IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Issue-8, Aug 2019 E-ISSN: 2347-2693

A Survey of Classification Methods and Techniques for Improving Classification Performance

M. Balasaraswathi[1*], A. Uthiramoorthy[2]

[1,2]Department of Computer Science, Rathinam College of Arts and Science Coimbatore

Corresponding Author: baladars@yahoo.co.in

DOI: https://doi.org/10.26438/ijcse/v7i8.233240 | Available online at: www.ijcseonline.org

Accepted: 13/Aug/2019, Published: 31/Aug/2019

Abstract - This paper surveys the state of the art techniques which have been reviewed to develop the overall classification methodology of this research work. The feature selection methods, traditional classification algorithms followed by a brief description of theoretical works on data mining are summarized. The major classification approaches and the techniques which is used for improving classification performance are analyzed. In addition, some important issues affecting classification performance are discussed. In this paper we have gone through the existing work in the area of classification which will allow us to have a fair evaluation of the progress made in the field of Classification.

Keywords: Machine Learning, Feature Selection methods, Classification.

I. INTRODUCTION

Classification is a process of mapping the given data item into one of predefined classes using the learning function. Classification, being a data analysis technique, extracts models unfolding important data classes and predicts future values. To discover and present knowledge in an understandable format, data mining uses classification techniques with image processing, machine learning, statistical and visualization techniques, and natural language processing. In literature, the classification algorithms mostly suffer from the memory resident and scalability problems. To overcome this, recent data mining research has concentrated in developing scalable and efficient classification techniques capable of handling large disk resident data. The performance metrics like accuracy, scalability, speed, comprehensibility, robustness, time and interpretability are used to evaluate the classification techniques.

The Data classification process includes two steps namely (Han and Kamber, 2006), Building the Classifier or Model: This step is the learning step or the learning phase. In this step the classification algorithms build the classifier. The classifier is built from the training set made up of database tuples and their associated class labels.

Using Classifier for Classification: In this step, the classifier is used for classification. To estimate the accuracy of classification rules, the test data is used. The classification rules can be applied to the new data tuples if the accuracy is considered acceptable.

The data mining tasks and the dimensionality of the data involved in the learning model has increased explosively in the past three decades. The curse of dimensionality remains a challenge to the existing learning models [2] [3]. The learning model tends to overfit in the presence of a large number of features, thereby reducing the performance. To address the problem of the curse of dimensionality, the feature selection methods which are an important branch in the machine learning and data mining research area have been studied.

II. FEATURE SELECTION METHODS

Feature selection [4] is the process of identifying and removing the irrelevant and redundant information present in the dataset. The objective is to select a small subset of important features from the original one by using the evaluation criteria, and to attain a good learning performance (e.g., higher accuracy in learning, lower computational cost, and better model interpretability).

Feature selection algorithms can be grouped into supervised [5] [6], unsupervised [7] and semi-supervised feature selection [8] [9] based on the applications. Supervised feature selection methods could be again classified into filter models, wrapper models and embedded models. The filter model distinguishes feature selection from classifier learning as the bias of the learning algorithm and does not link with the bias of feature selection algorithm. This depends on the general qualities of the training data including distance, consistency, dependency, information, and correlation.

John et al. [10] proposed the filtering models that are used to extract features from the data without any learning model. The wrappers make use of the learning techniques which is used to evaluate the features that are most useful.
Lal et al. [11] introduced an embedded technique that merges the feature selection process and the classifier construction. Among these supervised feature selection methods, applying a filter algorithm increases the performance of classification algorithms and also reduces the computer processing time. In the recent years, a number of performance criteria have been developed for filter based feature selection such as Fisher score, Trace Ratio Criterion, Relief, Relief-F and Correlation based Feature Selection (CFS).

Proposed a Fisher score which is a popularly used supervised feature selection method. It chooses each feature separately related to their scores under the Fisher criterion, which leads to a suboptimal subset of features.
Proposed a generalized Fisher score to select optimal set of features for multiple kernel learning problem. It aims to discover a subset of features, which increases the lower bound of the general Fisher score. The result of feature selection problem is formulated as Quadratically Constrained Linear Programming (QCLP). It is determined by using cutting plane algorithm, in which iteration of a multiple kernel learning problem is solved by multivariate ridge regression and projected gradient descent algorithm. The experimentation result on benchmark data sets concludes that the Fisher score method performs better when compared to state-of- the-art feature selection techniques.

Designed a hybrid feature selection method, which combines the procedure of Fisher score and Genetic Algorithm (GA). It uses the features' Fisher score to create the initial population of GA. The GA uses the initial population to present feature selection with exclusive approach for reference. Four different data sets of Sonar, WDBC, Arrhythmia, and Hepatitis are chosen to measure the performance of feature selection algorithm. 1-Nearest Neighbor (1-NN) classifier is developed to classify the selected features data sets, and correspondingly, obtain the classification accuracy results of 72.36%, 95.64%, 72.04%, and 87.83% with ten-fold cross validation. The experimentation results demonstrated that the hybrid algorithm is fit for removing redundant features, and irrelevant features when compared with the performance of Fisher score, GA, and Fisher score with GA.

