Feature Selection for Classification of Abnormalities in Medical Images – A Review

Suja K V, Rajkumar K K

Department of Information Technology, Kannur University, Kannur, Kerala, India

Corresponding author: Suja K V, Email: sujakv@kannuruniv.ac.in

In medical imaging, Computer Aided Diagnosis (CAD) has become one of the major research topics but is still in the infancy stage due to the lack of its full potential for applications to analyze the lesions obtained from various modalities. Pattern recognition and computer vision plays a significant role in clinical procedures for detecting and diagnosing different human diseases through the processing and analyzing of images acquired through various medical imaging modalities. In many cases of medical applications having high dimensional data characterized by huge number of features require large amount of memory and computation power. In order to tackle this problem, the aim is to construct a combination of feature that builds a unique model to provide better classification performance and accuracy. In this paper, we have conducted a survey on widely used approaches for feature selection and analyzed the purpose to investigate the strength and weakness of existing methods used in different types of modalities of images. Most of the work discussed in this literature review faces many limitations such as accuracy, cost, time and storage when dealing with huge amount of data. Our prime intention is to tackle these problems by building a uniform modal for feature selection to rank the features which are extracted from different medical image modalities to detect and diagnose the abnormalities present in those images in a most efficient way.

Keywords: Dimensionality reduction, Feature selection, Feature ranking, Medical Imaging.

2023. In Saroj Hiranwal & Garima Mathur (eds.), *Artificial Intelligence and Communication Technologies*, 29–38. Computing & Intelligent Systems, SCRS, India. https://doi.org/10.52458/978-81-955020-5-9-3

1. Introduction

The fast and continuous advancement in computerized medical image envision, advances in analysis methods and Computer Aided Diagnosis (CAD) led to the development of medical imaging as one of the most important fields [1]. The medical images acquired from digital signals such as Computed Tomography (CT), X-Rays, Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), ultrasound has given rise to the development of automatic processing and analysis of medical images [2]. The process of retrieving meaningful information embedded in these images generates huge volume of data. Interpreting and analyzing the characteristics of images by using those data is one of the challenging tasks due to the size of data [3]. Diagnosing and prediction of the disease in the early stage become too tedious job due to the size of the data set which leads classification more complex. To reduce the complexity of classification, the volume of data is to be reduced. Dimensionality reduction is one such method that reduces the volume of the data by eliminating redundant and less significant information from the original dataset by representing it in a lower dimension using dimensionality reduction techniques to retain the most essential information [4]. Presently there are various algorithms to solve this problem by developing an automated CAD system which consists of a compact set of feature selection techniques and classification. The ability to detect and classify pathological patterns in medical images in a more effective and unique manner is the key feature of this system.

Feature extraction is a method that replaces the original feature by transforming high dimensional feature space into a lower dimension that represents the complete information regarding the dataset [5]. Feature selection is a vital step in dimensionality reduction which is used for selecting the best minimal subset of original features based on some specific criteria without altering the information of original features [6] [7]. Maximization of relevance and minimization of redundancy by retaining the minimum number of features for the estimation process is the objective of feature selection [8] [9]. It is to be evolved as a new feature selection method by ranking among the different feature used in detection and classification process. The efficiency of machine learning/classification can be increased by removing irrelevant features and retaining the most relevant feature based on the rank of the feature obtained from the feature ranking method [10]. So our prime intention is to tackle these problems, by building a uniform method for feature selection by ranking the features on different modalities of images and to detect and diagnose the abnormalities present in an image. Such a ranking algorithm will be most effective for detecting and diagnosing abnormalities present in medical image dataset. Rough set theory, hybrid genetic algorithm and fuzzy rough set are some of the mathematical tools that can be utilized for ranking the features from high dimensional to minimal number of features that improve performance of classification by providing more accurate result.

The outline of the paper is structured as detailed: section 2 presents the general procedure for feature selection to improve the classification algorithm’s performance. Section 3 presents classification of feature selection methods, in section 4 the literature review of related works that have been performed on different modalities of medical images to analyze various feature selection algorithms are presented. Section 5, explains the merit and demerits of the existing methods followed by the proposal of new model to overcome the demerit identified during these surveys.

2. General Procedure for Feature Selection

Feature selection is the most important technique that is applied for pre-processing task in data mining and has become the crucial element of machine learning algorithm [11]. Feature selection algorithms eliminates redundant and irrelevant features and sort out the minimum number of impressive features that maintain the intent of original features [12][13].

