Turkish Journal of Computer and Mathematics Education *Vol.11 No.3 (2020), 2024-2032*

Research Article

Performance analysis of symptoms classification of disease using machine learning algorithms

Jayesh Mohanrao Sarwade[1], Harsh Mathur[2]

Department of Computer Science & Engineering[12]
Madhyanchal Professional University, Bhopal, India[12]
jayesh.sarwade@gmail.com[1], harshmathur786@gmail.com[2]

Abstract

Classification of acute disease detection symptoms is a big issue in an automatic healthcare system. The automatic healthcare system provides telemedicine for rural and urban citizens worldwide. The classification algorithm estimated the disease behaviors and suggested the medicine for the patients. Machine learning offers various algorithms for mining. The bucket of machine learning techniques is association rule mining, clustering, classification and regression. The classification of medical disease data is very intensive. Most authors applied NB algorithms for better detection of disease. This paper presents the experimental analysis of medical disease data based on various machine learning algorithms. The clustering and classification technique play a significant role in health care databases. The process of clustering defines the process of similar patterns based on iteration. Instead of data, the classification technique guided the grouping technique based on specific guidance. The central complex problem in the classification process is a selection of features of disease data. Most disease data features are similar, and the applied classifier detects the false value of classification. The importance of feature selection in medical disease data classification is paramount. The process of analysis is based on experimental. The experiment uses MATLAB software and reputed disease datasets from the UCI machine learning site.

Keywords: - Disease Detection, Classification, Machine learning. WBC, CVD, MATLAB

I. Introduction

The growth of digital medical data is very vast and the process of data categorization based on symptoms is big challenge for an automatic detection of disease. Machine learning techniques, in particular, traditional data mining techniques are used to raise the accuracy and efficiency of medical data analysis. Due to the large size of the data, these techniques are not suitable for collecting, storing, and analyzing these datasets [2]. In medical data mining, for example, consultants prepare reports regarding their patients in order to give an accurate and efficient decision on their patients' health. This discovered information is available for consultants and patients to access in order to reach an accurate diagnosis [3,4]. The process of combining different clustering output (cluster ensemble or clustering Aggregation) emerged as an alternative approach for improving the quality of the Results of clustering algorithms. It is based on the success of the combination of supervised classifiers. Given a set of objects, a cluster ensemble method consists of two principal steps: Generation, which is about the creation of a set of partitions a of these objects, and Consensus Function, where a new partition, which is the integration of all partitions obtained in the generation step, is computed. Over the past years, many clustering ensemble techniques have been proposed, resulting in new ways to face the problem together with new fields of application for these techniques. Besides the presentation of the main methods, the introduction of taxonomy of the different tendencies and critical comparisons among the methods is really important in order to give a practical application to a survey. Thus, due to the importance that clustering ensembles have gained facing cluster analysis, we have made a critical study of the different approaches and the existing methods.

with the help of Machine learning, data mining is be-coming a must in healthcare industry, it could be used to generate models that describe necessary classes, either using descriptive functions such as clustering, to identify previously unknown facts [9], or using classification and prediction techniques for instance, to predict chronic dis-eases like diabetes [10], cardiovascular [11], or neurological diseases like Parkinson's [12]. Coding a machine-learning project from the ground up is a good way to put the theory into practice, but it's a time consuming and a frustrating task, especially that mathematics represents a large part of every machine learning algorithm. But, with the development of software engineering, companies and researchers are putting in our hand's new tools and plat-forms that facilitate data mining applications to promote the scientific research. Many data mining platforms are available for free, availability means competition, competition means better products, but which one to choose? The thing that motivated us to make this study by comparing different data mining tools using machine learning techniques to classify cardiovascular disease. This paper basically focusses on naïve bias, decision tree and other algorithm for the detection of critical and lifestyle disease such as diabetes, blood pressure and heart disease. The remaining paper discusses, as in section II—the Related work of disease classification. Section III discuss machine learning algorithms. Section IV discusses the experimental analysis of algorithms. Furthermore, finally, conclude in section V.

II. Related Work

Medical big data classification and prediction are an important issue. In recent years, many researchers have made significant efforts to increase the benefits of these data. Utilizing an experimental analysis some contributions of authors describe here.

In this [1] author proposed a selection of features that yields three subsets of features to choose the best features from the three feature subsets, correlation-based ensemble feature selection is used. The classifier's performance was trained and tested using the ten-fold cross-validation technique. The Wisconsin Diagnostic Breast Cancer dataset has a precision of 98.47 percent, whereas the Hepatitis dataset has a precision of 95.51 percent.

In this [2] author proposed a new particle swarm optimization-based method for optimizing the random subspace in order to construct an ensemble classifier They create a random subspace by iteratively clustering input data, then optimize all of the data clusters formed. The outcome is compared to state-of-the-art ensemble classifier techniques. A statistical significance test is also performed, followed by an analysis.

