algorithms while Figure 12 shows the results for dataset splitting ratios of 80% - 20% training and testing accuracy, sensitivity, and specificity and 80%-20% obtained from eight different ML algorithms.

Figure 4. Performance evaluation of 80% - 20% training and testing for chronic kidney disease

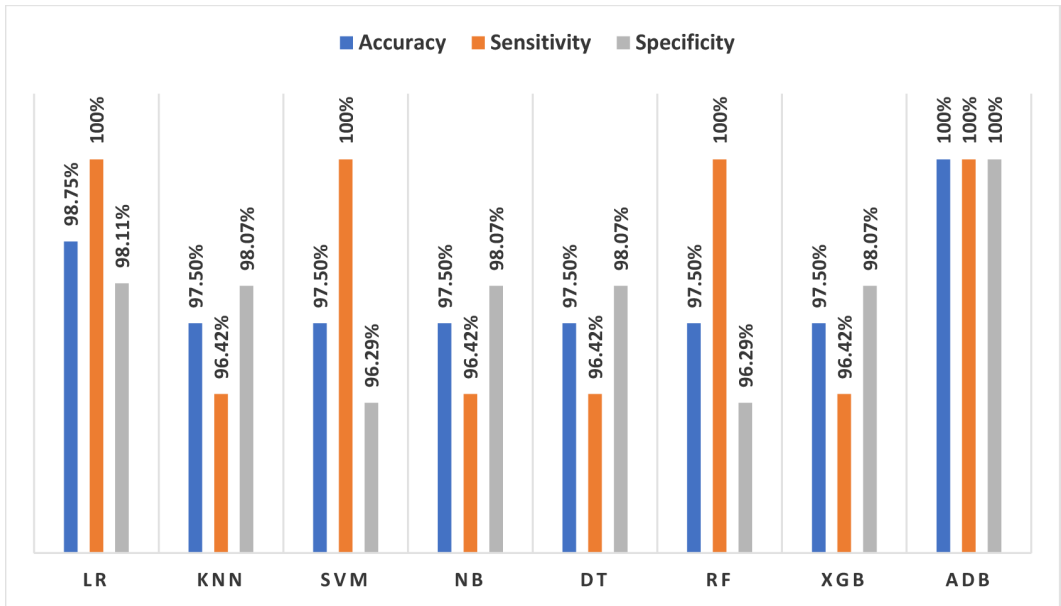


Figure 5. Performance evaluation of 70% - 30% training and testing for breast cancer disease

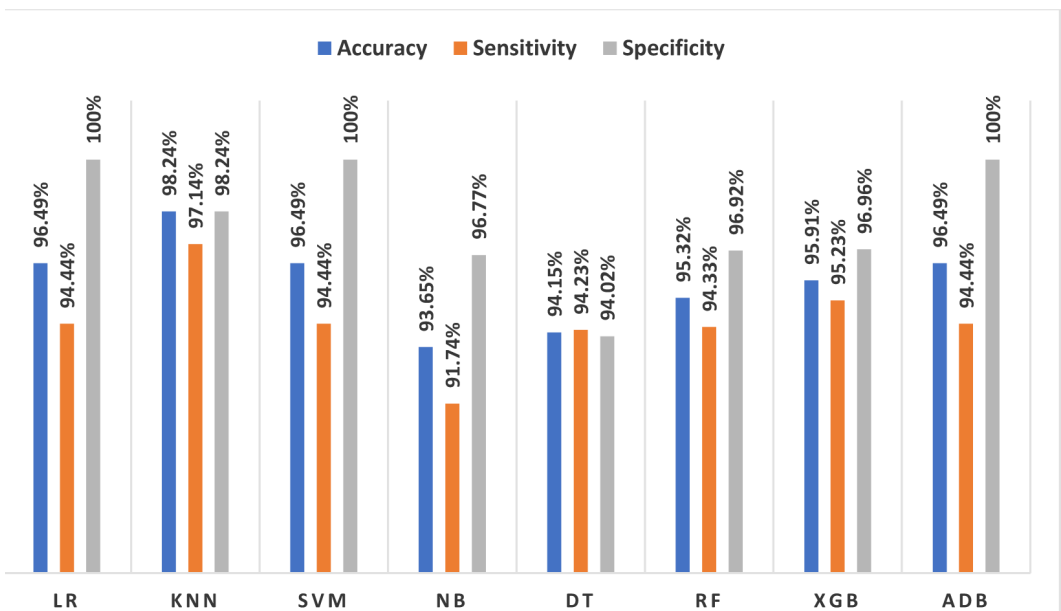


Figure 6. Performance evaluation of 80% - 20% training and testing for breast cancer disease

