CHAPTER-II

Review of Literature
CHAPTER-II

REVIEW OF LITERATURE

2.1 Introduction

An essential aspect of a research project is the review of the related literatures. Such a review represents the step of the scientific method. The survey of the literature is a crucial aspect of the planning of the study and the time spent in such a survey invariably is a wise investment.

The review of literature is a crucial step which invariably minimizes the risk of dead-ends, rejected topics, rejected studies, wasted effort, trial and error activity oriented towards approaches already discarded by previous investigators, and even more important erroneous findings based on a faculty research design. It also provides comparative data on the basis of which to evaluate and interpret the significance of one’s findings.

2.2 Data Mining and Machine Learning

Data mining is the process of extraction of useful patterns or knowledge from huge amount of data. According to Data mining is defined as the process of discovering patterns in data. Various alternative names for data mining is used like Knowledge mining from data, Knowledge Discovery from Data (KDD), knowledge extraction, data / pattern analysis, data archaeology, data dredging, information harvesting etc.

In the area of data mining, various studies were conducted over the years for feature selection, and classification, optimization of problems of medical diagnosis. In machine learning, supervised learning is a commonly used method for tackling
medical problems. The literature review presented in this section is divided into four parts:

(i) Data mining and machine learning
(ii) Feature Selection Techniques
(iii) Classification Techniques
(iv) Optimization Techniques

Detrano et al., (1989) discussed a new discriminant function model for estimating probabilities of angiographic coronary disease was tested for reliability and clinical utility in 3 patient test groups. This model, derived from the clinical and noninvasive test results of 303 patients undergoing angiography at the Cleveland Clinic in Cleveland, Ohio, was applied to a group of 425 patients undergoing angiography at the Hungarian Institute of Cardiology in Budapest, Hungary (disease prevalence 38%); 200 patients undergoing angiography at the Veterans Administration Medical Center in Long Beach, California (disease prevalence 75%); and 143 such patients from the University Hospitals in Zurich and Basel, Switzerland (disease prevalence 84%).

Knerr et al., (1990) Discussed simple SVM binary classifier to operate as a multi-class classifier using methods such as one-against-all, one-against-others, and directed acyclic graph.

Pawlak (1991) Propose almost every aspects of rough set theory, granulation, feature selection, rule construction has been explored.
Miller et al., (1992) reviewed the current applications of neural networks to medical imaging and signal processing. As is evident from the literature neural networks have already been used for a wide variety of tasks within medicine.

Krishnapuram et al., (1993) Discuss clustering algorithms, partitioning algorithms, clustering methods, possibility theory. It also discusses fuzzy set theory with membership update equations and it discuss possibilistic approach to fuzzy clustering, it has the undesirable tendency to produce coincidental clusters.

Wolberg et al., (1994) an interactive computer system evaluates and diagnoses based on cytologic features derived directly from a digital scan of Fine-Needle Aspirate (FNA) slides. A consecutive series of 569 patients provided the data to develop the system and an additional 54 consecutive, new patients provided samples to test the system. The projected prospective accuracy of the system estimated by tenfold cross validation.

Fayyad (1996) coined that the term Knowledge Discovery in Databases, or KDD for short, refers to the broad process of finding knowledge in data, and emphasizes the “high-level” application of particular data mining methods.

Clark Glymour et al., (1997) Discussed some statistical themes and lessons that are directly relevant to data mining and attempts to identify opportunities where close cooperation between the statistical and computational communities might reasonably provide synergy for further progress in data analysis.

Agrawal et al., (1998) state Data mining applications place special requirements on clustering algorithms including: the ability to find clusters embedded in subspaces of high dimensional data, scalability, end-user comprehensibility of the
results, non-presumption of any canonical data distribution, and insensitivity to the order of input records.

Koh et al., (2005) This article explores data mining applications in healthcare. In particular, it discusses data mining and its applications within healthcare in major areas such as the evaluation of treatment effectiveness, management of healthcare, customer relationship management, and the detection of fraud and abuse. It also gives an illustrative example of a healthcare data mining application involving the identification of risk factors associated with the onset of diabetes.

An excellent survey of SVMs can be found in, and a more recent book is by Cristianini et al., (2000). SVMs revolve around the notion of a “margin” either side of a hyperplane that separates two data classes. Maximizing the margin and thereby creating the largest possible distance between the separating hyperplane and the instances on either side of it has been proven to reduce an upper bound on the expected generalisation error.

Genton (2001) described several classes of kernels. Sequential Minimal Optimization (SMO) is a simple algorithm that can, relatively quickly, solve the SVM QP problem without any extra matrix storage and without using numerical QP optimization steps at all (Platt 1998). SMO decomposes the overall QP problem into QP sub-problems.

Keerthi et al., (2002) suggested two modified versions of SMO that are significantly faster than the original SMO in most situations. Finally, the training optimization problem of the SVM necessarily reaches a global minimum, and avoids ending in a local minimum, which may happen in other search algorithms such as
neural networks, Support Vector Machines (SVMs) is a supervised non-parametric statistical learning technique, therefore there is no assumption made on the underlying data distribution.

Valentini et al., (2003) described bagged ensembles of Support Vector Machines (SVM) and feature selection algorithms to the recognition of malignant tissues. Presented results show that bagged ensembles of SVMs are more reliable and achieve equal or better classification accuracy with respect to single SVMs, whereas feature selection methods can further enhance classification accuracy.

Ying Tan et al., (2004) This paper presents a mechanism to train support Vector Machines (SVMs) with a hybrid kernel and minimal Vapnik-Chervonenkis (VC) dimension. After describing the VC dimension of sets of separating hyperplanes in a high-dimensional feature space produced by a mapping related to kernels from the input space, we proposed an optimization criterion to design SVMs by minimizing the upper bound of the VC dimension. This method realizes a structural risk minimization and utilizes a flexible kernel function such that a superior generalization over test data can be obtained.