In this [3] author proposed the hybrid optimization algorithm PS-GWO, which combines GWO and PSO, makes a unique contribution. The retrieved features are then exposed to a well-known deep learning technique known as modified DBN, which optimizes the activation function and number of hidden neurons using the same established hybrid approach to increase heart diagnosis accuracy. According to the analysis, the accuracy of the developed PS-GWO-DBN is 60%, 52.5 percent, 35%, and 35% higher than NN, KNN, SVM, and DBN.

In this [4] author presented the decision level, classifiers such as the GRNN and PNN are used. On the GRNN and PNN ensembled models, the mentioned cluster-based one-shot learning was found to be more effective in distinguishing COVID-19 images from the other three classes. Experiments have also shown that the model outperforms current deep learning architecture.

In this [5] author proposed the five alternative meta-heuristic optimization techniques for feature selection and feature weighting in artificial neural networks that were recently introduced the chimp optimization algorithm, tunicate swarm algorithm, bear smell search algorithm, antlion optimization algorithm, and modified antlion optimization algorithm are among the algorithms included. All of these algorithms, on the other hand, have been found to be more effective in feature selection and feature weighing operations.

In this [6] author investigated the HGWO-C4.5 was thoroughly based on real-time patient data, which included elements that influence RA prediction by combining RA and non-RA data. To compare and contrast the HGWO-C4.5 approach with other meta-heuristics-based methods such as GWO-based C4.5, PSO-based C4.5, and individual C4.5 methods. The outcome of cross-validation demonstrates that HGWO-C4.5 attained an overall average accuracy of 86.36 percent from three other techniques, which was 6%–14% greater than the individual predictors' accuracy.

In this [7] author present the framework (C-HMOSHSSA) for MOSHO and SSA gene selection the challenge of maintaining convergence and diversity in real-world optimization situations with many objectives is common. The outcome reveal that the proposed technique outperforms existing state-of-the-art procedures significantly.

In this [8] author Using the attribute-class, attribute-significance, and attribute-attribute relevance measures to select a subset of attributes that are most relevant, significant, and non-redundant from a pool of different attribute subsets in order to predict the presence or absence of different diseases in a medical dataset, the R-Ensemble, a parameter free greedy ensemble attribute selection method.

In this [9] author evaluate the learning model, they included a new operator for features weighting, enhanced the mutation and crossover operators, and introduced nested cross-validation into the GA process. The outcome show that GEFES can generalize the mentioned multipopulational intelligent evolutionary algorithm across a wide range of dataset sizes, including two-class and multi-class datasets. They got 95.83 percent, 97.62 percent, 99.02 percent, 98.51 percent, and 94.28 percent categorization accuracy on average.

In this [10] author identifying the stage of patients' condition, a k-Nearest-Neighbors technique is used, in which a genetic algorithm is used for efficient feature selection to reduce dataset dimensions and improve classifier pace. this means that the algorithm could be used to find a correlation between clinical data and data mining techniques to help with lung cancer staging detection.

In this [11] author diagnosis by comparing data mining algorithms to predict liver disease based on an ELTA strategy. As a outcome, data mining models such as random forest MLP neural network, Bayesian networks, SVM, and PSO-SVM are compared using the ELTA technique. Random Forest, MLP Neural network, Bayesian network, SVM, and PSO-SVM models had an average estimated accuracy of 87.35 percent, 78.91 percent, 66.78 percent, 76.51 percent, and 95.17 percent, respectively.

In this [12] author proposed chaotic Darcy optimization-based feature selection approach is described to pick meaningful features, and this method selects the 31 most discriminative features in the completed HCC dataset. The discussed chaotic Darcy optimization-based HCC survival classification approach yielded a 0.9879 accuracy rate.

In this [13] author aims to categories labels based on internal features' properties, whereas Mutual Congestion is based on frequency. The strategies mentioned here were tested on eight binary medical datasets. The outcome show that X variance works well with conventional datasets, but Mutual Congestion significantly enhances the accuracy of high-dimensional datasets.

In this [14] author propose A unique multi-objective feature selection strategy based on PSO is proposed. There are three primary phases to the process being presented. The original features are displayed as a graph representation model in the first step. the outcome of five medical datasets show that the presented method is more efficient and effective than prior similar methods.

In this [15] author propose a unique fuzzy rule miner (ANT FDCSM) derived from an ant colony meta-heuristic for diabetes patient diagnosis. To increase the performance of the standard ant colony optimization induced decision tree classifier, a few key changes. SW FDCSM is a hybrid of attribute SW and a new fuzzy variation of the well-known distinct class split measure.