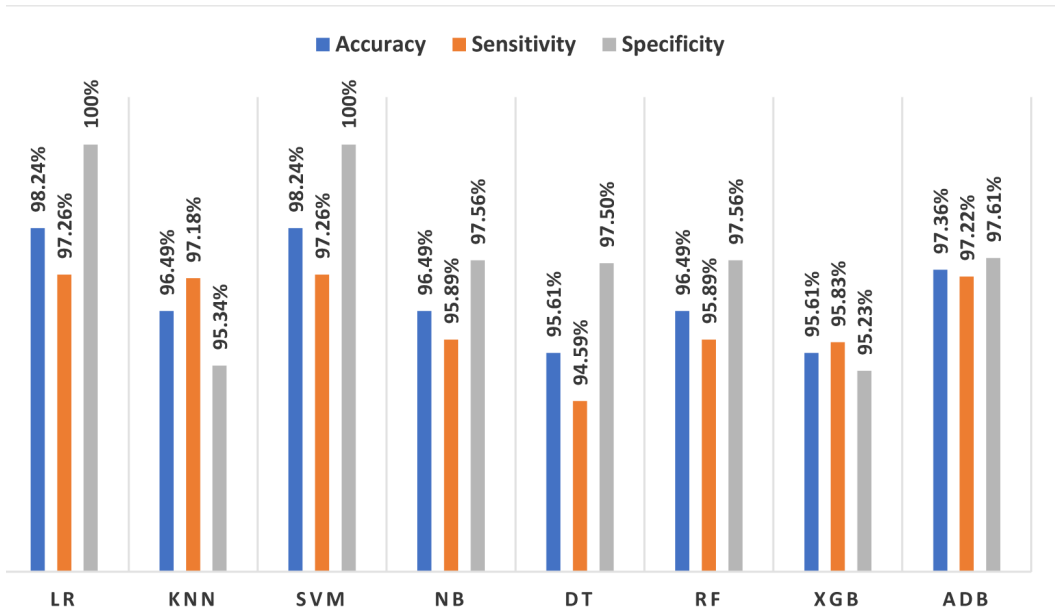


Figure 7. Performance evaluation of 70% - 30% training and testing for heart rate disease

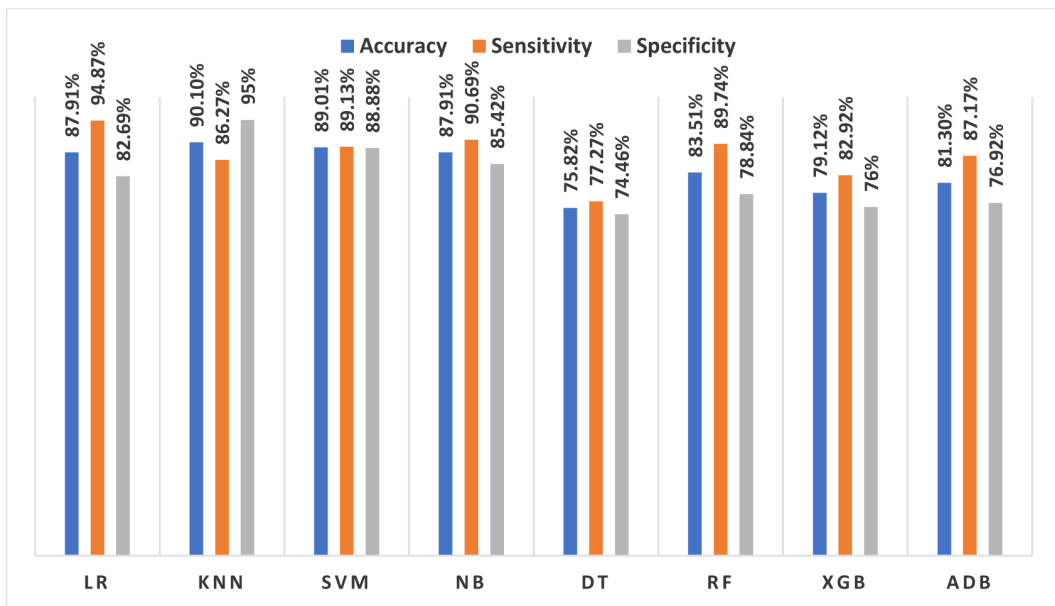


Figure 8. Performance evaluation of 80% - 20% training and testing for heart rate disease

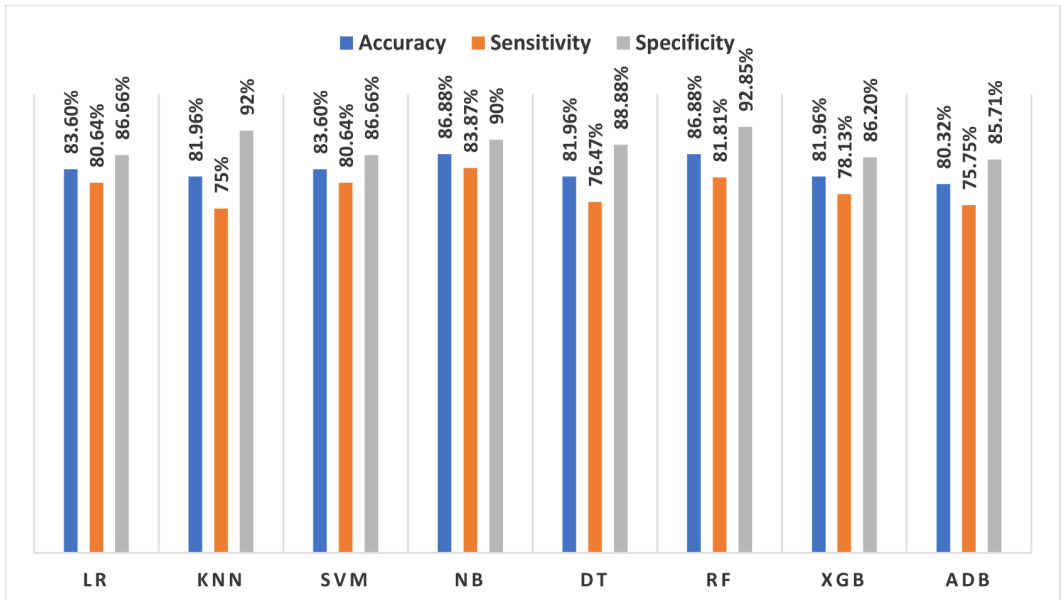
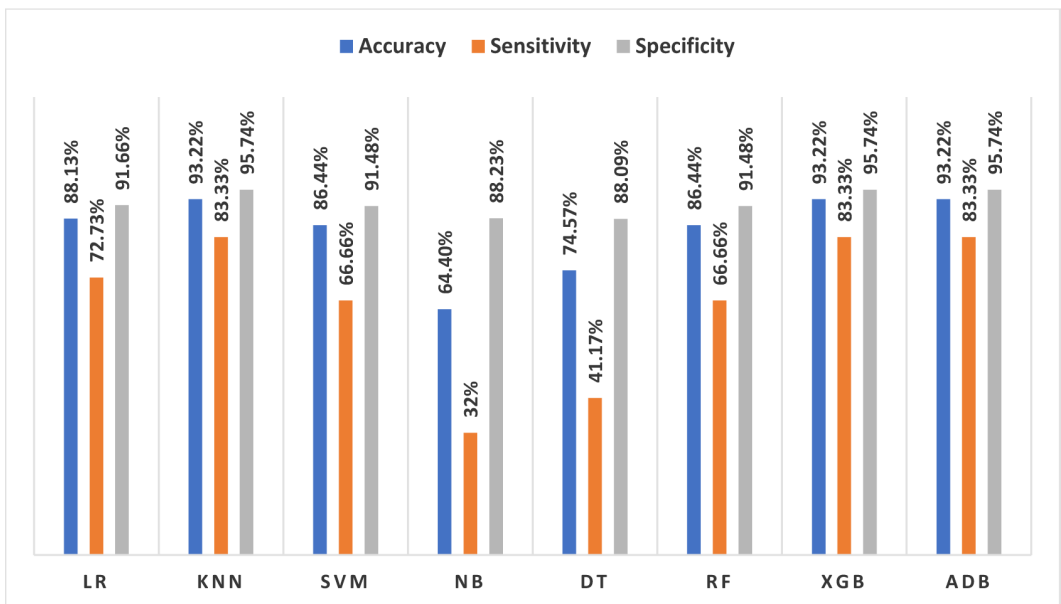


