Performance Analysis of Particle Swarm Optimization for Feature Selection

*Amos O. Bajeh, Bukola O. Funso and Fatima E. Usman-Hamza

Department of Computer Science, University of Ilorin, Ilorin, Nigeria
{bajehamos | usman-hamza.fe}@unilorin.edu.ng | ajengbeomolara@gmail.com

Abstract— One of the key task in data mining is the selection of relevant features from datasets with high dimensionality. This is expected to reduce the time and space complexity, and consequently improve the performance of data mining algorithms for tasks such as classification. This study presents an empirical study of the effect of particle swarm optimization as a feature selection technique on the performance of classification algorithms. Two dataset from different domains were used: SMS spam detection and sentiment analysis datasets. Particle swarm optimization is applied on the datasets for feature selection. Both the reduced and raw dataset are separately classified using C4.5 decision tree, k-nearest neighbour and support vector machine. The result of the analysis showed that the improvement of classifier performance is case-dependent; some significant improvements are noticed in the sentiment analysis datasets and not in the SMS spam dataset. Although some marginal effect are observed on performance, it implies that with particle swarm optimization features selection the space complexity is reduced while maintaining the accuracy of the classifiers.

Keywords—classification, feature selection, machine learning, particle swarm optimization, text mining

1 Introduction

Text Classification (TC) is a means of knowledge engineering by which expert knowledge on classifying a document is automated so that documents can be classified into their individual suitable categories according to pre-defined class distinctions (Dang & Ahmad, 2014; Singh, 2016; Zelaia, Alegria, Aregi & Sierra, 2011). TC is a type of supervised machine learning in which algorithms learn from examples to carry out new classification. There are two dimensions of TC task, one is to classify documents to only a single category while the other is to classify document into more than one category (Korde & Mahender, 2012; Sharma, 2017; Bajeh, Alabidun & Sadiku, 2018). One of the tasks involved in TC is feature selection for reducing high dimension datasets (Jindal, 2015).

Text data usually comes with very high dimensions containing both relevant and irrelevant data that can be degrading to text classification tasks (Aurangzeb, Baharum, Lam Hong & Khairullah , 2010; Jindal 2015). Thus, the need to reduce the dimensions of such data before classification is necessary to reduce computational time and space complexity without reducing accuracy of classification algorithms (Sutha & Tamilselvi, 2015). Particle Swarm Optimization (PSO) among all other known traditional feature selection algorithms such as the filter-based, wrapper-based and embedded approaches, is an evolutionary algorithm for both optimization and feature selection and does not converge into a local optimum but global optimum. (Vashishtha, 2016). It is a meta-heuristic optimization algorithm that is more particular about the selection of most relevant attributes in a large search space for classification purpose.

This study applied PSO algorithm for feature selection of relevant attributes from textual dataset from two different domains with a view to determine its impact on some classification algorithms. After the reduction of the dimensions of the data using PSO, three classification algorithms were used to categorize both the reduced and the original text data and their performance were evaluated for comparison.

*Corresponding Author

The remaining part of this paper is organised as follows. Section 2 presents the concept of classification algorithms and feature selection; the classification algorithms and feature selection methods considered in this study are discussed. Section 3 discusses studies on the use of PSO and its variants for dimensionality reduction and text classification. Section 4 presents the research methodology including the dataset and the performance evaluation measures employed in this study. The experimental results showing the performances of the three classifiers with and without the use of PSO for feature selection are presented and discussed in section 5 and section 6 concludes the paper.

2 Classification Algorithms

Classification is one of the major problem areas addressed by data mining and machine learning techniques and it basically involves developing models that can be used to place new items into one of predefined classes depending on the features of the new item (Gareth, Daniela, Trevor. & Robert, 2013; Bajeh et al., 2018; Gorade & Deo, 2017; Kaur & Grewal, 2016; Ashari, Paryudi & Tjoa, 2013). This section discusses the concept of feature selection, PSO and the three classification algorithms (C4.5 decision tree, k-nearest neighbour and support vector machine) considered in this stud. The three algorithms are considered for this study because of the results of Ashari, Paryudi, and Tjoa (2013) and Jodas, Marranghello, Pereira, and Guido (2013) in which these algorithms showed good performances.