Proposed a new Hidden Markov Model (HMM) based feature selection scheme for Face Recognition (FR). In this theory, HMM technique is used to model the classes of face images. A set of Fisher scores is computed by using partial derivative analysis of the parameters which is evaluated in each HMM. These Fisher scores are again merged with some conventional features like log-likelihood and appearance based features to form new feature vectors which collectively uses the strengths of both local and holistic features of human face. Linear Discriminant Analysis (LDA) is then used to evaluate the feature vectors for FR. Improvements in performance measures are analyzed between traditional HMM model and Fisher score based HMM method which uses appearance based feature vectors. The experimentation work shows that, by minimizing the number of models involved in the training and testing stages of LDA, the proposed feature selection scheme could obtain a high discriminative power with a lesser computational time when compared to traditional HMM based FR system.

Proposed a new Feature Selection (FS) method based on fisher criterion and genetic optimization, named as FIsher criterion and Genetic (FIG), so as to handle the Computed Tomography (CT) Imaging Signs of Lung diseases (CISL) recognition problem. In the FIG feature selection method, the Fisher criterion is practiced to calculate feature subsets, based on this genetic optimization algorithm which is proposed to identify an optimal feature subset from the candidate features. The FIG method is proposed to choose the features for the CISL recognition from the variety of features, including the bag of visual words based on the histogram of oriented gradients, the wavelet transform based features, the local binary pattern, and the CT value histogram. Then, the selected features works with every five generally used classifiers such as Support Vector Machine (SVM), Bagging (Bag), Naïve Bayes (NB), K Nearest Neighbor (KNN), and AdaBoost (Ada) to categorize the Regions Of Interests (ROIs) in lung CT images into the CISL categories. To measure the results of FIG feature selection method and CISL recognition approach, the fivefold cross validation experiments are performed on a set of 511 ROIs collected from real lung CT images. Among the particular category of classifiers, the proposed FIG based feature selection method attained higher recognition performance for selected features as well as the complete set of features. Moreover, the results of FIG method is compared with the FS method based on classification Accuracy Rate Genetic optimization (ARG). The results reveal the computational efficiency and effectiveness of FIG over ARG.

The major limitation of this fisher criterion is that it cannot deal with redundant features as it validates the features independently. The research work [17] proposed the Fisher score and Laplacian score which are the mostly used filter methods for feature selection and they belong to the common graph based FS framework. In this graph based FS framework, a feature subset is chosen that depends on the subset level score which is computed in a trace ratio form. As the number of feature subsets available is very large, and it have higher computational cost to discover optimal subset of feature subset with the higher subset level score. Though, the feature subset is chosen based on the feature score it does not guarantee the best of the subset score. suggested that the optimization of the subset level score, and designed a novel algorithm to successfully discover the global optimal feature

subset so as to increase the subset level score. The experimental results demonstrate that the accuracy of the proposed algorithm is high when compared with conventional methods for feature selection.

In the case of feature selection problem, trace ratio optimization is extensively used in the recent works [19] [20]. Since it directly reflects the similarity of data points. Usually, there is no direct solution to the original trace ratio problem. Previous works have specified that the trace ratio problem is solved in an iterative manner.

Developed an efficient Iterative Trace Ratio (ITR) Score algorithm to discover the optimal solutions. This algorithm can be suitably extended to its related kernel version for dealing with the nonlinear problems. Finally, the results of proposed ITR Score algorithm are measured on the extensive simulations of real world datasets. The results demonstrate that the proposed ITR Score algorithm is able to provide higher improvements when compared to other supervised and unsupervised algorithms.

Introduce a new Trace Ratio-Linear Discriminant Analysis (TR-LDA) for dementia diagnosis. An improved ITR algorithm (iITR) is introduced to solve the TR-LDA problem. This novel method can be combined with enhanced missing value imputation algorithm and used for the diagnosis of the nonlinear datasets in various real world medical diagnosis problems. Wide ranging simulations are conducted to measure the efficiency of the proposed work. The results conclude that the proposed TR-LDA method obtains higher accuracy for recognizing the demented patients when compared with the other existing algorithms.

Propose a new effective algorithm to discover the feasible optimal solution for trace ratio problem. From this algorithm, an orthogonal constrained semi-supervised learning framework is developed. This algorithm incorporates unlabeled data into training stage in order to preserve the discriminative structure as well as geometrical structure fixed in the original dataset. Theoretical analysis demonstrates that there exists a particular relationship among linear and nonlinear methods. An extensive simulation on the synthetic dataset and the real world dataset are presented to show the accuracy of the proposed algorithm. The result reveals a higher accuracy when compared to the existing algorithms.