Humar and Novruz (2008) proposed a hybrid neural network that includes Artificial Neural Network (ANN) and Fuzzy Neural Network (FNN) to predict diabetes and heart diseases. The proposed model is able to support good accuracy values for Pima Indians diabetes data-set and Cleveland heart disease data-set, respectively.

Mingrui Zhang et al., (2008) investigated the performance of class imbalance in training data using neural network for medical diagnosis. Two algorithms have
been used standard Back-Propagation (BP) and Particle Swarm Optimization (PSO). The results show that BP outperforms the PSO for small and large data-sets.

Fabien Lauer et al., (2008) have proposed different formulations of the optimization problem along with Support Vector Machines (SVMs) for classification task. They have exposed the utility of concerns on the incorporation of prior knowledge into SVMs in their review of the literature. The methods are classified with respect to the categorization into three categories depending on the implementation approach via samples, in the kernel or in the problem formulation. They considered two main types of prior knowledge that can be included by these methods like class invariance and knowledge on the data.

Este et al., (2009) have introduced a new classification technique based on Support Vector Machines which is based on a flow representation that expresses the statistical properties of an application protocol. The classification mechanism presents a relatively high complexity during the training phase, especially due to the tuning process of the involved configuration parameters. They have applied the proposed technique to three different Datasets and almost in all cases, they found the accuracy of the classifier is very good with classification results (True Positives) going over the 90% mark and in general low False Positive rates.

Chih-Hung Wu et al., (2010) have proposed HGASVM which can help innovators or firms in identifying (classifying) and searching critical documents that can assist their strategic decision making process. The contribution of their study is that the proposed algorithm is an effective patent classification system that can ensure the continuous and systematic use of patent information in a company’s decision making processes.
John Taylor et al., (2011) have presented a review of optimization techniques used for training SVMs. They have shown how to instantiate the KKT conditions for SVMs. Along with the introduction of the SVM algorithms, the characterization of effective kernels has also been presented, which is helpful to understand the SVMs with nonlinear classifiers. For the optimization methodologies applied to SVMs, they have reviewed interior point algorithms, chunking and SMO, coordinate descent, active set methods and Newton’s method for solving the primal, stochastic subgradient with projection and cutting plane algorithms. They believe that the optimization techniques introduced in this paper can be applied to other SVM-related research as well.

Chin Heng Wan et al., (2012) have implemented a new text document classifier by integrating the K-Nearest Neighbor (KNN) classification approach with the Support Vector Machine (SVM) training algorithm. The proposed Nearest Neighbor-Support Vector Machine hybrid classification approach is coined as SVM-NN which avoids a major problem of the KNN in determining the appropriate value for parameter K in order to guarantee high classification effectiveness.

Metwally (2013) proposed a decision support tool for the detection of breast cancer based on three different classifiers, namely, Single Decision Tree (SDT), Boosted Decision Tree (BDT) and decision tree forest. They claimed that BDT performed better than SDT with 98.83% and 97.07% accuracy respectively. The demerit of decision tree classifier in medical diagnosis is imbalancement and cost sensitivity problem.

Ratnakar et al., (2013) proposed a solution based on genetic algorithm for selection of optimal set of attributes for prediction of heart diseases and Naïve Bayes
technique to generate relationships amongst the attributes using the concepts of conditional probability. Using GA, 13 attributes are reduced to 6 which are further fed to Naïve Bayes Classifier.

Chitra et al., (2013) claimed that the application of ANN can be time-consuming due to the selection of input features for the multi-layer perceptron. It is very slow training process and clinicians find it difficult to understand how its classification decisions are taken and cannot interpret the results easily.

Masethe (2014) performed a comparison of various data mining algorithms on WEKA tool for the prediction of heart attacks to find the best method of prediction. The algorithms used are J48, REPTREE, Naïve Bayes, Bayes net and CART with prediction accuracy.

Yuehjen et al., (2014) proposed a several hybrid models to predict heart disease such as Logistic Regression (LR), Multivariate Adaptive Regression Splines (MARS), Artificial Neural Network (ANN) and Rough Set (RS). The performance of the proposed models is better than artificial neural network.

Beheshti et al., (2014) proposed a new meta-heuristics approach named Centripetal Accelerated Particle Swarm Optimization (CAPSO). The proposed approach is used to enhance the performance of ANN learning and accuracy.

Archana et al., (2015) introduced a novel Hybrid Prediction Algorithm with Missing value Imputation (HPM-MI). The authors proposed a hybrid model based on K-means clustering with Multilayer Perceptron. The performance of the proposed algorithm is investigated on three benchmark medical Datasets namely Pima Indians Diabetes, Wisconsin Breast Cancer and Hepatitis from the UCI Repository of
Machine Learning. The results are very strong and the proposed model works fine when numbers of missing value are large in the data-set.

Bajaj et al., (2015) applied three data mining classification algorithm for heart disease such as decision tree, split validation and apply model. The proposed system is able to reduce medical mistakes and provide high accuracy prediction.

2.2.1 Feature Selection Techniques

Vapnik (1979) presented a set of labeled data instances and the SVM training algorithm aims to find a hyperplane that separates the dataset into a discrete predefined number of classes. Learning refers to the iterative process of finding a classifier with optimal decision boundary to separate the training patterns (in potentially high-dimensional space) and then to separate simulation data under the same configurations (dimensions).

Bezdek (1981) This book include mathematical theories underlying problems in cluster analysis, cluster validity, feature selection and classifier design, applications in medical diagnosis. And it introduced the basic idea of determining the fuzzy clusters by minimizing an appropriately defined functional and have derived iterative algorithms for computing the membership functions for the clusters.

Shortliffe (1987) It emphasizes in particular that decision-support programs are intended to serve as tools for trained practitioners who retain ultimate responsibility for determining diagnostic and therapeutic strategies that summarizes the current status of computer-based medical decision support.