2.1 Decision Tree

A decision tree is an inverted tree-shaped structure in which each branch of the tree leads to a decision based on the features of the item being classified and the internal nodes (aside the leaf nodes) are points where decisions are determined. Decision tree instance classification is done by sorting instances from the root node to some leaf node. Each node in the tree specifies the test of some attribute of the instance, and each branch descending from that node corresponds to one of the possible values for this attribute (Banu, 2017; Singh, Leavline, Valliyappan, & Srinivasan., 2015). Several decision tree algorithm have been developed, this study uses the C4.5 decision three.

2.2 K-Nearest Neighbour

K-NN is an instance based classifier which classifies an instance based on the category of its neighbour. K-NN classifies a new instance into the category that the majority of the k neighbours of the instance. It uses distance function to determine the k nearest neighbours items in the trained model and the new/test case being classified (Imandoust & Bolandraftar, 2013; Lavanya, & Divya, 2017; Akintola, Balogun, Lafenwa-Balogun & Hameed, 2018; Bajeh et al., 2018).

2.3 Support Vector Machine

This is a supervised machine learning method that sort data into categories using decision surface called hyper plane. SVM requires a positive and negative training set to find the hyper plane decision surface which best separates the positive from the negative data in n-dimensional. Support vectors are the data instances closest to the decision surface (Aurangzeb et al., 2010; Bajeh et al., 2018; Fathima & Manimeglai, 2012; Lavanya & Divya, 2017; Gareth et al., 2013). The performance of SVM classification is not affected by the removal from the training dataset, instances that does not belong to the support vectors (Heide, Gerhard, & Marc-andré, 2002). The hyper plane is located at the maximum distance margin between the plane and the support vectors of two close categories being differentiated by SVM. This classification method is robust in the case of high dimensional data.

2.4 Feature Selection

Classification problems usually involve high-dimensional datasets consisting of thousands of instances (records) and each of which may be represented by several descriptive features also referred to as attributes or variables. Feature selection is applied to reduce data dimensionality by selecting the most relevant set of features before employing data mining techniques such as classification (Sutha & Tamilselvi, 2015). There are various types of feature selection methods such as information gain which uses entropy for selecting most relevant features (Abu & El-Henawy, 2017), relief F (Novaković, Strbac & Bulatović, 2011; Durgabai, 2014), gain ratio and chi-square. Predominantly, the correlation-based techniques are used for feature selection. Correlation–based approach identifies the association between sample data features and the response variable for classification (Doshi & Chaturvedi, 2014). This study considers a meta-heuristic based algorithm, the particle swarm optimization (PSO), as a feature selection technique.

2.5 Particle Swarm

PSO is a meta-heuristic population-based globalized search algorithm (Saini, Rohaya, Awang, Zakaria & Sulaiman, 2014; Durga, Lalitha & Application, 2015) inspired by the behaviour of flock of birds and school of fishes. It is an optimization technique which discovers a solution through several iterations. It works by a population consisting of a set of particles. Each particle is associated with a position and a velocity. The particles position and velocity are updated using simple mathematical equations, (1) and (2) respectively. It contains 2 best positions known as local best and global best (Singh et al., 2015).

$$x_{i,j}(t+1) = x_{i,j}(t) + v_{i,j}(t+1) \quad (1)$$

$$v_{i,j}(t+1) = wv_{i,j}(t) + r_1c_1[p_{i,j}(t) - x_{i,j}(t)] + r_2c_2[g(t) - x_{i,j}(t)] \quad (2)$$

where $x_{i,j}(t)$ is the position of a particle (candidate solution) at time t, and $v_{i,j}(t)$ is its velocity at time t. c_1 and c_2 are the acceleration coefficients and w is the inertia coefficient. r_1 and r_2 are real value coefficients that ranges between 0 and 1. Equations (1) and (2) update the position and velocity of the particles as they move towards the global optimum. The new position is computed by adding the current position and the new velocity of each particle (equation (1)). The new velocity is determined as the summation of the current velocity $v_{i,j}$; the vector between the current position (x_{ij}) and its personal best position ($P_{i,j}$) (i.e. $P_{i,j}$ - x_{ij}); and the vector between the current position and the global best g(t) i.e. g(t)-x_{ij}(t). These vectors move each particle to a better position closer to the global optimum.