Figure 9. Performance evaluation of 70%-30% training and testing for Parkinson disease



For epileptic seizures disease, Figure 13 shows the results for dataset splitting ratios of 70% - 30% training and testing accuracy, sensitivity, and specificity obtained from eight different ML algorithms while Figure 14 shows the results for dataset splitting ratios of 80% - 20% training and testing accuracy, sensitivity, and specificity and 80%-20% obtained from eight different ML algorithms.

Figure 10. Performance evaluation of 80%-20% training and testing for Parkinson disease

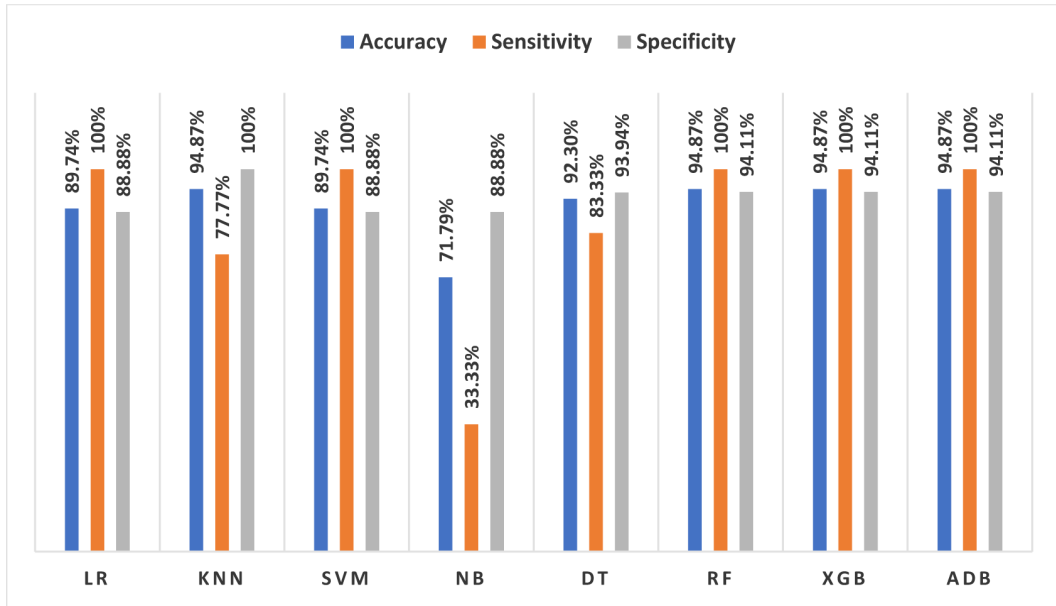
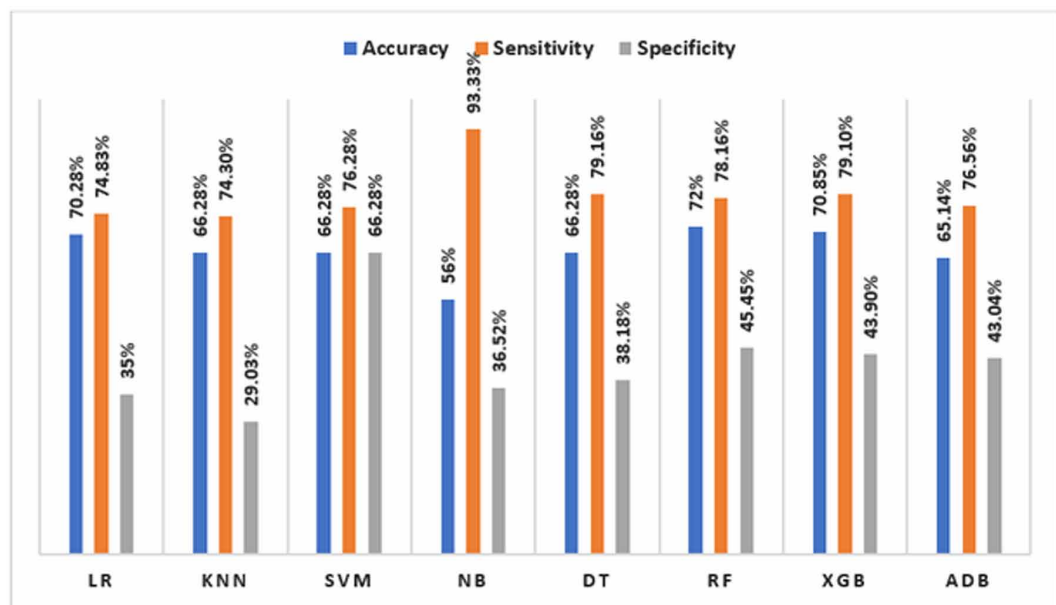