Jain et al., (1988) the book emphasizes informal algorithms for clustering data and interpreting results. It deals as exploratory data analysis and is being applied in a
variety of engineering and scientific interdisciplines such as biology, psychology, medicine, marketing, computer vision and remote sensing.

Dubois et al., (1990) this paper discussed the notion of rough sets; fuzzy sets and it clearly declare the difference between fuzzy sets and rough sets.

Boser et al., (1992) this work describe a training algorithm that automatically tunes the capacity of the classification function by maximizing the margin between training examples and class boundary optionally after removing some typical or meaningless examples from the training data.

Burges (1998) describe Support Vector Machines (SVMs) are the newest supervised machine learning technique for pattern recognition.

Eisen et al.,(1998) presented a system of cluster analysis for genome-wide expression data from DNA microarray hybridization is described that uses standard statistical algorithms to arrange genes according to similarity in pattern of gene expression. The output is displayed graphically, conveying the clustering and the underlying expression data simultaneously in a form intuitive for biologists.

Duda et al., (1999) this book provides comprehensive coverage of both statistical classification theory and computer analysis, Bayesian decision theory, supervised and unsupervised learning, discriminant analysis and clustering.

Tavazoie et al., (1999) Technologies to measure whole-genome mRNA abundances and methods to organize and display such data are emerging as valuable tools for systems-level exploration of transcriptional regulatory networks.
Hierarchical clustering has also been used to organize genes into hierarchical dendograms on the basis of their expression across multiple growth conditions.

Veropoulos et al., (1999) state the maximum margin allows the SVM to select among multiple candidate hyperplanes, for many datasets, the SVM may not be able to find any separating hyperplane at all because the data contains misclassified instances.

Scholkopf et al., (1999) described kernels are a special class of function that allow inner products to be calculated directly in feature space, without performing the mapping. Once a hyperplane has been created, the kernel function is used to map new points into the feature space for classification.

Zhu et al., (2002) The selection of an appropriate kernel function is important, since the kernel function defines the feature space in which the training set instances will be classified. The problem can be addressed by using a soft margin that accepts some misclassifications of the training instances.

Dudoit et al., (2002) this article compares the performance of different discrimination methods for the classification of tumors based on gene expression data. The methods include nearest-neighbor classifiers, linear discriminant analysis, and classification trees. Recent machine learning approaches, such as bagging and boosting, are also considered. The discrimination methods are applied to datasets from three recently published cancer gene expression studies.

This work is presented by Miin-Shen et al., (2002) and in which they employed MRI segmentation techniques to differentiate abnormal and normal tissues in Ophthalmology using fuzzy clustering algorithms. Applying the best-known fuzzy
C-Means (FCM) clustering algorithm, a newly proposed algorithm, called an Alternative Fuzzy C-Means (AFCM), was used for MRI segmentation in Ophthalmology.


Huang et al., (2003) how do popular learning algorithms, such as decision trees and Naive Bayes, compare in terms of the better measure AUC. How does recent Support Vector Machine (SVM) compare to traditional learning algorithms such as Naive Bayes and decision trees and compare accuracy and AUC of Naive Bayes, C4.4 and C4.5 to the recently developed SVM on the datasets from the UCI repository.

Goodwin et al., (2003) Health care information systems tend to capture data for nursing tasks, and have little basis in nursing knowledge. Opportunity lies in an important issue where the knowledge used by expert nurses (nursing knowledge workers) in caring for patients is undervalued in the health care system.

Pawan Lingras et al., (2003), presented reviewed fuzzy and rough hybridization efforts. The review could be broadly categorized into logical and black box approaches. The logical approaches could be further subdivided into theoretical, supervised learning, feature selection, and unsupervised learning. The black box approaches consisted of neural and evolutionary computing.
Daxin Jiang et al., (2004) have discussed various types of clustering techniques such as gene based clustering, such as K-means, Hierarchical, Self organizing map and model based clustering, Sample based clustering.

Lingras et al., (2004) Propose a variation of the k means clustering algorithm based on properties of rough sets. It represents clusters as intervals or rough sets. It also describes the design of an experiment including data collection and the clustering process.

Kuhu et al., (2005) proposed a new model called Possibilistic-Fuzzy C-Means (PFCM) model. PFCM produces memberships and possibilities simultaneously, along with the usual point prototypes or cluster centers for each cluster. PFCM is a hybridization of Possibilistic C-Means (PCM) and Fuzzy C-Means (FCM) that often avoids various problems of PCM, FCM and FPCM. PFCM solves the noise sensitivity defect of FCM, overcomes the 44 coincident clusters problem of PCM and eliminates the row sum constraints of FPCM.

Dan Li et al., (2005), described an Algorithms Based on Fuzzy Set and Rough Set Theories, they presented missing data imputation methods based on clustering, one of the most popular techniques in KDD. They combined clustering with soft computing, which tended to be more tolerant of imprecision and uncertainty, and apply fuzzy and rough clustering algorithms to be dealt with incomplete data. The experiments showed that a hybridization of fuzzy set and rough set theories in missing data imputation algorithms led to the best performance among four algorithms, i.e., crisp K-Means, fuzzy K-Means, rough K-Means, and rough fuzzy K-means imputation algorithms.
Pal et al., (2005) Proposed the possibilistic fuzzy c means model. It produces memberships and possibilities simultaneously along with the usual point prototypes or cluster centers for each cluster. It is a strong candidate for fuzzy rule based system identification.

Mantero et al., (2005) SVMs are particularly appealing in the remote sensing field due to their ability to successfully handle small training data sets, often producing higher classification accuracy than the traditional methods.

Lifang Zhou et al., (2006) presented an attribute reduction method based on fuzzy rough set which was applied for the result obtained by PCA method and the recognition process used neural network ensemble. The method avoided losing of information caused by dispersing before rough set attribute reduction. It was mainly quotes attribute reduction of fuzzy rough sets to deal with the face data. The method escaped the dispersing of data so that it could reduce information loss. Moreover it was more effective than the face recognition based on rough set.