In this [16] author proposed A hybrid GA-ABC algorithm for feature selection and classification employing classifier ensemble techniques, based on a genetically based artificial bee colony algorithm. In comparison to other feature selection methods, the discussed model GA-ABC-EL demonstrates an increase in classification accuracy of more than 90% in the reported outcome.

In this [17] author proposed Differential evolution with fuzzy rough set feature selection and ant colony optimization with fuzzy rough set feature selection techniques are two feature selection algorithms that handle multi-objective optimization problems.

In this [18] author proposed a model for identifying neurological disorders such as glaucoma and Parkinson's disease, as well as cancerous diseases such as breast cancer. Using a modified glowworm swarm optimization technique, the presented approach aims to improve the efficiency of the adaptive neuro-fuzzy inference system.

In this [19] author utilized in the GA to identify the most important traits that are associated with heart disease. The GA–experimental SVM's outcome are compared to those of other feature selection methods as Relief, CFS, filtered subset, Info gain, Consistency subset, Chi squared, one attribute based, Filtered attribute, Gain ratio, and GA.

In this [20] author proposed Associative neural networks are used to train multi-layer neural networks. The aberrant jaw fracture characteristic is then successfully classified by the classifier. The system's efficiency is then assessed using the mean square error rate, precision, recall, and accuracy.

In this [21] author used a systematic review methodology based on a three-dimensional search space, which included disease diagnosis, psychiatric disorders, and categorization procedures. The discipline, concepts, and procedures utilized to diagnose various psychiatric diseases are discussed. This study will serve as a road map for those interested in diving deeper into the mining of various psychological diseases.

In this [22] author provide a severe problem when it comes to establishing an efficient cluster. They suggested a remedy to the degradation in clustering during data partitioning in their study. The firefly algorithm is used to create the initial clusters. The outcome suggest that the presented method achieves effective ensemble clustering with higher clustering accuracy than traditional methods.

In this [23] author proposed a two-stage strategy for feature selection was developed. The gene ranks for various filter methods are extremely varied, and the usefulness of rankings is dataset dependent. They begin by evaluating the union and intersection of the top-n properties of Relief, chi-square, and symmetrical uncertainty to create an ensemble of filter algorithms.

In this [24] author classify a technique uses a DBN to categories photos into four stages: normal, early, moderate, and severe. The accuracy of the mentioned MGS-ROA-DBN is 30.1 percent greater than NN, 32.2 percent higher than KNN, and 17.1 percent higher than SVM and DBN, according to the outcome. Similarly, the developed MGS-ROA-DBN is 13.8 percent better than PSO, 5.1 percent better than GWO, 10.8 percent better than WOA, and 2.5 percent better than ROA in terms of accuracy.

In this [25] author proposed artificial neural networks and Naive Bayesian classifiers can be used to conduct classification. However, based on the measurement, they can establish the positivity and negatives of PCOS at an early stage. The mentioned model's accuracy is 98.63 percent, while precision and specificity are both 100 percent, F-measure is 68.76 percent, and recall is 55 percent.

In this [26] author overcome the feature selection challenge; a unique binary whale optimization algorithm (BWOA) was developed. When there is no heuristic information that can lead the search to the optimal minimal subset, BWOA is extremely useful and appealing.

In this [27] author extracted features from the chest CT scans using a Convolutional Neural Network (CNN) architecture in the first module. They used a guided FS methodology in the first stage of FS, employing two filter methods for the initial screening of the features produced from the CNN model MI and Relief-F. The DA was employed in the second step to narrow down the most important features.

In this [28] author present a framework for creating a CDSS that handles the problem of feature selection and class imbalance. The dataset is balanced at the data level in this framework, and feature selection is done using a wrapper technique. the classifier had a 0.65 MCC, 0.84 F1, and 82.46 percent accuracy; for the TSD, it had a 0.74 MCC, 0.87 F1, and 86.88 percent accuracy; and for the PID dataset, it had a 0.78 MCC, 0.89 F1, and 89.04 percent accuracy.

In this [29] author improve the learning methods of fuzzy clustering-based neural networks by combining a probabilistic model with its learning mechanism and developing ensemble neural networks classifier design methodology for generating the network structure.

In this [30] author proposed the CMBA-SVM is a unique intelligence model that combines a chaos-mapped bat algorithm (CMBA) and a support vector machine (SVM). The CMBA method's cooperation was successful in resolving the SVM parameter tuning challenges.

III. Machine Learning Algorithms

The machine learning algorithm is a pillar of medical disease data classification. The multiple features of medical data minimize the classification ratio of the classifier. The sampling of medical data and feature selection methods improves classification accuracy and precision. This section describes three classification algorithms applied to classify medical disease data.

Decision Tree algorithm (DT)

The decision tree is a set of algorithms such as ID3, the successor of ID3(C4.5), classification and regression tree (CART). For example, the process of data classification applied the ID3 algorithm.