For some applications the optimal solution is not obtained clearly by using trace ratio and fisher score function. A filter method is proposed and that works by sampling an instance randomly on the original dataset, and then discovers its nearest neighbor from the same and different class. The values of the features of the nearest neighbors are compared with the dataset samples and it is used to revise the relevance scores for each feature. This process is repeated until it reaches a user specified number of instances m. Features with weight, which is greater than the threshold τ, are considered as the suitable feature to the target variable.
Introduced a Relief algorithm for random selection of data instances for feature weight calculation. Monte Carlo Approach is used for random selection of instances in the Relief function. The accuracy of proposed algorithm is experimented with Cotton Disease datasets and implemented in Waikato Environment for Knowledge Analysis (WEKA) tool. Naïve Bayes (NB) and J48 are used as the classifiers and classification results are measured in terms of classification accuracy and size of feature subspace to show the efficiency of the Relief algorithm.

Found the new approach on feature selection on Relief, based on Median Variance model and is named as LAS-Relief algorithm. This algorithm alleviates the feature weights assessment when compared to mean variance based Relief algorithm and is considered to be an enhanced booming algorithm for feature selection. The random selection of instances in the data sets will direct to the variation of weight evaluation.

This in turn leads to poor evaluation accuracy. The new LAS-Relief algorithm incorporates the median variance in the feature weight assessment by removal of the inappropriate features in the feature space. The feature weight is intended by selection of the instances at random. This algorithm makes the result more stable and more precise on classification. They also introduced the new distance function for calculating the distance between the two instances is a squared Euclidean distance.

Developed an innovative feature selection algorithm by introducing the spatially weighted variation of Relief and is called as Sigmoid Weighted ReliefF Star (SWRF*). This algorithm is applied on synthetic Single Nucleotide Polymorphism (SNP) data sets. They reported that SWRF achieve well on SNP data than Relief-F.

In the LAS-Relief algorithm, noisy values of attributes or features are not taken into account in the feature assessment. The noisy features in the feature set powerfully affect the selection of nearest neighbors. This in turn affects the feature weight evaluation as well as the accuracy of classification. This issue is solved by introducing the new Relief algorithm.

Although, the Relief algorithm has comparatively unique drawback that the feature weight may vary with the instances, in most of the occasion, the instances possessed are indiscriminate. Furthermore, with regard to the Relief algorithm, the frequency in sampling could not be predicted. Hence, Relief algorithm is inconsistent and minimizes the accuracy of anticipated results found a quantum bio inspired Estimation of Distribution Algorithm (EDA) for CFS. The presented algorithm mincorporates the Quantum computing concepts, vaccination process with the Immune Clonal

Selection (QVICA) and EDA. It is engaged as a search technique for CFS to discover the feasible feature subset from the features space. It is enacted and calculated on benchmark dataset Knowledge discovery in Databases (KDD)- cup99 and correlated with the GA algorithm. The attained outcome proved the ability of QVICA- with EDA to acquire better feature subsets with a minimum length, greater fitness values and in a minimum computation time. Relative studies are carried out on CFS and different methods in the past on different data domains.

Table 1. Feature selection Methods - Findings From Literature Review

Year	Author	Technique	Observations
2012	Gu et al	Fisher score	To select optimal set of features for multiple kernel learning problem. The result is formulated as QCLP and performs better when compared to up to date feature selection techniques.
2016	Zhou	Hybrid feature selection	Hybrid algorithm is fit for removing redundant and irrelevant features. It combines fisher score and GA.
2015	Liu et al	*FIsher criterion* and *Genetic* (FIG)	Identifies an optimal feature subset from the candidate features. It attains higher recognition performance for selected features.
2011	Zhao et al	Iterative Trace Ratio (ITR) Score Algorithm	Suitably extended to its related kernel version for dealing with the nonlinear problems. ITR Score algorithm provides higher improvement on extensive simulations of real world datasets.
2013	Zhao et al	Trace Ratio-Linear Discriminant Analysis (TR- LDA)	It is used for the diagnosis of the nonlinear datasets. Exhibiting higher accuracy in recognizing the dementia patients. Still higher computational efficiency is needed.
2015	Rosario and Thangadurai	Relief algorithm	Handles random selection of data instances for feature weight calculation. Accuracy of proposed algorithm is experimented with Cotton Disease datasets. Performance is measured in terms of size of feature subspace and classification accuracy.
2012	Stokes and Visweswaran	Sigmoid Weighted ReliefF Star (SWRF*)	SWRF performs well on Single Nucleotide Polymorphism (SNP) data sets ReliefF algorithm is inconsistent and minimizes the accuracy of anticipated results.
2003	Yu and Liu	Correlation based Feature Selection (CFS)	Efficiency and effectiveness of method is demonstrated using real world data of high dimensionality. Efficiently identify both irrelevant and redundant features with less time complexity than subset search algorithms. Drawback is that the most of the filter methods are invariant in nature.