Krzysztof Michalak et al., (2006) reported the issue of high dimensionality of data is often considered important in classification problems. To lower data dimensionality, feature selection methods are often employed. To select a set of features that will span a representation space that is as good as possible for the classification task, one must take into consideration possible interdependencies among the features.

Masulli et al., (2006) It demonstrates possible novel applications of the possibilistic clustering approach by introducing a soft transition from the possibilistic to the probabilistic models.
Mitra et al., (2006) This study paper introduce a novel clustering architecture in which several subsets of patterns can be processed together with an objective of finding a common structure. It is developed by integrating the advantages of both fuzzy sets and rough sets.

In data mining, According to Han and Kamber (2006), Classification is the process of finding a model that describes and distinguishes data classes or concepts for the purpose of being able to use the model to predict the class of objects whose class label is not known. The model is based on the analysis of data objects whose class label is known.

Crone et al., (2006) have discussed the influence of different preprocessing techniques of attribute scaling, sampling, coding of categorical as well as coding of continuous attributes on the classifier performance of decision trees, neural networks and support vector machines. The impact of different preprocessing choices is assessed on a real world dataset from direct marketing using a multi-factorial analysis of variance on various performance metrics and method parameterizations.

Yanxiong Peng et al., (2007) In this work, evaluated different gene selection methods for biomarker discovery, including some traditional statistical methods and several newly developed methods aimed to obtain maximum relevance and minimum redundancy. The hybrid approach which combines filter and wrapper methods, in which we use the feature estimation from the filter step as the heuristic information for the wrapper step can reduce the effect of the over fitting problem and achieve the goal of maximum relevance with minimum redundancy. With these advantages, the hybrid approach may be a good candidate for biomarker discovery from microarray datasets.
Jensen et al., (2007) investigates a novel approach based on fuzzy-rough sets, fuzzy rough feature selection (FRFS), that addresses these problems and retains dataset semantics. FRFS is applied to two challenging domains where a feature reducing step is important; namely, web content classification and complex systems monitoring. The utility of this approach is demonstrated and is compared empirically with several dimensionality reducers.

Hu et al., (2007) Propose a simple and efficient hybrid attribute reduction algorithm based on a generalized fuzzy rough model. In this derive several attribute significance measures based on the proposed fuzzy rough model and construct a forward greedy algorithm for hybrid attribute reduction. This technique proven that the number of the selected features is the least but accuracy is the best.

Tsang et al., (2008) focused on the attributes reduction with fuzzy rough sets. After scrutinizing the preceding works on attributes reduction with fuzzy rough sets, the formal concepts of attributes reduction with fuzzy rough sets are introduced and completely study the structure of attributes reduction.

Sankar K. Pal (2008), described Data mining and knowledge discovery in databases, which had drawn the attention of researchers significantly, had been explained from the view-point of pattern recognition. The concept of rough-fuzzy computing was given more emphasis. Three examples of judicious integration, viz., rough-fuzzy case generation, rough-fuzzy C-means and rough-fuzzy C-medoids were explained along with their merits. Problems of rough-fuzzy clustering in protein sequence analysis and segmentation of brain MR images were considered. Fuzzy granulation through rough-fuzzy computing, Rough Fuzzy Knowledge Encoding and Uncertainty Analysis and performing operations on fuzzy granules provided both
information compression and gain in computation time; thereby making it suitable for data mining applications.

Chih-Cheng Hung et al., (2008), proposed a hybrid rough K-Means algorithm for image classification. The rough set theory was used to establish the lower and upper bound for data clustering in the K-Means algorithm. Then, the Particle Swarm Optimization (PSO) was employed to optimize the solutions of the rough K-Means algorithm. The combined algorithm was called the Rough K-Means PSO algorithm.

Jensen et al., (2010) proposed a novel hybrid approach for fuzzy-rough set rule induction. By performing feature selection and rule induction simultaneously, the generated rule sets were guaranteed to be compact and transparent. The experimental results showed that the method performed very well against a range of leading classifiers. The Quick rules induction algorithm did not employ any post-processing procedures to improve rule quality. It was likely that such optimizations would improve classification accuracy further.

Jensen et al., (2009) stated Fuzzy-rough set-based feature (FS) selection has been shown to be highly useful at reducing data dimensionality but possesses several problems that render it ineffective for large datasets. This paper proposes three new approaches to fuzzy-rough FS-based on fuzzy similarity relations. In particular, a fuzzy extension to crisp discernibility matrices is proposed and utilized. Initial experimentation shows that the methods greatly reduce dimensionality while preserving classification accuracy.

Yan Wang et al., (2009) Propose an efficient algorithm called as Feature Forest algorithm for generation of the reducts of a medical dataset. In the algorithm,
the given dataset is transformed into a forest to form discernibility string that is the concatenation of some of features and the disjunctive normal form is computed to reduct features based on feature forest.

Ye Yuling (2009) proposed Feature selection based on rough set theory could keep the system information and decrease the number of the attributes effectively. Sequential reduct algorithm bought some degree of tolerance to the noise which was effective to real online systems. In the process of sequential reduct, the most “significant” attributes of the current subsystem were selected to make up of the sequential reduct, which improved the SVM’s prediction accuracy.

Karabatak et al., (2009) this paper presents an automatic diagnosis system for detecting breast cancer based on Association Rules (AR) and Neural Network (NN). In this study, AR is used for reducing the dimension of breast cancer database and NN is used for intelligent classification. The proposed AR + NN system performance is compared with NN model. The dimension of input feature space is reduced from nine to four by using AR. In test stage, 3-fold cross validation method was applied to the Wisconsin breast cancer database to evaluate the proposed system performances.

Zhao et al., (2010) put forward a rule-based classifier which is constructed through generalization of Fuzzy Rough Set (FRS) called as Generalization of Fuzzy Rough Sets (GFRS) by adhering to a new notion called as consistence degree it is employed as the decisive value to retain the discernibility information similarly as it is in the procedure of rules induction.