The general procedure for feature selection can be explained in four steps such as: Generation and Evaluation of Subset, Criteria for Stopping and Validation of result [10]. Based on a certain searching strategy a heuristic search procedure called subset generation generates candidate feature subset [14]. There are two basic issues such as successor generation and search organization to determine candidate subset. Successor generation mechanism proposed possible successor candidates of the current hypothesis. Forward, backward, compound, weighting and random methods are the five different operators for successor generation [15][16]. The first operator forward, starts with null feature and then features are added successively to the selected feature [10]. The second operator backward, eliminates the least significant feature and makes the performance of the model best with full set of features in the start [17]. In the third operator compound, 'k' consecutive forward step and 'l' consecutive backward step is applied until a stopping criterion is reached, the fourth one weighting, represent all the features in the result to a certain degree by iteratively sampling the available set of instances and the operator random is a group which can possibly produce any other state in a single step by considering all the four operators until a minimum criterion is attained [10][17]. The search organization is liable for the feature selection process with a precise approach by considering search algorithms like exponential, sequential or random search [10]. In the second step, the candidate feature subset generated is evaluated and compared against the previous subset generated using evaluation criterion like dependence, divergence, interclass distance, information or uncertainty, probability of error and consistency [15][14]. If the score of new subgroup happens to be superior, it then substitutes the preceding one. The subset generation and evaluation process is repeated until the stopping criteria are reached [12]. The stopping criteria can be set as predefined minimum number of features or minimum classification error rate, maximum number of iterations or if the addition or removal of features to the subgroup do not produce a major difference or including all features in the problem [10]. Finally, the validation of best features that are selected must be carried out by different tests on both the original set and the selected candidate subset and comparing the results of contesting methods using real world, artificial datasets, or both [18].

The feature selection algorithm that evaluates subset of features from original features may be univariate or multivariate [11]. The performance of each feature is evaluated individually in univariate scheme, whereas in multivariate scheme subsets of features are evaluated [7][11][19]. Multivariate method is capable of handling redundant features; they may result in less predictive performance due to over fitting [19]. Most of the feature selection algorithms adopt some kind of search algorithm, varying in search strategy and evaluation measure [20].

3. Classification of Feature Selection Method

The processes of choosing a subset of relevant features without any transformation by maintaining the real meaning of original features is called as feature selection [8][21][13]. It leads to better prediction accuracy, lower computational cost, limits the size of the data, less execution time and reduced storage [19]. Generally, for the given 'n' features, the feature selection method is to identify the optimum subset among 2n probable options [12][21]. A substantial amount of feature selection algorithms has already proposed and achieved the state-of-the art results in many different fields such as image processing, data mining, bioinformatics, text categorization etc. [14]. A good feature subset evolved from the feature transformation method will be highly correlated with the decision feature compared to the original feature set [22]. In order to identify such decision features from the original feature set, there exist several different methods such as filter method, wrapper method and embedded/hybrid methods. In general, filter method is as a feature ranking whereas embedded and wrapper methods are used for feature subset selection [12][21]. Genetic Algorithm (GA), sequential backward selection, sequential forward, Particle Swarm Optimization (PSO) are the commonly used feature selection algorithms [23].

a. Filter Method

This method is the oldest feature selection method which is independent of any classification model during the feature selection process [9] [24]. Filter method rely on the characteristics of the data alone to assess the importance of feature and uses ranking technique as the principal criteria to rank the features [11][25]. A typical filter method consists of two steps: A score is assigned to each feature according to some feature evaluation criteria and then selecting the features having score above a threshold value in the first step and high ranked features are retained by eliminating low scored feature for further classification in the second step [11][25][26][27]. Filter methods have the advantage that it can easily scalable with high dimensional dataset for high processing speed and efficient result on execution with minimum time [24][28]. The main disadvantage of this method is that it ignores the interaction between classifiers and feature dependencies [28].