III. CONVENTIONAL CLASSIFICATION METHODS

Classification is a machine learning algorithm and it is used to classify the data instances into a set of predetermined classes (Balasaraswathi, M, 2017). The classification algorithms are categorized as mathematical and statistical techniques. Some methods are Decision Trees (DT), Naïve Bayes (NB), K Nearest Neighbour (KNN) classifier, Neural Network (NN) classifier, and Support Vector Machines (SVM) classifier which are discussed in the coming section.

Proposed few feature selection algorithms such as Stepwise Feature Selection (SFS), Sequential Floating Forward Search (SFFS), and Principal Component Analysis (PCA), with two major classifiers such as Fisher's Linear Discriminant Analysis (FLDA) and Support Vector Machine (SVM). The dataset samples were collected from multidimensional feature spaces from Multivariate Gaussian Distribution (MVGD) function with equivalent or uneven covariance matrices. The results conclude that the PCA based feature selection algorithm does not perform better when compared to other SFS and SFFS algorithms. All of these FS algorithms produce higher accuracy results for SVM with Radial Kernel when compared to FLDA with less number of training samples, whereas FLDA performs better with huge number of training samples.

Evaluated the performance of the various classifiers such as Random Tree (RT), Quinlan Decision Tree (QDT) algorithm (C4.5), and KNN algorithm on a Wisconsin Breast Cancer (WBC) dataset samples. This dataset sample is collected from the UCI machine learning repository which consists of eleven features and 106 data instances. The results of these classifiers are measured in terms of accuracy and other performance measures. These algorithms predict the presence of WBC and the related breast tissue situations with the purpose of reducing the risk of occurrence of cancer in future. Furthermore the significance of feature selection is to enhance the accuracy of classification algorithms. RT produces higher classification accuracy than the other two classification algorithms QDT and KNN for all the training data under multiple classes.

Decision tree is a popular technique for both induction research and data mining, which is predominantly utilized for model classification and prediction. ID3 algorithm is one of the commonly used algorithms in the decision tree.

Proposed an efficient classifier which was perfectly trained in order to classify oncogenic data. The Lymphographic dataset is completely used in machine learning techniques to train the classifier with the help of feature selection and classification algorithms. Feature Selection is a supervised method which tries to choose a subset of the predictor features that is focused on the information gain. The Lymphography dataset contains 18 predictor attributes and 148 instances with the class label which has four unique values. This analyses the potential of sixteen classification algorithms on the Lymphographic dataset which validates the classifier to correctly perform multiclass categorization of medical data. The fact determines that the Random Tree algorithm and the Quinlan's C4.5 algorithm give 100 percent classification accuracy with all predictor features and also with the feature subset chosen by the Fisher Filtering feature selection algorithm. Furthermore, Relief based FS algorithm provides better results for Radial Basis Function(RBF) algorithm which enhances the classifier accuracy by 1.35%. It can also be mentioned that the C4.5 algorithm provides more efficient classification since the decision tree size generated is smaller than the Random Tree.

Illustrated the general ideas of decision tree in data mining and proposed a study about ID3's inclining to select the attributes with multiple values, and then a new decision tree algorithm that merges ID3 and Association Function(AF) is presented. The results prove that the proposed algorithm can overwhelm ID3's disadvantages effectively and acquire reasonable and effective rules.

Developed an algorithm that is focused on the expectation information entropy and association function rather than the traditional information gain. In the proposed algorithm, the expectation information entropy is modified with the improved association function and the count of the attributes values. The result reveals that the improved algorithm acquired reasonable and more effective rules.

Designed an improved algorithm based on the information entropy and attribute weights. It generally combines the Taylor's theorem and Attribute Similarity theorem to reduce the evaluation of Entropy. To discover the importance of attribute weights and to attain amended information gain, so as to the attribute selection criteria. The comparison results proved that the algorithm will be able to develop in terms of speed of classification, specifically it enhances the accuracy of rules, and generates more extended practical rules for applications.
Proposed a new decision based ID3 classification algorithm, which chooses attribute value with the maximum gains as the experiment attribute of its datasets, creates decision making node, and splits them. It consists of frequent logarithmic process that concerns the effectiveness of creating Decision Tree (DT), when there are a huge number of dataset samples. Therefore one should modify the selection criterion of dataset attributes, with the Taylor principle to transform the algorithm to decrease the quantity of data computation and the creation time of DTs and consequently increase the accuracy of the ID3 classifier.