The input processing of training phase is data sampling technique for classifier. While fitness function selects the initial input of ID3 algorithm.

1. Sampling of data of sampling technique
2. Split data into two parts training and testing part
3. Apply entropy function for training a sample value
4. Using 2/3 of the sample, fit a tree the split at each node

   For each tree. .

   - Calculate classification of the available 1/3 using the tree, and calculate the misclassification rate = out of Gain.
5. For each variable in the tree
6. Compute Error Rate: Calculate the overall percentage of misclassification
   - Variable selection: Average increase in Gain error over all trees and assuming a normal division of the increase among the trees, decide an associated value of feature.
7. Resulting classifier set is classified
   - Finally to estimate the entire model, misclassification

Conversion of binary attribute in actual value

Naive Bayes (NB) algorithm

Naive Bayes model is compatible for very large datasets to build and for further analysis. This model is very simple and sophisticated classification method, and it performed well even in complicated scenarios. By using Bayes theorem, calculate the posterior probability using the equation below:

$P(a/y) = (P(y/a)P(a))/P(y)$……………..(1)

where P(a/y) indicates the posterior probability of class, P(a) represents the class prior probability, P(y/a) shows the likelihood which is the probability of predictor given class and P(y) indicates the predictor's prior probability.

Random forest(RF) algorithm

The Random Forest (RF) is tree structure collection of base classifiers. The medical data have many numbers of features. The medical dataset contains many unwanted features, only few features of data is predictive. The RF classifier apply probability to select relevant features. The RF algorithm is set of training data D and NF features describe here.

1. begin f1,f2,f3,………..,fn is sample features estimated by probability.
2. Each features-built decision tree. Estimate all possible probability-based N features.
3. Merge all unpair tree H1(X), H2(X) into random forest ensemble

1. Randomly select "n" features from total "N" features.
2. Where n << N
3. Among the "n" features, calculate the node "n" using the best split point.
4. Categorize the node into daughter nodes using the best split.
5. Repeat 1 to 3 steps until "l" number of nodes has been reached.
6. Build forest by repeating steps 1 to 4 for "n" number times to create "n" number of trees.

IV Experimental Analysis

To evaluate the performance of different classification algorithm with MALTAB software. The version of software is R2014a, and the configuration of system is I7 processor, 16GB RAM and windows10 operating system. The MATLAB provides the basic support library file of NB and DT classification algorithm. But the other function of classifier defines and programmed with function file and compile with library file. For the process of detection applied UCI machine learning dataset. The process of sample of data applied 10 cross folds for the processing of prediction and measurement of parameters such as accuracy and precision. The name of datasets mentions in table. Table also mention the feature selection approach for the classification. the standard parameters

$$\text{Accuracy} = \frac{TP}{TP+TN} \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

$$\text{Precision} = \frac{TP}{TP+FP} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Table: 1Average experimental results of the cross validation obtained by different classifiers NB and DT performed on reduced attribute set generated by different attribute selection methods SQR, USQR, DMQR performed on four medical datasets.

Datasets	**Feature selection methods**	**NB Accuracy**	**RF Precision**
WBC	SQR[13]	95.70	94.16
	USQR[14]	95.99	93.99
	DMQR[15]	95.84	94.56
Lung Cancer	SQR[13]	71.87	69.86
	USQR[14]	75.02	73.03
	DMQR[15]	81.25	79.94
Diabetes	SQR[13]	75.91	72.13
	USQR[14]	77.21	74.59
	DMQR[15]	79.54	77.24
Chronic kidney	SQR[8]	95.75	91.46
	USQR	87.75	85.69
	DMQR	39.25	37.27

Table: Average experimental results of the cross validation obtained by different classifiers NB and RF performed on reduced attribute set generated by different attribute selection methods EMFS, HGWO, PSO performed on four medical datasets.

Datasets	**Feature selection methods**	**NB Accuracy**	**RF Precision**
ILP	EMFS[10]	50.08	69.63
	HGWO [6]	54.42	71.79
	PSO[3]	69.56	78.65
Breast cancer	EMFS[10]	72.65	75.64
	HGWO [6]	76.49	79.65
	PSO[3]	78.65	81.21
Dermatology	EMFS[10]	66.99	68.12
	HGWO [6]	68.24	72.15
	PSO[3]	69.45	73.69
Hepatitis	EMFS[10]	80.64	83.21
	HGWO [6]	82.94	85.33
	PSO[3]	85.73	87.01