b. Wrapper Method

Unlike filter method, Wrapper method uses the achievement of learning algorithm as the estimation criteria [21][29]. This approach is better in defining optimal features rather than relevant features [26]. It evaluates its goodness by applying classification algorithms like Naive Bayes, Support Vector Machine etc. for each subset generated or evaluating the subsets based on the performance of a clustering algorithm, which is considered as a black box evaluator [8][22]. Therefore, different selection procedures and classification techniques will produce different sets of optimized feature sets [22]. It has been empirically proven that wrapper method obtains subset of features with better performance than filter method. One of the limitations of wrapper method is that when dealing with large number of features it takes high computation time for generating subset of features [8] [12]. Wrappers are too expensive to be employed in large dataset as all attribute set have to be tested with trained classifiers that make the process of features selection too slow [30]. Wrapper method is categorized as sequential selection algorithms and heuristics search algorithms. These include Ant Colony Optimization (ACO), GA, PSO [26].

c. Hybrid/ Embedded Method

To achieve better classification performance, hybrid method combines best properties of filter and wrapper methods [22]. Initially, filter method is applied to obtain several candidate subset by reducing the feature space. Once the set of candidate subset are generated then wrapper method is applied on these subsets to identify the best subset from the candidate subsets [8]. For evaluating the quality of features and for selecting the features, learning algorithm is employed in this method. It is computationally less expensive than wrapper approaches and is much slower than filter approaches [19].

4. Literature Review

The primary intention of this literature survey to identify the various feature selection algorithm based on feature ranking to select the best minimum subset of original feature to lead better classification performance. The conventional feature extraction techniques extract different number of features either by feature selection or transformation that degrades the performance while classifying the dataset in terms of accuracy, cost, time and storage. This type of feature extraction reduces the number of features to a limited extend, dealing those set of features for classification degrade the overall performance of the classification. Further the feature extraction/transformation methods extract a limited set of features by transforming original feature set into a different domain by applying dimensionality reduction technique on the extracted feature. So our primary concern is to identify and reduce optimal number of features based on the relevance of those feature ranking criteria lead better classification performance for detecting the abnormalities present in the medical images without changing the domain of the features.

A novel feature ranking technique was proposed by [20] using attribute frequency information in rough set theory for reduct computation. This mechanism is based on two heuristic reduct computation algorithms: optimal reduct and approximate reduct. The optimal reduct algorithm is developed using the significance of attributes as heuristics and also the information of discernibility matrix. Approximate reduct algorithm is applicable to large datasets and increase the classification accuracy and performance. Sampling and feature ranking mechanism are used to make the algorithm applicable to large datasets. This algorithm works in two phases- generating phase and testing phase. In first phase, several samples of original dataset are selected and count the frequency of every attribute using discernibility matrix. These frequencies are later sorted. In phase two, sampling method is used in which more samples are used together with frequencies counted in phase 1 to produce a good approximate reduct for very large datasets. Experimental results on 45 UCI datasets show that optimal reduct is applicable to middle-sized datasets and approximate reduct is applicable to large datasets. It also shows that an attribute reduction of 75% is possible and the error increases only by 2.7%.

A novel approach using two-stage hierarchical ranking procedure was proposed [31] for the feature selection and ranking with abductive network training algorithm based on Group Method of Data Handling (GMDH). In first phase of the algorithm, features are ranked in groups by the order in which they are selected by a GMDH learning algorithm and in the second phase features within each group are ranked by iterating the method for only the features within the group used as model inputs. The lists of ranked features derived from second phase are utilized to decide the subset of an optimum features, which possess minimum classification error rate on a dedicated evaluation set. For heart disease and breast cancer data set, an optimum subset had given 54% and 56% feature reduction and it increases the overall classification accuracy from 82.5% to 85% and 96.5% to 97.5% respectively. For Dimensionality reductions introduced in both datasets have no major reduction in the area under ROC curve. Feature reduction methods developed was tested with only two data sets; it has to be tested with different learning algorithms using other medical datasets too.

A rough set based feature reduction algorithms such as Quickreduct (QR) and Modified Quickreduct (MQR) was introduced by [32]. In Quickreduct algorithm, the reduction of attributes is done by comparing equivalence relations generated by sets of attributes which results only vertical reduct where only the unwanted attributes are removed, whereas in Modified Quickreduct algorithm, by eliminating the objects, the size of the information system can also be minimized horizontally. The proposed algorithm is tested using Mini MIAS images and identified that both these methods reduce the dataset efficiently without losing essential information. They also identified that among the two methods, Modified Quickreduct reduces the dataset to minimal number of attributes when compared with Quickreduct. The significance of attributes and objects are not considered in this work while removing those elements from the system.