Liu Wenjun et al., (2010), analyzed the factors of leading to the accident blackspots; then, using forward an algorithm to obtain the weight of each factor in the
accident black-spots combining rough set and fuzzy theory; Sorting the affected degree of each factor of accident black spots.

Jyoti Soni U.A (2011) conducted a research which intends to provide a survey of current techniques of knowledge discovery in databases using data mining techniques that are in use in today’s medical research particularly in Heart Disease Prediction and has concluded that the number experiments that have been conducted to compare the performance of predictive data mining technique on the same dataset and the outcome reveals that Decision Tree outperforms and sometime Bayesian classification is having similar accuracy as of decision tree but other predictive methods like KNN, Neural Networks, Classification based on clustering are not performing well.

Xu et al., (2011) Propose a fuzzy rough attribute reduction algorithm based on mutual information. This paper focus an approximate replacement of the mutual information, from both maximum relevance and maximum significance is raised. It improves the efficiency and decreases the complexity of the classical algorithm.

Pachgade et al., (2012) The new approached method provides efficient outlier detection and data clustering capabilities in the presence of outliers. The proposed method first finds out the user defined number of clusters with the mean of Euclidean plus Manhattan Distance then outlier are detected from each cluster. A Hybrid method can cluster the data according to user need and find the outliers that differ from the other data in the dataset.

Ankit et al., (2012) have presented an overview of data mining and its techniques which have been used to extract interesting patterns and to develop
significant relationships among variables stored in a huge dataset. It analyze the lung
cancer data available from the SEER program with the aim of developing accurate
survival prediction models for lung cancer. Several supervised classification methods
were used on the pre-processed data along with various data mining optimizations and
validations.

Fan et al., (2012) proposed the algorithm dynamically calculates heuristic
information based on the significance of feature to guide search. Experimental are
carried out on some standard UCI datasets. The results demonstrate that, in terms of
solution quality and computational effort, proposed algorithm can get better results
than other intelligent swarm algorithms for attribute reduction.

Noor Diana et al., (2013) stated that a classification technique, also known as a
classifier is a systematic approach in developing a classification model from a set of
input data. There are many classifiers such as Decision Tree (DT), Neural Networks
(NN), naive Bayesian, Support Vector Machine (SVM) and Rough Set Theory (RST).
Each classifier uses learning algorithms to discover the most appropriate model for
the relationship between an attribute set and class labels of input data.

Jenzi (2013) Conducted a study to develop a heart disease prediction system
by using data mining techniques with the aimed of identify useful patterns of
information from the medical data for quality decision making. The Decision Tree
and Naive Bayes as Data mining techniques were used on building his classifier
model used for predicting heart disease so that the result obtained from the classifier
enabled to establish significant patterns and relationships between the medical factors
relating to heart disease.
Rahman et al., (2013) used the various classification techniques like MLP, Bayes net, Naïve Bayesian, J48 graft, Fuzzy lattice reasoning, JRip, fuzzy inference system, adaptive neuro fuzzy inference system on three data mining tools TANAGRA, WEKA and MATLAB.

Yılmaz Kayaa (2013) offered a new hybrid medical decision support system based on Rough Set (RS) and Extreme Learning Machine (ELM) for the diagnosis of hepatitis disease. The RS-ELM consists of two stages. Initial stage, redundant features have been removed from the dataset through RS approach. Then, classification process has been implemented through ELM by using remaining features. The Hepatitis dataset taken from UCI machine learning repository has been used to test the hybrid model.

Sridevi et al., (2014) developed a feature selection algorithm Modified Correlation Rough Set Feature Selection (MCRSFS) that predicts both diagnosis and prognosis when compared with several data mining classification algorithms. In their approach, features are selected based on rough set with different starting values of reduction in stage one and in stage two features are selected from the reduced set based on the CFS.

Gaurav Taneja (2014) stated that a data mining is a technique that deals with the extraction of hidden predictive information from a large database. It uses sophisticated algorithms for the process of sorting through large amounts of Datasets and picking out relevant information. Data mining (the Analysis step of the Knowledge Discovery in Databases process, or KDD), a relatively young and interdisciplinary field of computer science, is the process of extracting Patterns from
large Datasets by combining methods from statistics and artificial intelligence with database management.

Boshra Bahrami (2015) carries a research on evaluating the different classification techniques in heart disease diagnosis. Classifiers like J48 Decision Tree, K Nearest Neighbors (KNN), Naive Bayes (NB), and SMO are used to classify dataset. After classification, some performance evaluation measures like accuracy, precision, sensitivity, specificity, F-measure and area under ROC curve are evaluated and compared.

2.2.2 Classification Techniques

Cristianini et al., (2000) presents comprehensive introduction to Support Vector Machines (SVMs), a new generation learning system based on recent advances in statistical learning theory. SVMs deliver state-of-the-art performance in real-world applications such as text categorization, hand-written character recognition, image classification, biosequences analysis.

Furey et al., (2000) discussed gene expression measurements, are being used to gather information from tissue and cell samples regarding gene expression differences that will be useful in diagnosing disease. It developed a new method to analyze this kind of data using Support Vector Machines (SVMs). This analysis consists of both classification of the tissue samples, and an exploration of the data for mis-labeled or questionable tissue results.

Burbidge et al., (2001) have shown that the Support Vector Machine (SVM) classification algorithm proves its potential for structure-activity relationship analysis. In a benchmark test, they compared SVM with various machine learning techniques currently used in this field. Among three tested artificial neural networks, they found that SVM is significantly better than all of these.

Giorgio Valentini (2002) have proposed classification methods, based on non-linear SVM with polynomial and Gaussian kernels, and Output Coding (OC), ensembles of learning machines to separate normal from malignant tissues, to classify different types of lymphoma and to analyze the role of sets of coordinately expressed genes in carcinogenic processes of lymphoid tissues. By using gene expression data from “Lymphochip” has shown that SVM can correctly separate the tumoural tissues, and OC ensembles can be successfully used to classify different types of lymphoma.