3 Related Works

Several studies using PSO for feature selection have been reported in the literature. This section presents a review of some of these studies. Al-ab and Al-taani (2017) in their study compared the use of a meta-heuristic algorithm (PSO) which combines informative scoring with semantic scoring to create a shorter version of an original text in Arabic document with two important evolutionary methods Genetic Algorithm (GA) and Harmony Search (HS). The results proved that PSO algorithm achieved a higher precision and F-score measure than the GA and HS approaches since it requires only 100 iterations to converge to a nearest best position whereas HS needs 100000 iterations. Devi, Rao, and Setty (2016) presented a study in which PSO and Genetic algorithm (GA) were used to select important features on a dengue dataset consisting of 18 attributes collected from 1275 patients. Decision tree was used as the classifier for the two feature selection algorithms. It was reported that the PSO combined with decision tree (DT) classifier gave a better classification result when compared to GA combined with decision tree classifier. The result also shows a decrease in the error rate of PSO+DT compared to GA+DT.

Vashishtha (2016) survey literatures for the accuracy of feature selection using traditional wrapper and filter based method compared with different variant of PSO on Vehicle and Sonar datasets with 18 features, 846 instances and 60 features, 208 instances respectively. It was reported that feature selection based on PSO gives better accuracy as compared to traditional approaches. Muthusamy, Polat, and Yaacob (2015) used two PSO based method to select features and enhance data so as to improve the recognition of emotion in speech and glottal signals. PSO based clustering and wrapper based PSO were the two algorithms proposed and when applied to three emotional speech databases, they reported that it was able to classify better when compared to previously published results.

Tu, Chuang, Chang and Yang (2007) used PSO to optimize 5 publicly available datasets and used SVM as their classifier, and in all cases PSO with SVM had better classification results than applying SVM only. Most of the studies use single dataset, only few used more than one dataset (Tu et al., 2007). Thus, there is no sufficient triangulation to extend the validity of their results, making their conclusion to be limited to the individual datasets used. This study uses two datasets from different domain to investigate the commonality of the result of using PSO for feature selection. This triangulation is to observe if the results will be consistent across different domain.

Also, several studies have used PSO for optimization and clustering: Cui, Beaver, Charles and Potok (2008) investigated the use of various dimensionality reduction algorithms with PSO as a clustering algorithm; Saini et al. (2014) presented a literature survey on the use of PSO and its variants in addressing the problem of human motion tracking; Sarkar & Roy, A. (2013) and Dahiya & Singh (2014) studied the application of PSO as a clustering algorithm in text mining; Pradesh (2013) presented an algorithm that can combine weighted principal component analysis as the dimensionality reduction process and PSO for clustering.

4 METHODOLOGY

Two text datasets sourced from University of California Repository (http://kdd.ics.uci.edu) were used in this study. The two text datasets are from different domains. The first dataset is the Short Messaging Service (SMS) spam collection which consist of 5,574 messages each having a binary classification of either harmless or spam; it has 4827 harmless instances and 747 spam instances. The second dataset is a sentiment analysis dataset which consist of product reviews from three websites: Amazon, yelp and imdb. The collected dataset consist of 3000 instances with each website contributing a thousand instances evenly partitioned into five hundred positive and five hundred negative reviews respectively.

The classification algorithms applied are the C4.5 decision tree, k-NN and SVM. The results of the three algorithms on the reduced dataset (using PSO) were compared with the results when they are applied on the original dataset that is not reduced. The experiments were performed using the Waikato Environment for Knowledge Analysis (WEKA) data mining tool. Figure 1 depicts the framework used in this study.