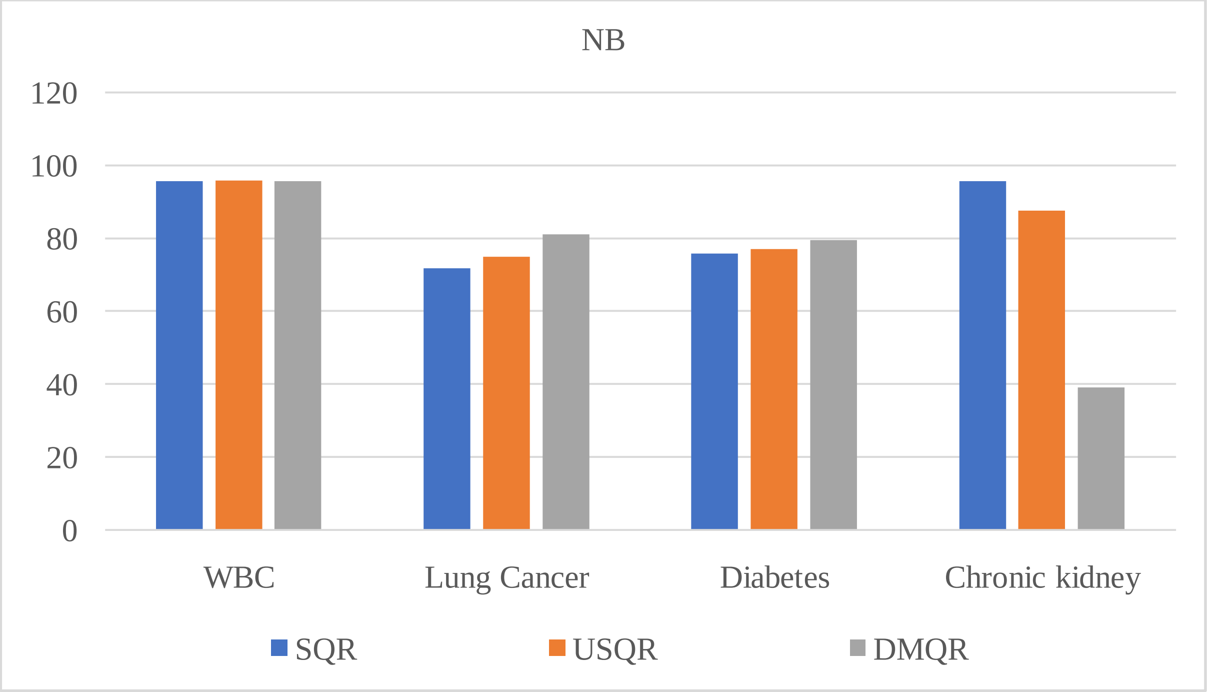

Fig: Shows that comparative result of different dataset WBC, Lung cancer, Diabetes and Chronic kidney achieved by NB classifier using SQR, USQR and DMQR.

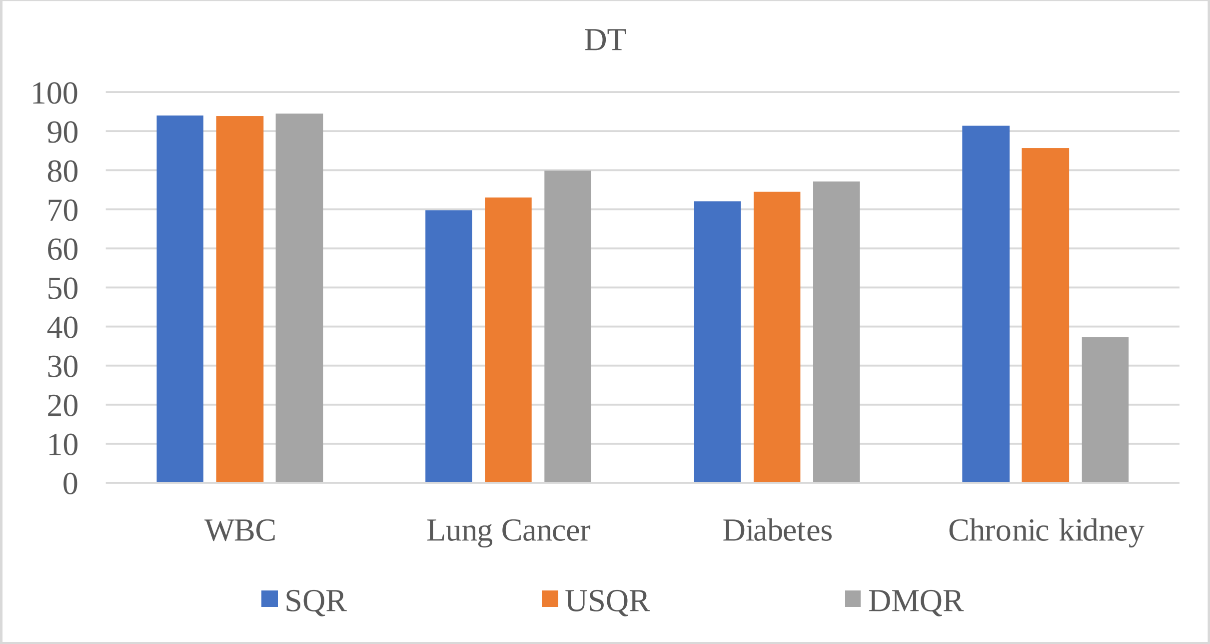

Fig: Shows that comparative result of different dataset WBC, Lung cancer, Diabetes and Chronic kidney achieved by NB classifier using SQR, USQR and DMQR.