Huang et al., (2003) How do popular learning algorithms, such as decision trees and Naive Bayes, compare in terms of the better measure AUC How does recent Support Vector Machine (SVM) compare to traditional learning algorithms such as Naive Bayes and decision trees and compare accuracy and AUC of Naive Bayes, C4.4, and C4.5 to the recently developed SVM on the datasets from the UCI repository.

Zhan et al., (2005) the complexity of the trained SVM is further reduced by approximating the separation hyper surface with a subset of the support vectors. Compared to the initially trained SVM by all samples, the efficiency of the finally-trained SVM is highly improved, without system degradation.
A training method to increase the efficiency of SVM has been presented by Maolong et al., (2005) for fast classification without system degradation. Experimental results on real prostate ultrasound images show good performance of their training method in discriminating the prostate tissues from other tissues and they claim that their proposed training method is able to generate more efficient SVMs with better classification abilities.

Shen et al., (2005) modified the conventional SVM, Robust SVM and one-class SVM respectively based on the idea from Online SVM, and their performances are compared with that of the original algorithms.

Nurettinacir et al., (2006) This paper presents a novel system for automatic recognition of Auditory Brainstem Responses (ABR) to detect hearing threshold. ABR is an important potential signal for determining objective audiograms. Its detection is usually performed by medical experts with often basic signal processing techniques. The proposed system comprises of two stages. In the first stage, for feature extraction, a set of raw amplitude values, a set of Discrete Cosine Transform (DCT) coefficients and a set of Discrete Wavelet Transform (DWT) approximation coefficients are calculated and extracted from signals separately as three different sets of feature vectors. These features are then selected by a modified adaptive method, which mainly supports to the input dimension reduction via selecting the most significant feature components. In the second stage, the feature vectors are classified by a Support Vector Machine (SVM) classifier which is a powerful advanced technique for solving supervised binary classification problem due to its generalization ability.
Sumit Bhatia et al., (2008) presents a decision support system for heart disease classification based on Support Vector Machine (SVM) and integer-coded Genetic Algorithm (GA). Simple Support Vector Machine (SSVM) algorithm has been used to determine the support vectors in a fast, iterative manner. For selecting the important and relevant features and discarding the irrelevant and redundant ones, integer-coded genetic algorithm is used which also maximizes SVM’s classification accuracy. The Cleveland heart disease database is used in this study and it consists of 303 cases divided in 5 classes, each with 13 diagnostic features.

Mehmet Fatih Akay (2009) breast cancer diagnosis based on a SVM-based method combined with feature selection has been proposed. Experiments have been conducted on different training-test partitions of the Wisconsin Breast Cancer Dataset (WBCD), which is commonly used among researchers who use machine learning methods for breast cancer diagnosis. The performance of the method is evaluated using classification accuracy, sensitivity, specificity, positive and negative predictive values, Receiver Operating Characteristic (ROC) curves and confusion matrix.

Govindarajan et al., (2009) have proposed radial basis function classifier that performs comparative cross validation technique for existing radial basis function classifier of direct marketing data set, also he conclude the have proposed method classification accuracy and prediction accuracy are determined where the prediction accuracy is comparatively high.

Jian Ni et al., (2009) propose the gait recognition algorithm adopt support vector machine based on hybrid kernel function and parameter optimization. Partial kernel function and overall kernel function are fitted to compose super-kernel function, so that the SVM obtain better generalization ability and generalization
ability. In terms of parameter selection, the text uses the objective function and combine PSO algorithm to select the best kernel parameter.

Alireza Osareh et al., (2011) presented, a Computer-Aided Diagnosis (CAD) framework for breast cancer is developed using application of supervised machine learning techniques to the classification of cancerous / noncancerous data. Here, attempt to explore several different feature selection and extraction techniques and combine the optimal feature subsets with various learning classification methods such as K-nearest neighbors, probabilistic neural networks and support vector machines classifiers.

Hui Ling Chen et al., (2011) proposed a Rough Set (RS) based supporting vector machine classifier (RS_SVM) is for breast cancer diagnosis. In the proposed method (RS_SVM), RS reduction algorithm is employed as a feature selection tool to remove the redundant features and further improve the diagnostic accuracy by SVM. The effectiveness of the RS_SVM is examined on Wisconsin Breast Cancer Dataset (WBCD) using classification accuracy, sensitivity, specificity, confusion matrix and Receiver Operating Characteristic (ROC) curves.

Naiyang Deng et al., (2012) book provides a clear introduction to SVMs from an optimization perspective. It Constructs SVMs for semi-supervised, knowledge-based, and robust classification problems. Also discuss SVMs, privileged, multi-class, multi-instance, and multi-label classification problems.

Wu et al., (2012) this paper studies two kernel function of support vector machine global kernel function (linear kernel function) and local kernel function (RBF kernel function)and presents combination kernel function of support vector
Based on the analysis of different kernel functions characteristics, this paper choose a hybrid kernel function it is a combination kernel function, it has good generalization performance and learning ability, so improved the performance of SVM. Through the experiment results comparing, results show that its performance is better than that of other SVMs constructed by ordinary kernel function.

Govindarajan (2013) have proposed RBF - SVM hybrid system is evaluated by conducting several experiments on voting database. He compared various classifiers such as radial basis function and support vector machine. Finally he concludes the hybrid RBF-SVM shows higher percentage of classification accuracy than the base classifiers.

Govindarajan (2014) have discussed an ensemble of classification methods for data mining applications like intrusion detection, direct marketing and signature verification of KDD ‘99 data set. He compared have proposed classifiers (hybrid RBF-SVM approach) into base classifiers and he conclude that the hybrid RBF-SVM shows higher percentage of classification accuracy than the base classifiers and enhances the testing time due to data dimensions reduction.