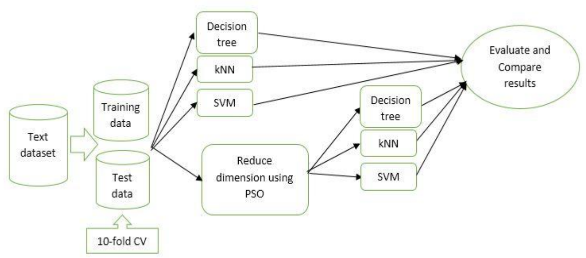

Fig. 1: Study Framework

The experiment involved the following: The PSO module for feature selection was executed for a period of 20 iterations to discover the vicinity of the optimal solution for a global search, and to minimize resource consumption. The result from the PSO module is then used as the input to the individual classifiers (C4.5, kNN and SVM) for the classification task. The population involves 20 particles with individual weight of 0.34 and an inertia weight (w) set as 0.33.

The datasets were tokenized and converted to "arff" format on the WEKA tool. The datasets were normalized and PSO was applied to reduce the dimension of the normalized text data. The training dataset consists of the 70% of each dataset to build the classifier models while the remaining 30% for evaluating the performance of the classifiers. 10-fold cross validation is used in the training of the classifiers in which the training dataset is randomly partitioned into 10 groups; the first 9 groups are used for training the classifier and the other set of data for testing the model correctness.

The classification evaluation parameters used are accuracy, recall and precision which are metrics derived from confusion matrix values: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Accuracy is the percentage of correctly classified instances. Precision measures the proportion of actual positives that are correctly classified within the predicted set. Recall measures the proportion of actual positives that are correctly classified within the actual set.

5 RESULTS AND DISCUSSION

The performance evaluation results of the classifiers with and without the use of PSO for feature selection are depicted in Figures 2 – 7 and compared in Table 1. The figures present the values of accuracy, precision and recall of the individual classifiers on each of the two datasets used.

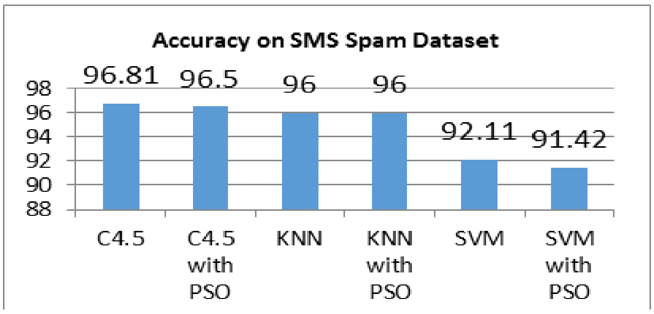

Fig. 2: Accuracy of classifiers on the SMS SPAM Dataset.

Figure 2 shows a very marginal decrease in the performance of C4.5 from 96.81% to 96.5%, and SVM from 92.11% to 91.42%. No effect on the accuracy of k-NN.

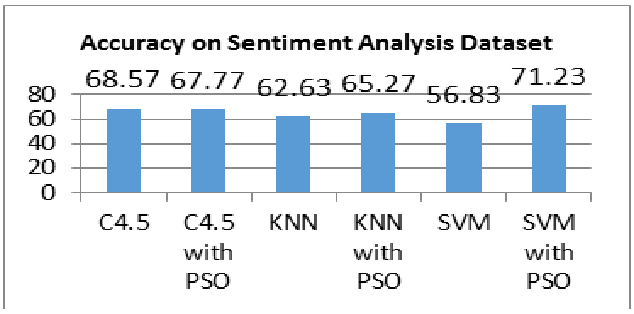

Fig. 3: Accuracy of classifiers on Sentiment Dataset

Figure 3 shows that for the sentiment dataset, the performances of C4.5 insignificantly reduced from 68.57 to 67.77 while that of K-NN and SVM insignificantly increased from 62.63 to 65.27 and significantly from 56.83 to 71.23 respectively.