Figure 11. Performance evaluation of 70% - 30% training and testing for Indian liver disease



5. DISCUSSION

After collecting and processing data, the six different datasets from UCI, the eight machine learning algorithms have been applied, the True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN) has been calculated. Based on these values, the performance of each classifier

Figure 12. Performance evaluation of 80% - 20% training and testing for Indian liver disease

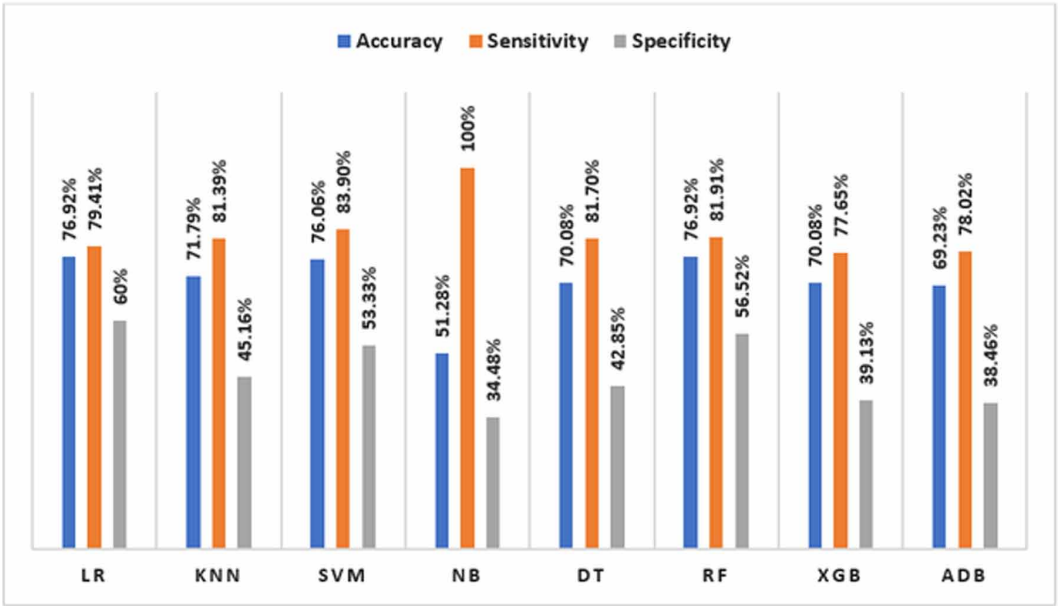


Figure 13. Performance evaluation of 70% - 30% training and testing for epileptic seizures disease

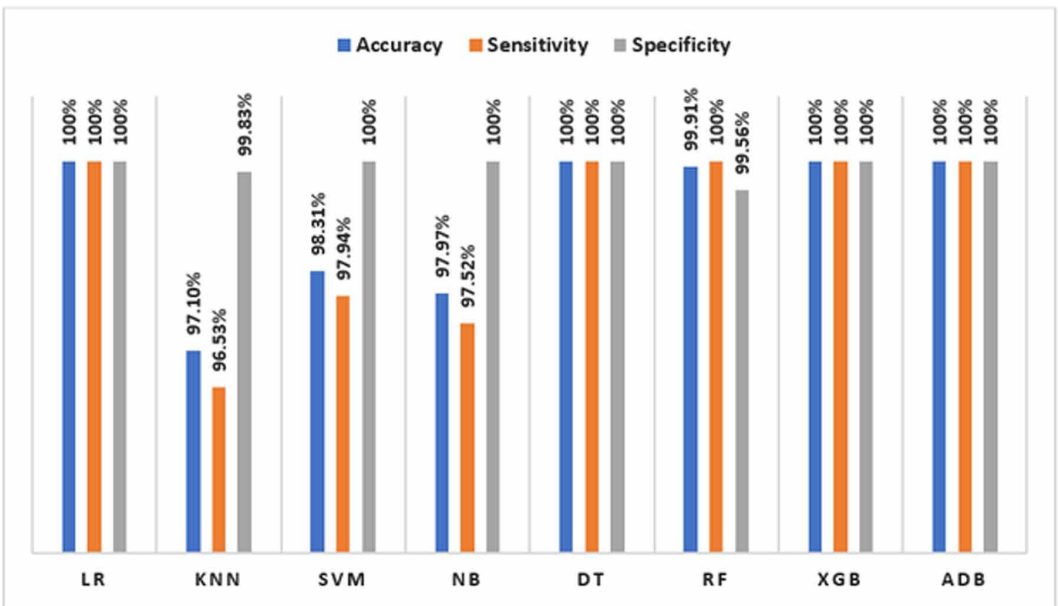
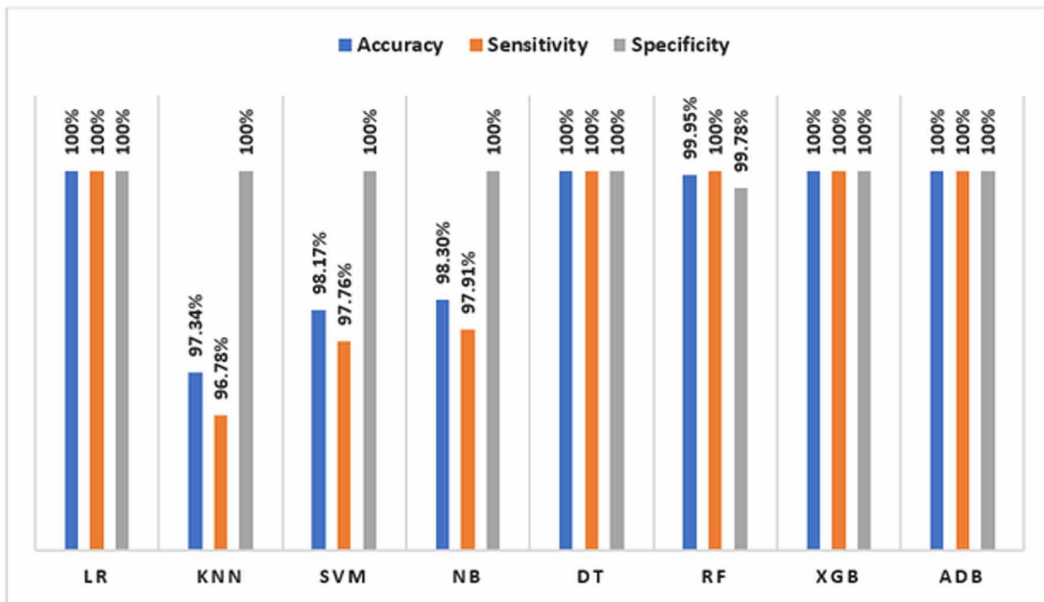