Saumya et al., (2014) identified and evaluated the most commonly used data mining algorithms as well-performing in cancer survivability prediction based on recent studies and it is used the well performing techniques artificial neural networks, decision trees, genetic / fuzzy algorithms, support vector machine.

Dennis et al., (2014) A Genetic Fuzzy System (GFS) was essentially a fuzzy system supplemented by a learning process based on a Genetic Algorithm (GA). In
different application domains, Fuzzy systems were revealed their capability to work out diverse kinds of problems.

Sarah et al., (2016) proposed an unsupervised anomaly detection technique for high-dimensional large-scale unlabelled datasets. The technique is a combination of a DBN and one-class SVM. While one-class support vector machines are effective at producing decision surfaces from well-behaved feature vectors, they can be inefficient at modeling the variation in large, high-dimensional datasets. The derived features from the training samples are taken as input to train the one-class SVM. Subsequently, the generated hybrid model from the two algorithms is used for Machine learning systems for the text mining of cancer-related clinical reports have not been adequately investigated. Napolitano et al., (2016) proposed a few procedures for the pre-processing of free-text breast cancer pathology reports, with the point of encouraging the extraction of data pertinent to cancer arranging.

Molina et al., (2016) conveyed that the general target was to classify time arrangement in light of the revelation of frequent patterns. Those patterns would be found in typical successions got from the time arrangement information by a method for a transient deliberation prepare. Firstly, they changed numeric time arrangement into typical time groupings, where the images expected to characterize the applicable space ideas.

Olga Tosas et al., (2016) work went for anticipating the patient release result on every hospitalization day by presenting another worldview developing a classification of event data streams. Most classification algorithms certainly accepted the estimations of every single prescient element to be accessible at the season of making the prediction.
Yanchun Zhang et al., (2016) gave an overview of late advances in biomedical image examination and classification from new imaging modalities, for example, Terahertz (THz) Pulse Imaging (TPI) and Dynamic Contrast-Enhanced Magnetic Resonance Images (DCE-MRIs) and recognized their underlining shared characteristics.

2.2.3 Optimization Techniques

David W. Opitz (1999) presented the traditional motivation behind feature selection algorithms which is to find the best subset of features for a task using one particular learning algorithm. Given the recent success of ensembles, however, they investigate the notion of ensemble feature selection in this work. This task is harder than traditional feature selection in that one not only needs to find features germane to the learning task and learning algorithm, but one also needs to find a set of feature subsets that will promote disagreement among the ensemble's classifiers. In their work, they present ensemble feature selection approach that is based on genetic algorithms. Their algorithm shows improved performance over the popular and powerful ensemble approaches of AdaBoost and Bagging and demonstrates the utility of ensemble feature selection. The huge number of genes makes it impossible to execute an exhaustive search. In this work, propose a Recursive Reature Elimination (RFE) algorithm named FCM-SVM-RFE for the gene selection task. In each step, similar genes are grouped into clusters by the fuzzy C-means clustering algorithm, and then a Support Vector Machine (SVM) is modeled in each cluster-induced space, the genes which contribute large to the margin width of the SVM are selected to survive to the next step.
Mazurowski et al., (2008) investigates the effect of class imbalance in training data when developing neural network classifiers for computer-aided medical diagnosis. The investigation is performed in the presence of other characteristics that are typical among medical data, namely small training sample size, large number of features, and correlations between features. Two methods of neural network training are explored: classical Back Propagation (BP) and Particle Swarm Optimization (PSO) with clinically relevant training criteria.

Wu et al., (2010) the research results demonstrate that our proposed hybrid genetic algorithm approach can optimize all the parameters of the SVM for developing a patent classification system with a high accuracy. The proposed HGA-SVM model is able to dynamically and automatically classify patent documents by recording and learning the experts' knowledge and logic. Finally, propose a new decision-making process for improving the development of the SVM patent classification and searching system.

Ladha et al., (2011) stated that feature selection is an important topic in data mining, especially for high dimensional datasets. This work presents an empirical comparison of feature selection methods and their algorithms. In view of the substantial number of existing feature selection algorithms, the need arises to count on criteria that enable one to decide adequately which algorithm to use in certain situations. This work reviews several fundamental algorithms found in the literature and assesses their performance in a controlled scenario.

Dilip et al., (2011) applied back-propagation neural network to predict neonatal disease diagnosis. The authors applied standard back-propagation on
different data-sets of neonatal disease. The accuracy of the proposed model is 75% with higher stability.

Milan et al., (2011) investigated several data mining algorithms to solve cardiovascular disease such as: Artificial Neural Network (ANN), Support vector Machine (SVM), Decision Tree and RIPPER classifier. The accuracy of the ANN, SVM, Decision Tree and RIPPER are 80.06%, 84.12%, 79.05% and 81.08% respectively. SVM is able to predict the cardiovascular disease with higher accuracy.

Kumari et al., (2011) investigates the powerful of hybridizing two computational intelligence methods viz., Gray Wolf Optimization (GWO) and Artificial Neural Networks (ANN) for prediction of heart disease. Gray wolf optimization is a global search method while gradient-based back propagation method is a local search one.

Chowdhury et al., (2011) utilized ANN in predicting neonatal disease diagnosis. The proposed technique comprised of Multi-Layer Perceptron with a Back propagation learning algorithm for training ANN and recognizing a pattern for the diagnosing and prediction of neonatal diseases. The data set consists of 94 samples of different symptoms parameter. The technique exhibits ANN based prediction of neonatal disease with an accuracy of 75% with 64 training set and, 15 test set and 15 validation test.

Vanisree et al., (2011) proposed a Decision Support System for diagnosis of Congenital Heart Disease. The proposed system is based on Back propagation neural network which is multi layered Feed Forward Neural Network, which is trained by a supervised Delta Learning Rule. The dataset consists of 200 samples with 36
attributes each depicting signs, symptoms and the results of physical evaluation of a patient. The proposed system used 80% data set for training and 20% for testing and achieved an accuracy of 90% and mean square error of 0.016.