Three rough set based algorithms namely Quick Reduct Algorithm (QR), Entropy Based Reduct Algorithm (EBR) and Relative Reduct Algorithm (RRA) for feature selections was proposed by [33]. Each attributes dependency is calculated using QR algorithm and the best candidate is selected by starting with a null set and then one attribute is added at a time that results in highest increase in the rough set dependency metric, until it produces the maximum possible value for the dataset. In the second approach, dataset is examined and the attributes that contribute the most gain in information is identified. Until the resulting subset entropy is equal to that of the entire feature set, the search for the finest feature subset is continued. The proposed third approach was to avoid the calculation of discernibility functions or positive regions, which is computationally expensive without optimization. They observed from the result that EBR and QR algorithms perform well and the performance for the GLCM is highest in the direction 450.

A novel Genetic Algorithm (GA) based Feature Selection (FS) framework for improving the efficiency and effectiveness of the Content Based Image Retrieval (CBIR) system was proposed by [12]. It employs a wrapper method and had established a set of evaluation functions for CBIR. Based on the ranking evaluation function, three new fitness functions were proposed and embedded with Genetic Algorithm. It is then compared with other algorithms such as GA-based FS algorithm for minimization of error of traditional classifiers. The result shows that the projected method considerably out performs all the others. The experimental results show that the practice of using huge quantity of features to denote a medical image can extremely decrease the content-based search's precision. The result also shows that the cost of query processing finally is also reduced. The technique projected here increases the precision of similarity searches and reduces the dimensionality of data significantly. This improves both the efficiency of the CBIR system as well as the efficacy of the access methods. Further the efficiency of the work can be improved by using local search in GA and discover for CBIR usage, the interaction among GA wrapper and filter methods and combining the textual information of the history of patient's clinical data and exams into the similarity search mechanism.

Alijla et al. [34] had applied fuzzy rough set method for choosing the most important texture features from mammogram images. Two steps of fuzzy rough set approaches are used in the feature selection process. In the first step, similar characteristics of objects is found with respect to the subclass of features and in the second step, to calculate the degree of dependency of objects on a feature, approximation decision concepts are used. Significance of selected features was evaluated by passing it into two common classifiers J48 and Vector Quantization Neural Network (VQNN). The result shows the classification accuracy of 94.22% is obtained using J48 classifier and 94.60% for VQNN classifier. To confirm the importance of other statistical features, further investigations are essential to attain enhanced accuracy of classification.

A hybrid method of feature ranking was proposed by Barakat [35], utilizing support vectors (SVs) of Support Vector Machines (SVMs). The idea behind this procedure is implemented in two steps so as to categorize and rank the features that differentiate least between positive and negative class. At first, the method identifies the subclass of features that least contribute to interclass separation. Then, correlation based feature selection technique is used to re-rank the selected features. Experimental result shows better performance in classification for most of the medical data set. One demerit of this method is that optimization of the SVM training parameters is a serious issue in finding the optimum separation between classes.

A volumetric feature evaluation to detect the pertinent features for the accurate discovery of depression by using Degree of Contribution (DoC) calculation algorithm was presented by [36]. The basis of the procedure is on the frequency of each feature contributed towards the preferred limit of accuracy to determine the ranking of the features selected from the structural Magnetic Resonance Imaging (sMRI) dataset. Features are all ranked using four Feature Selection (FS) techniques and the proposed feature selection is centred on the value of DoC that shows the accurate assistance using the DoC ranking to FS. Final feature ranking in DoC algorithm is produced using various evaluators-classifiers, after the multi-rule evaluation. Before final ranking is done, it considers various possible combinations and this is the advantage of algorithm, which makes it produce more accurate result. To estimate the efficiency of the projected DoC, forty-four volumetric features from different brain regions were collected. The accuracy (ACC) result of 88.23% was obtained by using this algorithm, which is 3% superior to the existing algorithms.

A hybrid forward selection technique to estimate more precisely the existence of cardiovascular disease was proposed by [22]. Objective of the proposed technique was to identify appropriate algorithm that produces reduced feature subset from huge dimension of data with enhanced diagnosing ability. The features are ranked using three feature selection algorithm such as forward feature inclusion, backward elimination and forward selection. It is then included in the feature subset by using wrapper method including SVM classifier. The experiment is carried out on arrhythmia, SPECT cardiac and heart disease dataset. The result obtained shows highest accuracy and the dimensionality of data is reduced with forward inclusion and back-elimination feature selection techniques for arrhythmia and heart disease datasets.