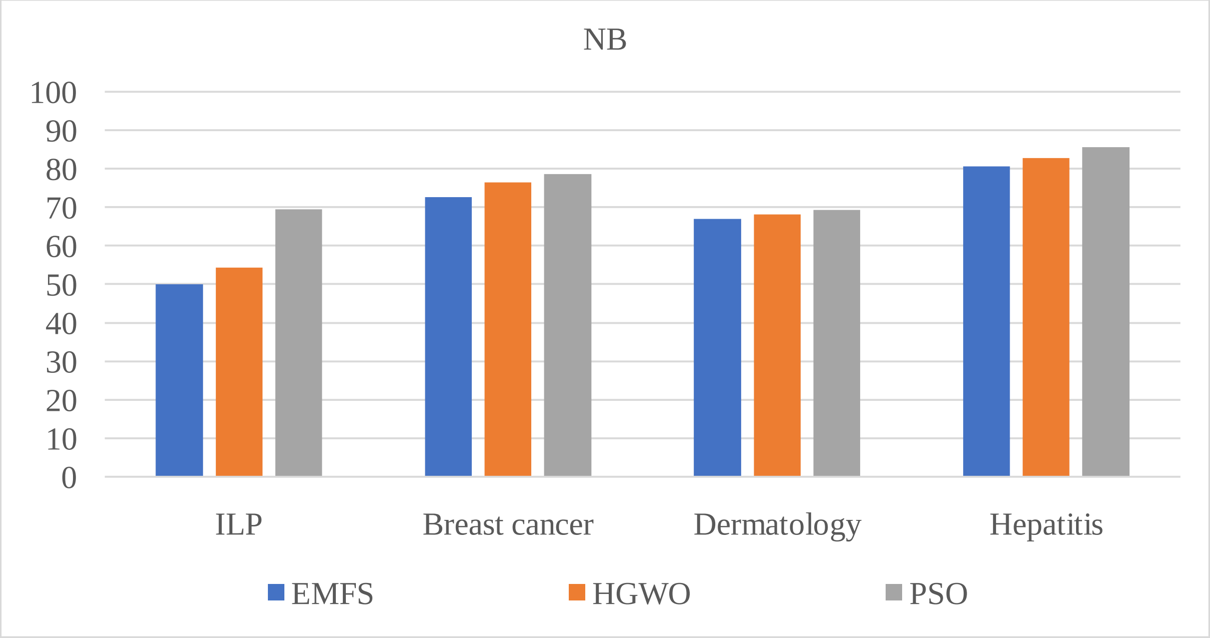

Fig: Shows that comparative result of different dataset ILP, Breast cancer, Dermatology and Hepatitis achieved by NB classifier using EMFS, HGWO and PSO.