Figure 14. Performance evaluation of 80% - 20% training and testing for epileptic seizures disease



has been evaluated using Accuracy, Sensitivity, and Specificity. Using the mentioned performance metrics each classifier evaluated in all datasets using the two ratios (70% training and 30% testing) and (80% training and 20% testing).

For the chronic kidney disease dataset, based on the results section, the results show that the best classifier among all was the ADB classifier according to both ratios as overall in both training and testing dataset. While, for breast cancer disease, using the results in the Results section, it is noticed that the KNN classifier is the best in the 70%:30% ratio, but LR and SVM classifiers give a better result in the ratio of 80%:20% training and testing, however in overall, both LR and SVM are the best classifiers in diagnosing kidney diseases.

For the heart rate disease dataset, established by the results in the Results section, it is remarked that the KNN classifier is the best if the ratio 70%:30% is used, while the RF classifier is better in the ratio of 80%:20%, However, the NB classifier is the best algorithm in predicting the disease as overall. On the other hand, for Parkinson’s disease, based on the results in the Results section, it is shown that the XGB and ADB among all algorithms used are the best classifiers as overall in both utilized ratios.

For liver disease, as the results are shown in the results section, the SVM classifier is the best in the ratio of 70%:30%, while LR gives a better result in the ratio of 80%:20%. Overall, the SVM classifier is the best in classifying the disease in both ratios. Furthermore, for epileptic seizures, with the results shown in the results section, it is observed that both XGB and ADB classifiers are giving better prediction results overall in both ratios training and testing.

To make a comprehensive comparison between all used classifiers, the prediction time for all testing patients in the dataset has been recorded for each classifier among all used datasets and it is shown in Table 14. It is noticed that the most classifiers recorded very short time (in milliseconds) on average among all testing patients while the shortest time is for XGBoost, logistic regression, and k-nearest neighbor respectively when comparing these classifiers time to other used classifiers.

Finally, overall as we noticed from the results section, we can rank the used classifiers based on their performance among all datasets in the following order: the first rank is for the XGB and

Table 14. Prediction time in milliseconds per-patient for all classifiers on the used datasets

| Dataset | Classifier | | | | | | | |
|----------------------------|---------------------|--------------------|------------------------|------------|---------------|---------------|---------|----------|
| | Logistic Regression | K Nearest Neighbor | Support Vector Machine | Naive Bays | Decision Tree | Random Forest | XGBoost | AdaBoost |
| Chronic Kidney Disease | 0.611 | 5.083 | 4.85 | 2.273 | 0.101 | 10.291 | 0.886 | 8.584 |
| Breast Cancer Disease | 2.294 | 7.027 | 1.975 | 14.471 | 12.39 | 12.616 | 0.83 | 11.354 |
| Heart Rate Disease | 1.779 | 6.028 | 4.118 | 15.954 | 5.543 | 1.395 | 0.977 | 7.686 |
| Parkinson Disease | 0.268 | 7.037 | 5.659 | 3.056 | 11.327 | 15.287 | 0.984 | 11.72 |
| Indian Liver Disease | 2.801 | 13.646 | 10.615 | 12.631 | 2.797 | 14.924 | 0.841 | 11.595 |
| Epileptic Seizures Disease | 0.862 | 3.348 | 3.11 | 8.427 | 3.769 | 11.245 | 0.766 | 8.101 |
| Average | 1.436 | 7.028 | 5.054 | 9.469 | 5.988 | 10.96 | 0.881 | 9.84 |

LR then the rest of the classifiers are close to each other in performance and may be considered in the second rank. So, we can use the XGB and LR classifiers more widely on medical classification problems, and in any future research, we can use them as a starting point.

6. CONCLUSION AND FUTURE WORK

In this paper, a study is conducted where a comparison between eight different classifiers are evaluated using six different medical datasets. The research is aimed toward finding the best classifier(s) that able to diagnose different medical conditions with high performances. The best classifiers were XGB and LR classifiers among all used datasets in terms of all used performance metrics (Accuracy, Sensitivity, Specificity, etc..). The findings of this research can be reflected in any future researches by selecting a certain set of classifiers as a starting point in any medical decision-making system. This will make researchers focus on part of classifiers instead of focusing on all classifiers type. While future work includes increasing the number of classifiers and using big datasets to make a comprehensive comparison between all available classifiers besides developing a graphical user interface (GUI) that shows all classifiers results and making an ensemble decision-making system.