For choosing the most significant features earlier than the derivation of classification predictors, a feature ranking algorithm was developed by [37]. The algorithm uses a ranking criterion which is based on a scoring function, to estimate the correlation measure between the classes and each feature. The effectiveness of the approach was tested by selecting few top-ranked features on some standard data sets of breast cancer with high dimensions. It is proved that ensemble feature selection can be used in large data sets that are intractable (inflexible or difficult) by some feature selection procedure. The study shows that the best ways to assess a classifier performance is Area under the curve. This approach enables the usage of all known feature ranking in large datasets. It also suggests that in many cases feature ranking can reduce the correlated variables. To check the robustness of proposed FS method, it can be applied for testing other datasets too.

A novel framework for feature reduction for choosing the most selective features that promises excellent analysis of breast masses using reduced dimension was proposed by [38]. The authors applied histogram equalization and nonlocal mean filter on each ROI as a pre-processing step and five feature ranking methods are applied for the 109 features extracted from each suspicious area. Highest scored feature was ranked first among other features that shared the same number of occurrences. Further reducing the set of features, they found the significant features that were utmost useful for classification. The database they used were retrieved from IRMA database and they found that only 49 out of the 109 extracted features are enough to attain the accuracy of 94.27% using Feed forward Neural Network classifier. If these 49 features can still be reduced more accuracy could have achieved for classification.

Feature selection methods based on effective distance was proposed by [39] by substituting conventional distance with effective distance. The proposed method consists of two steps: At first, to build a bi-directional network, an effective distance is obtained by developing a sparse representation-based algorithm. Secondly, they developed three novel unsupervised filter-type feature selection methods including two Effective Distance-based Sparsity Scores (EDS)S-1, and EDSS-2) and Effective Distance based Laplacian Score (EDLS) by using effective distance as the similarity measure. Result of the experiment was tested on UCI machine learning repository data sets. The clustering on K-means algorithm and classification on a strategy of 5-fold cross validation on a series of datasets are used to determine the efficiency of the proposed methods. This method is compared with other methods using conventional Euclidean distance. Further the authors suggested that the proposed computation method could be used in spectral clustering and dimensionality reduction.

A Genetic Algorithm (GA) approach for hybrid content based medical image retrieval system for the selection of dimensionally compact set of features was presented by [40]. The system was implemented in three phases. In the phase I, three different algorithms are used to extract the dynamic features from the images. In the phase II, to find the feature vector, GA based feature selection is done using a hybrid approach of "Artificial Bee Colony Algorithm" and "Branch and Bound Algorithm" is applied on three different images such as brain tumor, breast cancer and thyroid images. In the phase III, diverse density-based relevance feedback method is used to increase the performance of the hybrid content based medical image retrieval system. The results of the experiment show that the GA driven image retrieval system chooses the best subset of feature to find the true set of images. In this work, even though they were able to retrieve most relevant images, some obvious relevant images were also missing. By using more advanced algorithms the relevancy can be improved.

A novel method for ranking features based on Markov networks in the multi-label setting was proposed by [41]. The method contains two steps: in step one, using Ising model containing only labels, a markov network is build. In step two, the effect of addition of a particular feature to the initial network based on score statistic is tested. Ising model is used in this work for ranking the features and it is done in three approaches: The approach first followed is centred on the Ising model with score statistic and constant interaction terms. The approach secondly followed is based on the Ising model with score statistic and feature-dependent interaction terms. Computationally most expensive is the third approach which is centred on fitting l1 standardized logistic regressions. All these feature ranking procedures order the features with respect to their importance. The experimental result on real and artificial data indicates that the projected techniques can overtake the traditional one. This approach can be suggested, particularly for datasets with reasonable number of labels and also for huge number of features. The key problem related with general Markov networks is by which method the estimation of parameters and the significance of features can be tested efficiently.