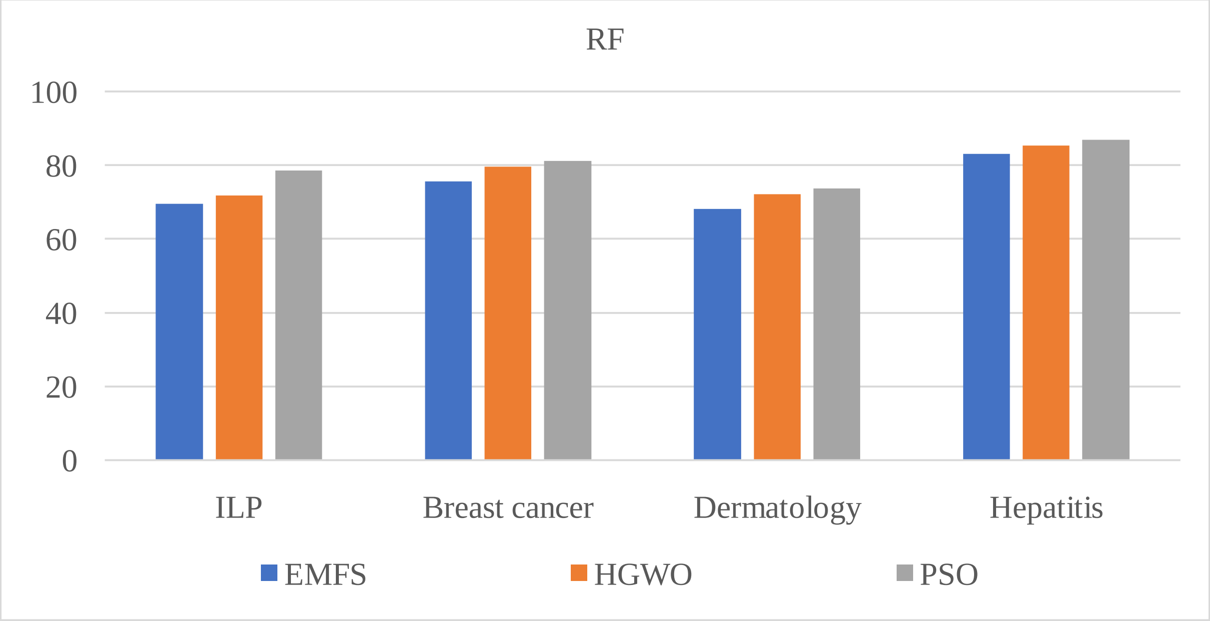

Fig: Shows that comparative result of different dataset ILP, Breast cancer, Dermatology and Hepatitis achieved by RF classifier using EMFS, HGWO and PSO.

V. Conclusion & Future Work

Machine learning algorithms effectively classify the medical disease data. The objective of this study is to finds a model for the diagnosis of diabetes, coronary heart disease and cancer among the available dataset. The dataset is chosen from online repositories UCI. The techniques of pre-processing applied are filled in missing values and removing correlated columns. Next, the classifier is applied to the pre-processed dataset, and then random forest models are constructed. Finally, the accuracy of the models is calculated and analyses are based on the efficiency calculations. NB Classification network shows the accuracy of 74.46, 82.35 and 63.74% for diabetes, coronary heart disease and cancer data. Similarly, classification with Random Forest model shows the accuracy of 74.03, 83.85 and 92.40. The accuracy outcome of Random Forest model for the three diseases is greater than the accuracy values of Naïve Bayes classifier. The multiple feature selection approach the efficiency of classification algorithms. The particle swarm optimization feature selector achieve more sample data for the classification. in future design hybrid classifier for the detection and classification of medical disease data.