REFERENCES

- Abbas, S. A., Rehman, A. U., Majeed, F., Majid, A., Malik, M. S. A., Kazmi, Z. H., & Zafar, S. (2020). Performance Analysis of Classification Algorithms on Birth Dataset. *IEEE Access : Practical Innovations, Open Solutions*, 8, 102146–102154. doi:10.1109/ACCESS.2020.2999899
- Abdar, M., Yen, N. Y., & Hung, J. C. S. (2018). Improving the Diagnosis of Liver Disease Using Multilayer Perceptron Neural Network and Boosted Decision Trees. *Journal of Medical and Biological Engineering*, 38(6), 953–965. doi:10.1007/s40846-017-0360-z
- Abdar, M., Zomorodi-Moghadam, M., Das, R., & Ting, I. H. (2017). Performance analysis of classification algorithms on early detection of liver disease. *Expert Systems with Applications*, 67, 239–251. doi:10.1016/j.eswa.2016.08.065
- Abdel-Zaher, A. M., & Eldeib, A. M. (2016). Breast cancer classification using deep belief networks. *Expert Systems with Applications*, 46, 139–144. doi:10.1016/j.eswa.2015.10.015
- Agarap, A. F. M. (2018). On breast cancer detection: An application of machine learning algorithms on the Wisconsin diagnostic dataset. *ACM International Conference Proceeding Series*, 1, 5–9. doi:10.1145/3184066.3184080
- Ak, M. F. (2020). A Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualization and Machine Learning Applications. *Health Care*, 8(2), 111. doi:10.3390/healthcare8020111 PMID:32357391
- Al Imran, A., Amin, M. N., & Johora, F. T. (2019). 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 2018*, 1–6. doi:10.1109/CIET.2018.8660844
- Ansari, M. F., AlankarKaur, B., & Kaur, H. (2021). A prediction of heart disease using machine learning algorithms. *Advances in Intelligent Systems and Computing*, 1200 AISC, 497–504. 10.1007/978-3-030-51859-2_45
- Aruna, S., Rajagopalan, S. P., & Nandakishore, L. V. (2011). Knowledge Based Analysis of Various Statistical Tools in Detecting Breast Cancer. *Computer Science and Information Technology*, 2, 37–45. doi:10.5121/csit.2011.1205
- Asri, H., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis. *Procedia Computer Science*, 83(Fams), 1064–1069. 10.1016/j.procs.2016.04.224
- B, G. S., Mittal, A., & Aggarwal, N. (2019). *Denoiser for Mammographic Images*. Springer Singapore. 10.1007/978-981-13-9939-8
- Basar, M. D., & Akan, A. (2018). Detection of chronic kidney disease by using ensemble classifiers. *2017 10th International Conference on Electrical and Electronics Engineering, ELECO 2017*, 544–547.
- Bhardwaj, A., & Tiwari, A. (2015). Breast cancer diagnosis using Genetically Optimized Neural Network model. *Expert Systems with Applications*, 42(10), 4611–4620. doi:10.1016/j.eswa.2015.01.065
- Charleonnann, A., Fufaung, T., Niyomwong, T., Chokchueypattanakit, W., Suwannawach, S., & Ninchawee, N. (2017). Predictive analytics for chronic kidney disease using machine learning techniques. *2016 Management and Innovation Technology International Conference, MITiCON 2016*, MIT80–MIT83. doi:10.1109/MITiCON.2016.8025242
- Chen, A. H., Huang, S. Y., Hong, P. S., Cheng, C. H., & Lin, E. J. (2011). HDPS: Heart disease prediction system. *Computers in Cardiology*, 38, 557–560.
- Das, R. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Systems with Applications*, 37(2), 1568–1572. doi:10.1016/j.eswa.2009.06.040
- De, S., & Chakraborty, B. (2020). Development of Chronic Kidney Disease Prediction System (CKDPS) Using Machine Learning Technique. *Lecture Notes on Data Engineering and Communications Technologies*, 10, 155–164. doi:10.1007/978-3-030-34080-3_18

- Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: A comparison of three data mining methods. *Artificial Intelligence in Medicine*, 34(2), 113–127. doi:10.1016/j.artmed.2004.07.002 PMID:15894176
- Devika, R., Avilala, S. V., & Subramaniaswamy, V. (2019). Comparative study of classifier for chronic kidney disease prediction using naive bayes, KNN and random forest. *Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC 2019, Iccmc*, 679–684. doi:10.1109/ICCMC.2019.8819654
- Durai, V. (n.d.). *Liver disease prediction using machine learning*. Academic Press.
- Eleyan, A. (2012). Breast cancer classification using moments. *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting, 1–4*, 1–4. Advance online publication. doi:10.1109/SIU.2012.6204778
- Ella, A., & Deepak, H. (2018). *International Conference on Innovative Computing and Communications (Vol. 2)*. Academic Press.
- Golande, A., & Pavan Kumar, T. (2019). Heart disease prediction using effective machine learning techniques. *International Journal of Recent Technology and Engineering*, 8(1), 944–950.
- Gupta, D., Julka, A., Jain, S., Aggarwal, T., Khanna, A., Arunkumar, N., & de Albuquerque, V. H. C. (2018). Optimized cuttlefish algorithm for diagnosis of Parkinson's disease. *Cognitive Systems Research*, 52, 36–48. doi:10.1016/j.cogsys.2018.06.006
- Haribaabu, V., Sivakumar, V., Selvakumarasamy, & Dixit, V. (2018). Prediction of heart disease risk using machine learning. *International Journal of Mechanical and Production Engineering Research and Development*, 8(2), 605–614. 10.24247/ijmperdfeb201867
- Harper, P. R. (2005). A review and comparison of classification algorithms for medical decision making. *Health Policy (Amsterdam)*, 71(3), 315–331. doi:10.1016/j.healthpol.2004.05.002 PMID:15694499
- Hussain, L. (2018). Detecting epileptic seizure with different feature extracting strategies using robust machine learning classification techniques by applying advance parameter optimization approach. *Cognitive Neurodynamics*, 12(3), 271–294. doi:10.1007/s11571-018-9477-1 PMID:29765477
- Jacob, J., Chakkalal Mathew, J., Mathew, J., & Issac, E. (2018). Diagnosis of Liver Disease Using Machine Learning Techniques. *International Research Journal of Engineering and Technology*, 5(4), 4011–4014. www.irjet.net
- Jain, A., Ratnoo, S., & Kumar, D. (2018). Performance comparison of classification algorithms for medical diagnosis. *Pertanika Journal of Science & Technology*, 26(2), 729–748.
- Jain, D., & Singh, V. (2020). A novel hybrid approach for chronic disease classification. *International Journal of Healthcare Information Systems and Informatics*, 15(1), 1–19. doi:10.4018/IJHISI.2020010101
- Joloudari, J. H., Saadatfar, H., Dehzangi, A., & Shamshirband, S. (2019). Computer-aided decision-making for predicting liver disease using PSO-based optimized SVM with feature selection. *Informatics in Medicine Unlocked*, 17, 100255. doi:10.1016/j.imu.2019.100255
- Karapinar Senturk, Z. (2020). Early diagnosis of Parkinson's disease using machine learning algorithms. *Medical Hypotheses*, 138(January), 109603. doi:10.1016/j.mehy.2020.109603 PMID:32028195
- Kavitha, R., & Kannan, E. (2016). An efficient framework for heart disease classification using feature extraction and feature selection technique in data mining. *1st International Conference on Emerging Trends in Engineering, Technology and Science, ICETETS 2016 - Proceedings*. doi:10.1109/ICETETS.2016.7603000
- Kotturu, P., Sasank, V. V. S., Supriya, G., Manoj, C. S., & Maheshwarredy, M. V. (2019). Prediction of chronic kidney disease using machine learning techniques. *International Journal of Advanced Science and Technology*, 28(16), 1436–1443. doi:10.17148/IJARCC.2018.71021
- Kumar, A., & Kolekar, M. H. (2014). Machine learning approach for epileptic seizure detection using wavelet analysis of EEG signals. *2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems, MedCom 2014*, 412–416. doi:10.1109/MedCom.2014.7006043

Kumar, P., & Thakur, R. S. (2020). Liver disorder detection using variable- neighbor weighted fuzzy K nearest neighbor approach. *Multimedia Tools and Applications*. Advance online publication. doi:10.1007/s11042-019-07978-3

Latha, C. B. C., & Jeeva, S. C. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Informatics in Medicine Unlocked*, 16(July), 100203. doi:10.1016/j.imu.2019.100203

Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., Mahendiran, T., Moraes, G., Shamdas, M., Kern, C., Ledsam, J. R., Schmid, M. K., Balaskas, K., Topol, E. J., Bachmann, L. M., Keane, P. A., & Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. *The Lancet. Digital Health*, 1(6), e271–e297. doi:10.1016/S2589-7500(19)30123-2 PMID:33323251

Mandal, I., & Sairam, N. (2014). New machine-learning algorithms for prediction of Parkinson's disease. *International Journal of Systems Science*, 45(3), 647–666. doi:10.1080/00207721.2012.724114

Mehtaj Banu, H. (2019). Liver disease prediction using machine-learning algorithms. *International Journal of Engineering and Advanced Technology*, 8(6), 2532–2534. doi:10.35940/ijeat.F8365.088619

Muhammad Usman, S., Khalid, S., & Aslam, M. H. (2020). Epileptic Seizures Prediction Using Deep Learning Techniques. *IEEE Access : Practical Innovations, Open Solutions*, 8, 39998–40007. doi:10.1109/ACCESS.2020.2976866

Omondigbe, D. A., Veeramani, S., & Sidhu, A. S. (2019). Machine Learning Classification Techniques for Breast Cancer Diagnosis. *IOP Conference Series. Materials Science and Engineering*, 495(1), 012033. Advance online publication. doi:10.1088/1757-899X/495/1/012033

Osman, A. H., & Alzahrani, A. A. (2018). New Approach for Automated Epileptic Disease Diagnosis using an Integrated Self- Organization Map and Radial Basis Function Neural Network Algorithm. *IEEE Access*, 1. 10.1109/ACCESS.2018.2886608

Oyewola, D., Hakimi, D., Adeboye, K., & Shehu, M. D. (2017). Using Five Machine Learning for Breast Cancer Biopsy Predictions Based on Mammographic Diagnosis. *International Journal of Engineering Technologies IJET*, 2(4), 142–145. doi:10.19072/ijet.280563

Pahwa, K., & Kumar, R. (2017). Prediction of heart disease using hybrid technique for selecting features. *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics, UPCON 2017*, 500–504. doi:10.1109/UPCON.2017.8251100

Pandey, K. K., Shukla, D., & Milan, R. (2020). A Comprehensive Study of Clustering Algorithms for Big Data Mining with MapReduce Capability. In *Lecture Notes in Networks and Systems* (Vol. 100, pp. 427–440). doi:10.1007/978-981-15-2071-6_34

Performance Evolution of Different Machine Learning Algorithms for Prediction of Liver Disease. (2019). *International Journal of Innovative Technology and Exploring Engineering*, 9(2), 1115–1122. doi:10.35940/ijtee.L3619.129219

Rahman, A. K. M. S., Javed Mehedi Shamrat, F. M., Tasnim, Z., Roy, J., & Hossain, S. A. (2019). A comparative study on liver disease prediction using supervised machine learning algorithms. *International Journal of Scientific and Technology Research*, 8(11), 419–422.