A filter-based feature selection approaches were proposed by [42] based on Evolutionary computation (EC), feature ranking techniques and information theory. By using a novel filter evaluation criterion this approach was successfully created on the concepts of ReliefF, Fisher Score and Mutual relevance. Further, an extensively used current filter based criterion called Mutual Information Feature Selection (MIFS) is also reformed as fitness function for single and multi-objective Differential Evolution (DE) to improve filter based approaches. To evaluate the feature selection approaches, one text classification data, one biomedical data and ten datasets from UCI machine learning repository are selected, including different numbers of samples, features, and classes. The experimental result shows that, DE based on the projected criteria overtook the existing criteria in terms of classification accuracy and the number of features. The result also shows that as a multi objective model, feature selection offers improved output in terms of classification accuracy and feature subset size.

A detailed literature survey is made on different algorithms of feature selection techniques using various medical imaging modalities. The researchers especially in the field of computer vision and pattern recognition face hurdles in building more accurate classification model for interpreting and analyzing medical images due to its huge volume of data. Most of the research papers discussed in the reviews utilizes Genetic Algorithm, Rough Set Theory, Fuzzy Rough Set for feature selection that reduce the huge volume of data. These techniques basically decrease the dimension of data by selecting the most relevant features from the high dimensional data. The ability to develop new features from large set of features and an enormous examination of the search space for new filter solution makes GA different from another feature selection algorithm [43]. Rough Set Theory (RST) can be used for feature extraction, feature selection, data reduction, pattern extraction, decision rule generation [44]. Fuzzy set are the tools that convert the concept of fuzzy logic into algorithms and as such, it can provide the computer with algorithms that extend the binary logic and make the computer to take decisions like human being [45]. Even though all these methods had established well desired outcome but requires more exploitation of all these methods for further improvement in this direction. For detecting abnormalities in medical images without changing the domain of the features such a model effectively provides much more classification accuracy and performance. So we aim to propose a novel model to tackle the problems when dealing with huge volume of data, by building a feature ranking algorithm which is an individual evaluation method that assigns weight for each and every feature according to its degree of significance. This can improve the efficiency of classification for predicting the abnormalities in various medical imaging modalities based on different methods such as Fuzzy-Rough Set, Rough Set Theory, Hybrid Genetic Algorithm.

5. Discussion and Conclusion

In this study, we reviewed various feature selection techniques such as fuzzy set, rough set, genetic algorithm etc in different modalities of images. Some researchers have even pooled the rough set theory with artificial intelligence approaches such as genetic algorithm, fuzzy logic, neural networks additional to the other methods that results in good classification accuracy and performance. Even though they were able to make improvement in classification performance, there were limitations such as cost, time and storage while dealing with huge volume of data. To overcome these limitations, it would be worthwhile if we develop an efficient ranking algorithm that determines the best features for attaining higher classification accuracy. Therefore, our primary intension is to devise and implement an efficient feature ranking algorithm, by combining the theory of Rough Set, Fuzzy-Rough Set, Hybrid Genetic Algorithm, Fuzzy Rough feature selection using Support vector Machine (SVM) and filter-based feature ranking techniques. This makes significant changes in the execution and the classification performance for medical images in different modalities by determining only the best feature with its significance for the classification model developed.

References

[1] Dougherty G, Image analysis in medical imaging: recent advances in selected examples, Biomedical Imaging and Intervention Journal, Review Article, 6(3): e32, (2010)

[2] Mary Joans S, J. Sandhiya, A Genetic Algorithm Based Feature Selection for Classification of Brain MRI Scan Images Using Random Forest Classifier. International Journal of Advanced Engineering Research andScience (IJAERS), Vol-4, Issue-5, (2017) https://dx.doi.org/10.22161/ijaers.4.5.20, pp 124-130

[3] Katerine Diaz-Chito, Jesus Martinez del Rincon, Marçal Rusinol, Aura Hernandez-Sabate, Feature Extraction by Using Dual-Generalized Discriminative Common Vectors, Journal of Mathematical Imaging and Vision, (2018), https://doi.org/10.1007/s10851-018-0837-6.

[4] Jiawei Han, Micheline Kamber, Jian Pei, Data Mining Concepts and Techniques. ISBN – 978-0-12-381479-1, 3rd Edition, Elsevier, 744 pages, (2011).

[5] Min Zhu, Jing Xia, Molei Yan, Guolong Cai, Jing Yan, Gangmin Ning, Dimensionality Reduction in Complex Medical Data: Improved Self-Adaptive Niche Genetic Algorithm. Computational and Mathematical Methods in Medicine, Hindawi Publishing Corporation, Article ID 794586, 12 pages, (2015).