References

1. Elgin Christo, V. R., H. Khanna Nehemiah, B. Minu, and Arputharaj Kannan. "Correlation-based ensemble feature selection using bioinspired algorithms and classification using backpropagation neural network." *Computational and mathematical methods in medicine* 2019 (2019).
2. Md. Jan, Zohaib, and Brijesh Verma. "Evolutionary classifier and cluster selection approach for ensemble classification." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 14, no. 1 (2019): 1-18.
3. AliyarVellameeran, Fathima, and Thomas Brindha. "A new variant of deep belief network assisted with optimal feature selection for heart disease diagnosis using IoT wearable medical devices." *Computer Methods in Biomechanics and Biomedical Engineering* (2021): 1-25.
4. Aradhya, VN Manjunath, Mufti Mahmud, D. S. Guru, Basant Agarwal, and M. Shamim Kaiser. "One-shot cluster-based approach for the detection of COVID–19 from chest X–ray images." *Cognitive Computation* (2021): 1-9.
5. Diaz, P. M., and M. Julie Emerald Jiju. "A comparative analysis of meta-heuristic optimization algorithms for feature selection and feature weighting in neural networks." *Evolutionary Intelligence* (2021): 1-20.
6. Sundaramurthy, Shanmugam, and Preethi Jayavel. "A hybrid grey wolf optimization and particle swarm optimization with C4. 5 approach for prediction of rheumatoid arthritis." *Applied Soft Computing* 94 (2020): 106500.
7. Sharma, Aman, and Rinkle Rani. "C-HMOSHSSA: Gene selection for cancer classification using multi-objective meta-heuristic and machine learning methods." *Computer methods and programs in biomedicine* 178 (2019): 219-235.
8. Bania, Rubul Kumar, and AnindyaHalder. "R-Ensembler: A greedy rough set based ensemble attribute selection algorithm with kNN imputation for classification of medical data." *Computer methods and programs in biomedicine* 184 (2020): 105122.
9. Sahebi, Golnaz, ParisaMovahedi, MasoumehEbrahimi, TapioPahikkala, JuhaPlosila, and Hannu Tenhunen. "GeFeS: A generalized wrapper feature selection approach for optimizing classification performance." *Computers in biology and medicine* 125 (2020): 103974.
10. Maleki, Negar, Yasser Zeinali, and SeyedTaghiAkhavanNiaki. "A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection." *Expert Systems with Applications* 164 (2021): 113981.
11. Joloudari, JavadHassannataj, Hamid Saadatfar, AbdollahDehzangi, and ShahaboddinShamshirband. "Computer-aided decision-making for predicting liver disease using PSO-based optimized SVM with feature selection." *Informatics in medicine unlocked* 17 (2019): 100255.
12. Demir, FahrettinBurak, TurkerTuncer, Adnan FatihKocamaz, and FatihErtam. "A survival classification method for hepatocellular carcinoma patients with chaotic Darcy optimization method-based feature selection." *Medical hypotheses* 139 (2020): 109626.
13. Alirezanejad, Mehdi, RasulEnayatifar, HomayunMotameni, and HosseinNematzadeh. "Heuristic filter feature selection methods for medical datasets." *Genomics* 112, no. 2 (2020): 1173-1181.
14. Rostami, Mehrdad, SamanForouzandeh, Kamal Berahmand, and Mina Soltani. "Integration of multi-objective PSO based feature selection and node centrality for medical datasets." *Genomics* 112, no. 6 (2020): 4370-4384.
15. Singh, Akansha, and Gaurav Gupta. "ANT_FDCSM: A novel fuzzy rule miner derived from ant colony meta-heuristic for diagnosis of diabetic patients." *Journal of Intelligent & Fuzzy Systems* 36, no. 1 (2019): 747-760.
16. Nagarajan, Senthil Murugan, V. Muthukumaran, R. Murugesan, Rose Bindu Joseph, and MeramMunirathanam. "Feature selection model for healthcare analysis and classification using classifier ensemble technique." *International Journal of System Assurance Engineering and Management* (2021): 1-12.
17. Meenachi, L., and S. Ramakrishnan. "Differential evolution and ACO based global optimal feature selection with fuzzy rough set for cancer data classification." *Soft Computing* 24 (2020): 18463-18475.
18. Balasubramanian, Kishore, and N. P. Ananthamoorthy. "Improved adaptive neuro-fuzzy inference system based on modified glowworm swarm and differential evolution optimization algorithm for medical diagnosis." *Neural Computing and Applications* 33, no. 13 (2021): 7649-7660.
19. 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, no. 6 (2019): 14777-14787.
20. Hashem, Mohamed, and Azza S. Hassanein. "Jaw fracture classification using meta heuristic firefly algorithm with multi-layered associative neural networks." *Cluster Computing* 22, no. 3 (2019): 7079-7086.
21. Kaur, Prableen, and Manik Sharma. "Diagnosis of human psychological disorders using supervised learning and nature-inspired computing techniques: a meta-analysis." *Journal of medical systems* 43, no. 7 (2019): 1-30.
22. Yuvaraj, N., and C. Suresh Ghana Dhas. "High-performance link-based cluster ensemble approach for categorical data clustering." *The Journal of Supercomputing* 76, no. 6 (2020): 4556-4579.

23. Ghosh, Manosij, SukdevAdhikary, KushalKanti Ghosh, AritraSardar, Shemim Begum, and Ram Sarkar. "Genetic algorithm based cancerous gene identification from microarray data using ensemble of filter methods." *Medical & biological engineering & computing* 57, no. 1 (2019): 159-176.
24. Jadhav, Ambaji S., Pushpa B. Patil, and Sunil Biradar. "Optimal feature selection-based diabetic retinopathy detection using improved rider optimization algorithm enabled with deep learning." *Evolutionary Intelligence* (2020): 1-18.
25. Maheswari, K., T. Baranidharan, S. Karthik, and T. Sumathi. "Modelling of F3I based feature selection approach for PCOS classification and prediction." *Journal of Ambient Intelligence and Humanized Computing* 12, no. 1 (2021): 1349-1362.
26. Tawhid, Mohamed A., and Abdelmonem M. Ibrahim. "Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm." *International Journal of Machine Learning and Cybernetics* 11, no. 3 (2020): 573-602.
27. Sen, Shibaprasad, SoumyajitSaha, Somnath Chatterjee, SeyedaliMirjalili, and Ram Sarkar. "A bi-stage feature selection approach for COVID-19 prediction using chest CT images." *Applied Intelligence* (2021): 1-16.
28. Sreejith, S., H. Khanna Nehemiah, and Arputharaj Kannan. "Clinical data classification using an enhanced SMOTE and chaotic evolutionary feature selection." *Computers in Biology and Medicine* 126 (2020): 103991.
29. Kim, Eun-Hu, Sung-Kwun Oh, WitoldPedrycz, and Zunwei Fu. "Reinforced fuzzy clustering-based ensemble neural networks." *IEEE Transactions on Fuzzy Systems* 28, no. 3 (2019): 569-582.
30. Sahu, Bibhuprasad, and SachiNandanMohanty. "CMBA-SVM: a clinical approach for Parkinson disease diagnosis." *International Journal of Information Technology* 13, no. 2 (2021): 647-655.