Rasheed, K., Qayyum, A., Qadir, J., Sivathamboo, S., Kwan, P., Kuhlmann, L., O'Brien, T., & Razi, A. (2020). Machine Learning for Predicting Epileptic Seizures Using EEG Signals. *RE:view*, ●●●, 1–15. doi:10.1109/RBME.2020.3008792 PMID:32746369

Revathy, S., Bharathi, B., Jeyanthi, P., & Ramesh, M. (2019). Chronic kidney disease prediction using machine learning models. *International Journal of Engineering and Advanced Technology*, 9(1), 6364–6367. doi:10.35940/ijeat.A2213.109119

Rusz, J., Cmejla, R., Ruzickova, H., & Ruzicka, E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *The Journal of the Acoustical Society of America*, 129(1), 350–367. doi:10.1121/1.3514381 PMID:21303016

- Saa, A. A., & Al-emran, M. (2018). Mining Student Information System Records to Predict Students'. *Academic Performance Amjad.*, 723(January), 229–239. doi:10.1007/978-3-319-74690-6
- Si, Y. (2020). Machine learning applications for electroencephalograph signals in epilepsy: A quick review. *Acta Epileptologica*, 2(1), 5–11. doi:10.1186/s42494-020-00014-0
- Siddiqui, M. K., Morales-Menendez, R., Huang, X., & Hussain, N. (2020). A review of epileptic seizure detection using machine learning classifiers. *Brain Informatics*, 7(1), 1–18. doi:10.1186/s40708-020-00105-1 PMID:32451639
- Sisodia, D. S., & Verma, A. (2018). Prediction performance of individual and ensemble learners for chronic kidney disease. *Proceedings of the International Conference on Inventive Computing and Informatics, ICICI 2017, Icici*, 1027–1031. doi:10.1109/ICICI.2017.8365295
- Sowmiya, C., & Sumitra, P. (2018). Analytical study of heart disease diagnosis using classification techniques. *Proceedings of the 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing, INCOS 2017*, 1–5. doi:10.1109/ITCOSP.2017.8303115
- Subas, Abdulhamit, & Alickovic, Emina, and J. K. (. (2017). Diagnosis of Chronic Kidney Disease by Using Random Forest Abdulhamit. *Springer Nature Singapore*, 6. Advance online publication. doi:10.1007/978-981-10-4166-2
- Swain, D., Ballal, P., Dolase, V., Dash, B., & Santhappan, J. (2020). An efficient heart disease prediction system using machine learning. In *Advances in Intelligent Systems and Computing* (Vol. 1101). doi:10.1007/978-981-15-1884-3_4
- Tazin, N., Sabab, S. A., & Chowdhury, M. T. (2017). Diagnosis of chronic kidney Disease using effective classification and feature selection technique. *1st International Conference on Medical Engineering, Health Informatics and Technology, MediTec 2016*, 1–6. doi:10.1109/MEDITEC.2016.7835365
- Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making*, 19(1), 1–16. doi:10.1186/s12911-019-1004-8 PMID:31864346
- Usman, S. M., Usman, M., & Fong, S. (2017). Epileptic Seizures Prediction Using Machine Learning Methods. *Computational and Mathematical Methods in Medicine*, 2017, 1–10. Advance online publication. doi:10.1155/2017/9074759 PMID:29410700
- Vijayarani, S., Dhayanand, S., Professor, A., & Research Scholar, M. P. (2015). Kidney Disease Prediction Using Svm and Ann Algorithms. *International Journal of Computing and Business Research*, 6(2), 2229–6166.
- W.H.S.D, G. (2017). Performance Evaluation on Machine Learning Classification Techniques for Disease (CKD). *IEEE*, 291–296. 10.1109/BIBE.2017.00056
- Wibawa, H. A., Malik, I., & Bahtiar, N. (2019). Evaluation of Kernel-Based Extreme Learning Machine Performance for Prediction of Chronic Kidney Disease. *2018 2nd International Conference on Informatics and Computational Sciences, ICICoS 2018*, 33–36. doi:10.1109/ICICoS.2018.8621762
- Williams, K., Adebayo Idowu, P., Ademola Balogun, J., & Ishola Oluwaranti, A. (2015). Breast Cancer Risk Prediction Using Data Mining Classification Techniques. *Transactions on Networks and Communications*, 3(2). Advance online publication. doi:10.14738/tnc.32.662
- Wroge, T. J., Özkanca, Y., Demiroglu, C., Si, D., Atkins, D. C., & Ghomi, R. H. (2019). Parkinson's Disease Diagnosis Using Machine Learning and Voice. *2018 IEEE Signal Processing in Medicine and Biology Symposium, SPMB 2018 - Proceedings*, 1–7. doi:10.1109/SPMB.2018.8615607
- Wu, C. C., Yeh, W. C., Hsu, W. D., Islam, M. M., & Nguyen, P. A., Poly, T. N., Wang, Y. C., Yang, H. C., & Li, Y. C. (2019). Prediction of fatty liver disease using machine learning algorithms. *Computer Methods and Programs in Biomedicine*, 170(January), 23–29. doi:10.1016/j.cmpb.2018.12.032 PMID:30712601

Ali Mohammad Alqudah received his BSc and MSc from Yarmouk University in 2015 and 2018 respectively. His research area is in the field of Biomedical Signal Processing, Image Processing and Analysis, Deep Learning, and Machine Learning. Currently he is a Teaching and Research Assistant at Department of Biomedical Systems and Informatics Engineering, Yarmouk University, Irbid, Jordan.

Qasem Qananwah is currently working as an Associate Professor at the Biomedical Systems and Medical Informatics department, he earned his PhD degree in Biomedical Engineering from Karlsruhe Institute of Technology, Germany. The MSc. degree was obtained in Electronics and communication from Jordan University of Science and Technology, Jordan. He received the BSc. degree in Electronics Engineering from Yarmouk University, Irbid, Jordan.