[6] Chunren Lai, Shengwen Guo, Lina Cheng, Wensheng Wang, A Comparative Study of Feature Selection Methods for the Discriminative Analysis of Temporal Lobe Epilepsy. Frontiers in Neurology, Article 633, Volume 8, (2017).

[7] Shobeir Fakhraei, Hamid Soltanian Zadeh, Farshad Fotouhi, Bias and stability of single variable classifiers for feature and selection. Expert Systems with Application, Elsevier, pp.6945-6958, (2014).

[8] Jovic A, K. Brkic and N. Bogunovic, A review of feature selection methods with applications. 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 1200-1205, (2015).

[9] Haoyan Guo, Yuanzhi Cheng, Dazheng Wang, Li Guo, A Practical Feature Selection Based on An Optimal Feature Subset and Its Application for Detecting Lung Nodules in Chest Radiographs. 6th International Conference on Biomedical Engineering and Informatics (BMEI), pp.501-518, (2013).

[10] Vipin Kumar, Sonajharia Minz, Featured Selection: A Literature Review, Smart Computing Review, Vol 4, No.3, pp.211-229, (2014).

[11] Jundong Li, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert P. Trevino, Jiliang Tang, Huan Liu, Feature Selection: A Data Perspective, ACM Computing Surveys, Vol. 50, No. 6, Article 94, (2017), https://doi.org/10.1145/3136625

[12] Sergio Francisco da Silva, Marcela Xavier Ribeiro, Joao do E S Batista Neto, Caetano Traina-Jr, Agma J M Traina, Improving the ranking quality of medical image retrieval using a genetic feature selection method. Published by Elsevier B V, Springer Briefs in Applied Sciences and Technology, Decision Support Systems (2011).

[13] Cao Truong Tran, Mengjie Zhang, Peter Andreae, Bing Xue, Lam Thu Bui. "Improving performance of classification on incomplete data using feature selection and clustering", Applied Soft Computing, 2018

[14] Bharat Singh, Nidhi Kushwaha, Om Prakash Vyas, A Feature Subset Selection Technique for High Dimensional Data Using Symmetric Uncertainty. Journal of Data Analysis and Information Processing, 2, pp. 95-105, http://dx.doi.org/10.4236/jdaip.2014.24012.(2014)

[15] Ladha L, T. Deepa, Feature Selection Methods and Algorithms. International Journal on Computer Science and Engineering (IJCSE), Vol. 3 No. 5, pp.1787-1797, (2011).

[16] Luis Carlos Molina, Evaluating Feature Selection Algorithms, Lecture Notes in Computer Science, 2002

[17] Luis Carlos Molina, Lluis Belanche, Angela Nebot, Feature Selection Algorithms: A Survey and Experimental Evaluation, 2002 IEEE International Conference on Data Mining, 2002. Proceedings., DOI: 10.1109/ICDM.2002.1183917, ISBN: 0-7695-1754-4

[18] Jasmina Novakovic, Perica Strbac, Dusan Bulatovic, Toward Optimal Feature Selection Using Ranking Methods and Classification Algorithms. Yugoslav Journal of Operations Research, pp.119-135, (2011).

[19] Jiliang Tang, Salem Alelyani and Huan Liu, Feature Selection for Classification: A Review. In Data classification: Algorithms and Applications, pp. 1-37, (2014).

[20] Keyun Hu, Yuchang Lu and Chunyi Shi, Feature ranking in rough sets. AI Communications IOS Press, pp.41-50, (2003).

[21] De Silva, Anthony Mihirana, Philip HW Leong, Grammar-based feature generation for time-series prediction. Springer Briefs in Applied Sciences and Technology-Computational Intelligence, Singapore, (2015).

[22] Swati Shilaskar, Ashok Ghatol, Feature selection for medical diagnosis: Evaluation for cardiovascular diseases, Expert Systems with Applications, Volume 40, Issue 10, Elsevier Ltd, pp.4146-4153, (2013).

[23] Geethu Mohan, M Monica Subashini, MRI based medical image analysis : Survey on brain tumor grade classification. Biomedical Signal Processing and Control, pp 139-161, (2018).

[24] Divya Jain, Vijendra Singh, Feature selection and classification systems for chronic disease prediction: A review, Egyptian Informatics Journal, 2018

[25] Pui Yi Lee, Wei Ping Loh, Jeng Feng Chin, Feature Selection in multimedia: The-state-of-the-art-review", Image and Vision Computing 67, pp.29-42, (2017).