Yao Liu et al., (2011) Among data classification algorithms, Discrete Particle Swarm Optimization (DPSO), a technique based on standard PSO has proved to be competitive in predicting breast cancer, and in this paper, we implement a classifier using DPSO with new rule pruning procedure for detecting lung cancer and breast cancer, which are the most common cancer for men and women. Experiment shows the new pruning method further improves the classification accuracy, and the new approach is effective in making cancer prediction.

Rashmi Rekha Sahoo et al., (2012) described that the Rough Set Theory (RST) is a mathematical approach for feature selection which does not require any additional information about the data. It has the capability to extract the relevant features smoothly from a high dimensional data set having noise, imprecise, vague and inessential information. In their work, diverse high dimensional gene expression Datasets are used for feature selection. Using their proposed rough set algorithm, feature selection has been done and a comparative study has been done on the accuracies of the classifiers before and after reducing the features of the data.

Gouda I. Salama et al., (2012) compared diverse classifiers like decision tree (J48), Multi-Layer Perception (MLP), Naive Bayes (NB), Sequential Minimal Optimization (SMO), and Instance Based for K-Nearest neighbor (IBK) on three diverse databases of breast cancer (Wisconsin Breast Cancer (WBC), Wisconsin Diagnosis Breast Cancer (WDBC) and Wisconsin Prognosis Breast Cancer (WPBC))
by using classification accuracy and confusion matrix based on 10-fold cross validation method.

Hannah Inbarani et al., (2014) discussed that the medical datasets are often classified by a large number of disease measurements and a relatively small number of records of patients. All these measurements (features) are not important or irrelevant/noisy. These features may be especially harmful in the case of relatively small training sets, where this irrelevancy and redundancy are harder to evaluate. On the other hand, this extreme number of features carries the problem of memory usage in order to represent the dataset. In this study, new supervised feature selection methods based on hybridization of Particle Swarm Optimization (PSO), PSO based Relative Reduct (PSO-RR) and PSO based Quick 55 Reduct (PSO-QR) are presented for the diagnosis of diseases. The experimental result on several standard medical datasets proves the efficiency of the proposed technique as well as enhancements over the existing feature selection techniques.

Laura Emmanuella et al., (2014) reported that feature selection methods select a subset of attributes (features) of a dataset and it is done based on a defined measure, invalidating the inessential and irrelevant ones. When a feature selection method is applied in a dataset, their aim is to improve the quality of the dataset representation.

Mirjalili et al., (2014) The GWO algorithm mimics the leadership hierarchy and hunting mechanism of grey wolves in nature. Four types of grey wolves such as alpha, beta, delta, and omega are employed for simulating the leadership hierarchy. In addition, the three main steps of hunting, searching for prey, encircling prey, and attacking prey, are implemented.
Girish Chandrashekar et al., (2014) described that plenty of feature selection methods are available in literature due to the availability of data with hundreds of variables leading to data with very high dimension. The work provides an overview of some of the methods present in literature. The objective was to provide a generic introduction to variable elimination which can be applied to a wide array of machine learning problems.

Vanaja et al., (2014) presented their work to discuss various feature selection algorithms applied on different datasets to select the relevant features to classify data into binary and multiclass in order to improve the accuracy of the classifier. Recent researches in medical diagnosis use the different kind of classification algorithms to diagnose the disease.

Subbulakshmi et al., (2015) In this work a new hybrid algorithm that integrates the proposed Self-Regulated Particle Swarm Optimization (SRPSO) algorithm with the Extreme Learning Machine (ELM) for classification problems is presented. To optimize the input weights and hidden biases and minimum norm least-square scheme, an improved PSO is used to analytically determine the output weights. The PSO is enhanced by incorporating a mechanism which diversifies the search behavior of the particles so that the algorithm finds much better solutions.

Sulaiman et al., (2015) This paper presents the use of a new meta-heuristic technique namely Gray Wolf Optimizer (GWO) which is inspired from gray wolves’ leadership and hunting behaviors to solve Optimal Reactive Power Dispatch (ORPD) problem. ORPD problem is a well-known nonlinear optimization problem in power system. GWO is utilized to find the best combination of control variables such as generator voltages, tap changing transformers’ ratios as well as the amount of reactive
compensation devices so that the loss and voltage deviation minimizations can be achieved to show the effectiveness of GWO technique compared to other techniques available in literature.

Mohamed et al., (2015) propose a multilayer neural network model tuned by Grey Wolf Optimization algorithm (GWO) is investigated and presented. GWO search algorithm is used to optimized all the connection of weights and biases for the artificial neural network.

Kaveh et al., (2016) In this study, the recently developed algorithm called GWO is used for optimum design of castellated beams. Three castellated beams are selected from literature to design by this method. GWO is a potential alternative optimization algorithm to solve castellated beam problems. It is observed that the optimization results obtained by the GWO algorithm have little difference with other methods for more design examples. The great advantages of the GWO are that the algorithm is simple, flexible, and easy to implement.

SelvaBhuvaneswari et al., (2017) The entire segmentation process of our proposed work comprises three phases: threshold generation with dynamic modified region growing phase, texture feature generation phase and region merging phase. by dynamically changing two thresholds in the modified region growing approach, the first phase of the given input image can be performed as dynamic modified region growing process, in which the optimization algorithm, firefly algorithm help to optimize the two thresholds in modified region growing. After identifying the abnormal tissues, the classification can be done by hybrid kernel-based SVM (Support Vector Machine). The performance analysis of the proposed method will be carried by K-cross fold validation method.
Li et al., (2017) An improved GWO, IGWO, was first proposed to identify the most discriminative features for major prediction. In the proposed approach, Particle Swarm Optimization (PSO) was firstly adopted to generate the diversified initial positions, and then GWO was used to update the current positions of population in the discrete searching space, thus getting the optimal feature subset for the better classification purpose based on SVM.