Proposed a new improved ID3 algorithm based on attribute reduction. During development of the features importance is given to highlight the features with fewer values and higher significance, reduce the attributes with higher values and lower significance. This improved ID3 algorithm makes use of feature importance toward increase Information Gain (IG) of features which has lesser attributes and results are compared with ID3 algorithm. The experimentation results demonstrate that the proposed improved ID3 algorithm performs better when compared to ID3 algorithm.

Naïve Bayes (NB) classifier is a straightforward Bayesian network classifier which has been applied effectively in various domains. In spite of the straightforwardness of the NB classifier and the restrictiveness of the independent assumptions between features, it is found to be more successful when compared to traditional classifiers.

Proposed a new NB based classifier and is used to increase the performance of the classifier, when compared to other caching methods such as Least-Recently-Used (LRU) and Greedy Dual-Size (GDS). NB is combined with these caching methods known as NB-GDS, NB-LRU and NB-DA. The results show that the proposed methods outperform the traditional caching methods on several web proxy datasets.

Introduced a new Class Dependent Feature Weighting (CDFW) using Naïve Bayes (NB) classifier based feature scaling algorithm. This CDFW-NB-RFE algorithm combines the procedure of CDFW and Recursive Feature Elimination (RFE). The experimentation results illustrate that the proposed CDFWNB-RFE algorithm performs better when compared to other feature ranking algorithms on text datasets.

Classified the imbalance between classes as the major problem, and developed a new method based on learning and sampling probability distributions. The experimental results show that the proposed Multinomial Naïve Bayes(MNB) performs better on Experiments Over A Standard Corpus (ENRON) on seven datasets and results are compared with Synthetic Minority Over- sampling TEchnique (SMOTE) and other classifiers.

Proposed a new Multinomial Naïve Bayes (MNB) based classifier on four text datasets, which is able to increase the performance using locally weighted learning. Tang et al. (2016) proposed a new information theory based FS algorithm which aims to select the features with their discriminative capability for classification. At initial stage of the work two information measures such as Kullback Leibler

Divergence (KLD) and Jeffreys divergence are computed for binary classification, and their asymptotic property is evaluated based on the type I and type II errors of a MNB classifier. The proposed FS algorithm ranks the unique features that aim to increase the classifier performance for text categorization.

Nearest Neighbor (NN) Classification is relatively straightforward, instances are categorized based on the class of their nearest neighbors. The K-Nearest Neighbor (KNN) classifier is a nonparametric classifier with the purpose of attaining higher accuracy for optimal value of k. In the KNN rule, a test example is allocated to the class that commonly represents the K nearest training datasets. KNN classification is categorized into two major steps such as structure less NN techniques and structure based NN techniques. In the former, the complete dataset samples are classified into training and testing data, and from training dataset the distance function is computed, the dataset point with lowest distance is named as NN. In the latter, the classification is performed based on structures of training samples such as like Orthogonal Structure Tree (OST), ball tree, K-D tree, and central line.

Proposed a new hybrid classification algorithm which combine the procedure of KNN and GA. GAs perform global exploration in huge and multimodal landscapes, and yields optimal results. Experimental results demonstrate that the proposed KNN-GA algorithm increases the accuracy of the classifier in diagnosis of heart disease.
Proposed a new KNN based classifier that facilitates healthcare professionals in the analysis of heart disease prediction. The results attempts to investigate whether integration of voting with KNN is able to increase its performance in the prediction of heart disease. The investigated results show that the proposed KNN classifier produces higher classification accuracy when compared to NN. Introduced a new KNN classification algorithm for economic forecasting datasets and the results are compared with conventional statistical methods and nonparametric methods.

Presented an evidence theory based KNN classification algorithm which is different the existing methods. This evidence theory introduces a two frequency estimations of prior probability such as Global and Local known as Global Estimations (GE) and, Local Estimations (LE) respectively. The GE is designed for a class is the prior probability of the class across the complete training samples depending on the frequency estimation, and on the other hand, the LE is designed for the prior probability of class in a specific neighborhood. By measuring the variation between the GE and the LE of each class, an imbalanced dataset problem is solved in this work. The experimental results of the proposed KNN algorithm were measured on two benchmark datasets, which concludes that the frequency estimations based KNN performs better than normal KNN algorithms.

Introduced a new Euclidean distance function to KNN classifier and the performance of the KNN classifiers with different distance functions are evaluated on three different medical datasets. These datasets includes categorical, numerical, and mixed types of data, and the results were compared with some standard distance functions such as cosine, Chi square, and minkowski with KNN classifier separately. The experimentation results demonstrated that the Chi square based distance function with KNN classifier performs better for all types of datasets. On the other hand, usage of cosine and Minkowsky distance function with KNN classifier produces worst results for mixed type of datasets.