[26] Pradnya Kumbhar, Manisha Mali, A Survey on Feature Selection Techniques and Classification Algorithms for Efficient Text Classification. International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064, Vol.5, Issue.5, pp.1267-1275, (2016).

[27] Huu-Thanh Duong, Vinh Truong Hoang, "Dimensionality Reduction Based on Feature Selection for Rice Varieties Recognition", 2019, 4th International Conference on Information Technology (InCIT), 2019

[28] Visalakshi S, V Radha, "A Literature Review of Feature Selection Techniques and Applications: Review of feature selection in data mining", IEEE International Conference on Computational Intelligence and Computing Research, (2014).

[29] Anthony Mihirana De Silva, Philip H. W. Leong. "Grammar-Based Feature Generation for Time Series Prediction", Springer Science and Business Media LLC, 2015

[30] Lenin Babu R, A novel feature selection technique for web page based medical information retrieval system. Ph.D Thesis, 151p, (2016).

[31] Abdel-Aal R.E, GMDH–based feature ranking and selection for improved classification of medical data. Journal of Biomedical Informatics 38, published by Elsevier, pp.456-468, (2005).

[32] Pethalakshmi A, K. Thangavel, P. Jaganathan, Mammography Feature Selection using Rough set Theory. IEEE 14th International Conference on Advanced Computing and Communication, pp.244-249, (2006).

[33] Raja K.T Keerthana, K Thangavel, Feature selection in mammogram image using rough set approach. IEEE Proceedings of National Conference on Innovations in Emerging Technology, pp .147-152, (2011).

[34] Basem O. Alijla; Ahamad Tajudin Khader, Lim Chee Peng, Mohammed Azmi Al- etar, Wong Li Pei, Fuzzy rough set approach for selecting the most significant texture features in mammogram images. IEEE Palestinian International Conference on Information and Communication Technology, pp.51-56, (2013).

[35] Nahla Barakat, Feature ranking utilizing support vector machines' SVs. IEEE Third International Conference on Innovative Computing Technology (INTECH), pp.401-406, (2013).

[36] Kuryati Kipli, Abbas Z Kouzani, An Algorithm for Determination of Rank and Degree of Contribution of sMRI Volumetric Features in Depression Detection. 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS) Osaka, (2013).

[37] Vitor Santos, Nuno Datia, M P M Pato, Ensemble Feature Ranking applied to medical data. Procedia Technology 17, 223-230, Published by Elsevier Ltd, (2014).

[38] Hajar Alharbi, Gregory Falzon, Paul Kwan, A novel feature reduction framework for digital mammogram image classification. IEEE 3rd IAPR Asian Conference on Pattern Recognition, pp.221-225, (2015)

[39] Mingxia Liu, Daoqiang Zhang, Feature Selection with Effective Distance, Neurocomputing, Elsevier publication, pp.100-109, (2016).

[40] Nagarajan G, R.I. Minu, B Muthukumar, V.Vedanarayanan, S.D. Sundarsingh, Hybrid Genetic Algorithm for Medical Image Feature Extraction and Selection. International Conference on Computational Modeling and Security (CMS 2016), Procedia Computer Science 85, Elsevier B.V Publication, pp. 455–462 (2016).

[41] Pawel Teisseyre, Feature ranking for multi-label classification using Markov networks. Neurocomputing, Elsevier B.V, pp.439-454, (2016).

[42] Emrah Hancer, Bing Xue, Mengjie Zhang, Differential evolution for filter feature selection based on information theory and feature ranking. Knowledge-Based Systems, Volume 140, Elsevier.B V, pp.103-119, (2017).

[43] Sindhiya S, S Gunasundari, A Survey on Genetic Algorithm Based Feature Selection for Disease Diagnosis System, Proceedings of IEEE International Conference on Computer Communication Systems (ICCCS'14), pp.164-169, (2014).

[44] Zdzisław Pawlak, Rough sets and intelligent data analysis, Information Sciences 147, pp: 1–12 ,Elsevier Science Inc, (2002).

[45] S.N Sivanandam, S.N Deepa, Principles of Soft Computing. John Wiley and Sons, ISBN: 978-81-265-2741-0, (2008)