Table 2.2 Conventional Classification Methods - Findings

From Literature Review

Year	Author	Technique	Observations
2010	Way et al	Feature Selection Algorithm-SFS,SFSS,PCA	Feature Selection produces higher accuracy results for SVM with radial kernel when compared to FLDA. FLDA performed better with huge number of training samples.
2011	Jacob and Ramani	Random Tree (RT), Quinlan Decision Tree (QDT), (C4.5), and KNN	RT produces higher classification accuracy than C4.5 and KNN for all the training data under multiple classes. They are not easily applicable to huge number of training samples.
2013	Jacob et al	RT and Quinlan's C4.5 algorithm	RT and Quinlan's C4.5 algorithm C4.5 algorithm provides high classification accuracy with all predictor features. Perfectly trained in order to classify lymphographic clinical data.
2010	Luo et al	Improved algorithm based on information entropy and attribute weights	Algorithm can improve the speed of classification. It enhances the accuracy. Higher computational time.
2009	Youn and Jeong	CDFW-NB-RFE Algorithm	Better when compared to other feature ranking algorithms on text datasets. Better computational efficiency is needed.
2012	Hara and Hayashi	E-Re-RX Algorithm	Deals with discrete and continuous attributes. It reduces the time complexity and produces good accuracy.

Proposed a new Decision Support System (DSS) based on the procedure of NN and data mining which consists of the three major steps. Initially, a new NN model is created to support DSS, and data mining algorithm is used to maintain DSS. Finally, both NN and data mining algorithm are combined to support DSS. The performance metrics are evaluated on the benchmark datasets.

Proposed a new NN classifier to forecast decomposition behavior of metal alloys and is trained on the fundamental laws with the purpose of mapping the alloy's composition with the decomposition rate. The corrosion information on corrosion acceptable and decomposition resistive alloys is presented together for both DC and AC corrosion experiments. The experimentation results on NN illustrate to classify and prioritize specific parameters such as pH, temperature, time of exposure, electrolyte composition, metal work, etc. and help to recognize the synergetic effects of the parameters and features on electrochemical potentials.

Introduced a new Ensemble-Recursive-Rule eXtraction (E-Re-RX) for Ensemble Neural Networks (ENNs). E-Re-RX is a successful rule extraction algorithm designed for dealing datasets that consists of both discrete and continuous attributes. Here, primary rules as well as secondary rules are created to manage only those data instances with the purpose of not satisfying the primary rules, and then these rules are integrated to original rule set. The experimental results reveal that the proposed ENNs classifier reduces the time complexity and produces higher accuracy when compared to traditional NNs neural networks.

Introduced the Map Reduce based Back propagation Neural Network (MRBNN) classification algorithm to classify the big data of mobiles into set of classes. Many experimentation works are performed using the Cloud Computing (CC) platform and results demonstrated that it provides higher performances in terms of efficiency, good scalability and anti-noise.
Seven different machine learning algorithms to predict the presence of chronic kidney disease of all the other models compared, Logistic regression, SGD Classifier and Random forest provide the best results.

IV. RESEARCH FINDINGS AND GAP

There is a need for an efficient Feature subsets selection technique and need for a completely automatic algorithm, which does not require any supervision in terms of threshold limits with to operate on the original feature space.The disadvantage of these Machine Learning algorithms is that they are time consuming and can be improved by using a metaheuristic optimization in the classification design.

V. CONCLUSION

From the review of the literature, it is noted that a lot of research on classification has arrived at improving the performance of classifiers. This paper presents the review of feature selection, conventional classifiers in a brief manner. In the traditional classifiers, the process of prediction is performed by learning the data associated with the problem of classification. These classifiers can effectively predict results in dynamic environments. The major limitations of these classifiers is that they are time consuming, so the meta-heuristic algorithms are used to find a solution among all possible ones. The future research therefore concentrates on overcoming the problems associated with the usage of traditional classification methods in Classification. Further enhancements are proposed to improve the performance of Classification.

REFERENCES

[1] Han, J. and Kamber, M., 2006. Classification and prediction. Data mining: Concepts and techniques, 347-350.
[2] Friedman, J., Hastie, T., and Tibshirani, R., 2001. The elements of statistical learning. Springer series in statistics.
[3] Liu, H., and Motoda, H., 2007. Computational methods of feature selection. CRC Press.
[4] Balasaraswathi, M., "Improved PSO based classifier for MultiClass Datasets", (Doctoral dissertation, Avinashilingam University, 2017.
[5] Weston, J., Elisseeff, A., Schölkopf, B., and Tipping, M., 2003. Use of the zero-norm with linear models and kernel methods. Journal of machine learning research, 3, 1439-1461.
[6] Song, L., Smola, A., Gretton, A., Borgwardt, K.M., and Bedo, J., 2007. Supervised feature selection via dependence estimation. In Proceedings of the 24th international conference on Machine learning, 823-830.
[7] Dy, J.G., and Brodley, C.E., 2004. Feature selection for unsupervised learning. Journal of machine learning research,845-889.
[8] Zhao, Z. and Liu, H., 2007. Semi-supervised feature selection via spectral analysis. In Proceedings of the 2007 Society for Industrial and Applied Mathematics (SIAM) International Conference on Data Mining, 641-646.
[9] Xu, Z., King, I., Lyu, M.R.T., and Jin, R., 2010. Discriminative semi-supervised feature selection via manifold regularization. IEEE Transactions on Neural networks, 21(7), 1033-1047.
[10] John, G. H., Kohavi, R., and Pfleger, K., 1994. Irrelevant features and the subset selection problem. Proceedings of the Eleventh International Conference on Machine Learning, 121–129.
[11] Lal, T., Chapelle, O., Weston, J., and Elisseeff, A., 2006. Embedded methods. Feature extraction, 137-165.
[12] Duda, R.O., Hart, P.E., and Stork, D.G., 2012. Pattern classification.John Wiley & Sons.
[13] Gu, Q., Li, Z., and Han, J., 2012. Generalized fisher score for feature selection. arXiv preprint arXiv:1202.3725.
[14] Zhou, M., 2016. A Hybrid Feature Selection Method Based On Fisher Score and Genetic Algorithm. Journal of Mathematical Sciences: Advances and Applications, 37,51-78.
[15] Chen, L., Man, H., and Nefian, A.V., 2005. Face recognition based on multi-class mapping of Fisher scores. Pattern Recognition, 38(6), 799-811.
[16] Liu, X., Ma, L., Song, L., Zhao, Y., Zhao, X., and Zhou, C., 2015. Recognizing common CT imaging signs of lung diseases through a new feature selection method based on Fisher criterion and genetic optimization. IEEE journal of biomedical and health informatics, 19(2), 635-647.
[17] He, X. F., Yan, S. C., Hu, Y. X., Niyogi, P., and Zhang, H. J., 2005. Face recognition using laplacian faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(3),328–340.
[18] Nie ,F., Xiang ,S., Jia ,Y., Zhang ,C., and Yan, S.,2008. TraceRatio Criterion for Feature Selection. In AAAI , 2, 671-676.
[19] Zhao, M., Zhang, Z., and Chow, T.W., 2011. ITR-Score algorithm: An efficient trace ratio criterion based algorithm for supervised dimensionality reduction. 2011 International Joint Conference on Neural Networks (IJCNN), 145-152.
[20] Arauzo-Azofra, A., Benitez, J.M., and Castro, J.L., 2004. A feature set measure based on relief. In Proceedings of the fifth international

conference on Recent Advances in Soft Computing, 104-109.
[21] Zhao, M., Chan, R.H., Tang, P., Chow, T.W., and Wong, S.W., 2013. Trace ratio linear discriminant analysis for medical diagnosis: a case study of dementia. IEEE signal processing letters, 20(5), 431-434.
[22] Zhao, M., Zhang, Z., and Chow, T.W., 2012. Trace ratio criterion based generalized discriminative learning for semi- supervised dimensionality reduction. Pattern Recognition, 45(4), 1482-1499.
[23] Rosario, S.F., and Thangadurai, K., 2015. RELIEF: Feature Selection Approach. International Journal of Innovative Research and Development, 4(11), 218-221.
[24] Baskar, S.S., and Arockiam, L., 2013. E LAS-Relief-A Novel Feature Selection Algorithm In Data mining. Compusoft, 2(12), 392-395.
[25] Stokes, M.E., and Visweswaran, S., 2012. Application of a spatially-weighted Relief algorithm for ranking genetic predictors of disease. BioData mining, 5(1), pp.1-11.
[26] Soliman, O.S., and Rassem, A., 2012. Correlation based feature selection using quantum bio inspired estimation of distribution algorithm. In International Workshop on Multi- disciplinary Trends in Artificial Intelligence, 318-329.
[27] Way, T.W., Sahiner, B., Hadjiiski, L.M., and Chan, H.P., 2010. Effect of finite sample size on feature selection and classification: a simulation study. Medical physics, 37(2), 907- 920.
[28] Jacob, S.G., and Ramani, R.G., 2011. Discovery of knowledge patterns in clinical data through data mining algorithms: Multi- class categorization of breast tissue data. International Journal of Computer Applications (IJCA), 32(7), 46-53.
[29] Jacob, S.G., Geetha Ramani, R., and Nancy, P., 2013. Discovery of knowledge patterns in lymphographic clinical data through data mining methods and techniques. Advances in computing and information technology, 129-140.
[30] Jin, C., De-Lin, L., and Fen-Xiang, M., 2009. An improved ID3 decision tree algorithm. 4th International Conference on Computer Science & Education, 2009 (ICCSE'09), 127-130.
[31] Chen, X.J., Zhang, Z.G., and Tong, Y., 2014. An improved ID3 decision tree algorithm. In Advanced Materials Research , 962, 2842-2847.
[32] Luo, H., Chen, Y., and Zhang, W., 2010, An Improved ID3 Algorithm Based on Attribute Importance-Weighted. 2nd International Workshop on Database Technology and Applications (DBTA), 1-4.
[33] Yang, F., Jin, H., and Qi, H., 2012, Study on the application of data mining for customer groups based on the modified ID3 algorithm in the e-commerce. 2012 International Conference on Computer Science and Information Processing (CSIP), 615-619.
[34] Yuxun, L., and Niuniu, X., 2010, Improved ID3 algorithm. 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), 465-468.
[35] Ali, M., Siarry, P., and Pant, M., 2012. An efficient differential evolution based algorithm for solving multi-objective optimization problems. European journal of operational research, 217(2), 404-416.t
[36] Youn, E., and Jeong, M.K., 2009. Class dependent feature scaling method using naive Bayes classifier for text datamining. Pattern Recognition Letters, 30(5), 477-485.
[37] Bermejo, P., Gámez, J.A., and Puerta, J.M., 2011. Improving the performance of Naive Bayes multinomial in e-mail foldering by introducing distribution-based balance of datasets. Expert Systems with Applications, 38(3), 2072-2080.
[38] Kibriya, A.M., Frank, E., Pfahringer, B., and Holmes, G., 2004, Multinomial Naive Bayes for Text Categorization Revisited. In Australian Conference on Artificial Intelligence, 3339, 488- 499.
[39] Deekshatulu, B.L., and Chandra, P., 2013. Classification of heart disease using k-nearest neighbor and genetic algorithm. Procedia Technology, 10, 85-94.
[40] Shouman, M., Turner, T., and Stocker, R., 2012. Applying k-nearest neighbour in diagnosing heart disease patients. International Journal of Information and Education Technology, 2(3), 220-223.
[41] Imandoust, S.B., and Bolandraftar, M., 2013. Application of k-Nearest Neighbor (KNN) approach for predicting economic events: Theoretical background. International Journal of Engineering Research and Applications, 3(5), 605-610.
[42] Wang, L., Khan, L., and Thuraisingham, B., 2008, An effective evidence theory based k-nearest neighbor (KNN) classification. In Proceedings of the 2008 IEEE/WIC/ ACM International Conference on Web Intelligence and Intelligent Agent Technology, 1, 797-801.
[43] Hu, L.Y., Huang, M.W., Ke, S.W., and Tsai, C.F., 2016. The distance function effect on k-nearest neighbor classification for medical datasets. SpringerPlus, 5(1304), 1-9.
[44] Qian, X., and Wang, X., 2009. A new study of DSS based on neural network and data mining. International Conference on E-Business and Information System Security, 2009 (EBISS›09),1-4.
[45] Kamrunnahar, M., and Urquidi-Macdonald, M., 2010. Prediction of corrosion behavior using neural network as a data mining tool. Corrosion Science, 52(3), 669-677.
[46] Hara, A., and Hayashi, Y., 2012. A new neural data analysis approach using ensemble neural network rule extraction. Artificial Neural Networks and Machine Learning–ICANN 2012, 515-522.
[47] Liu, Z., Li, H., and Miao, G., 2010. MapReduce-based backpropagation neural network over large scale mobile data. 2010 Sixth International Conference on Natural Computation (ICNC), 4, 1726-1730.
[48] Bhawna Sharma1, Sheetal Gandotra, Utkarsh Sharma, Rahul Thakur, Alankar Mahajan, A Comparative Analysis Of Different Machine Learning Classification Algorithms For Predicting Chronic Kidney Disease, International Journal of Computer Sciences and Engineering, E-ISSN: 2347-2693, Vol.-7, Issue-6, June 2019

Authors Profile

Mrs. M. Balasaraswathi completed M.Sc.,. in Computer Science from Bharathiar University in 1996 and M.Phil., in Computer Science from Alagappa University in the year 2008. She has done Ph.D in Computer Science from Avinashilingam University, Coimbatore in 2018 and is working as Head of Computer Science Dept. in Rathinam College of Arts and Science, Coimbatore. She has done Ph.D in Computer Science from Avinashilingam University, Coimbatore in 2018. She has around 20 years of teaching experience at the post graduate and under graduate levels and worked in abroad INTI College Malaysia for a period of 5 years. Her Specialization is Data Mining and published papers in National and International Journals.

A.Uthiramoorthy is an Assistant Profes of Computer Science at Rathinam College of Arts and Science (Autonomous), Coimbatore, India. His research interests include Data Mining, Data Analytics, and Network Security. He received his M.Phil. (Data Mining in Computer Science), Madurai Kamaraj University and Completed MCA (Computer Applications) from Periyar University, Salem,TN, India. He has more than twelve years of teaching experience in Computer Science. He has published 4 research articles in leading journals, and more than 20 presented papers in National and International conferences.