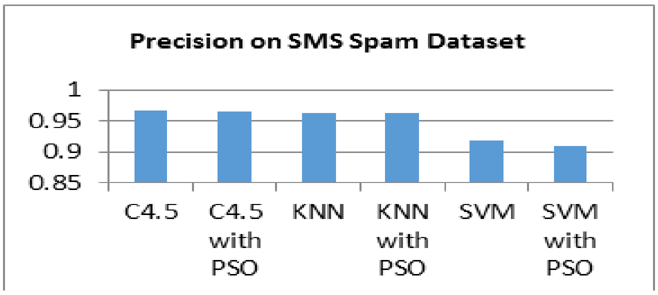

Fig. 4: Precision of classifiers on SMS Spam Dataset.

Figure 4 shows that the application of PSO for feature selection has no significant effect on the precision value of classifiers.

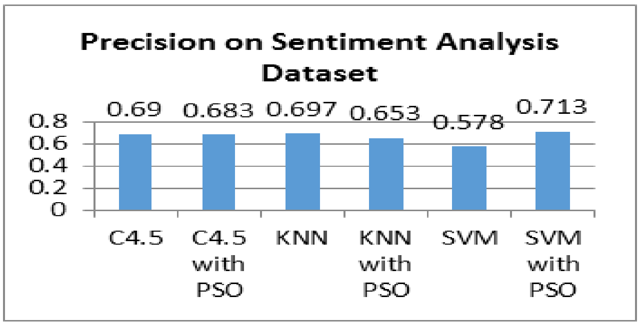

Fig. 5: Precision of classifiers on Sentiment Dataset

Figure 5 depicts a significant improvement in the precision of only SVM from 57.8% to 71.3%.

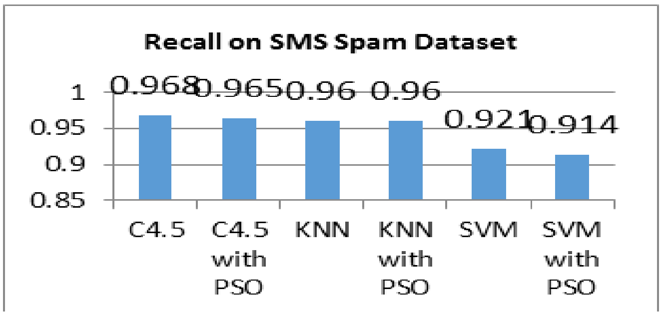

Fig.6: Recall of classifiers on SMS Spam Dataset

Figure 6 shows that the application of PSO for feature selection has no significant improvement on the recall value of the classifiers.

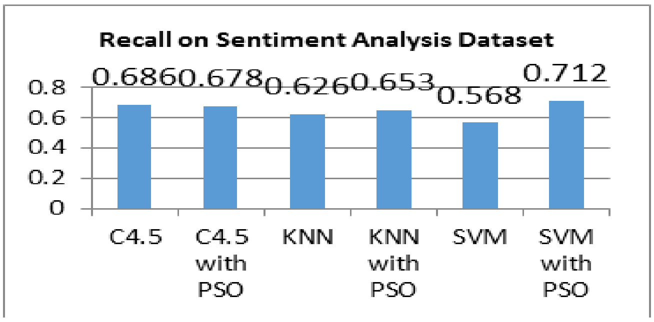

Fig. 7: Recall of classifiers on Sentiment Dataset.

From Figure 7, another significant improvement is seen on the recall of SVM when PSO is applied.

Table 1 presents a summary of the performance of the classifier depicted in Figures 2 – 7; it shows the difference in the performance of the classifiers with and without PSO feature selection. The table shows that feature selection does not significantly improve performance at all times as commonly believed. We also observed that performance is case-based i.e., every dataset has different performance levels. The classifiers had no performance improvement on the SMS spam dataset while some improvements were noticed in some instances of the classifiers on the sentiment analysis dataset. Thus, although the application of PSO for feature selection did not improve the predictive accuracy of the classifiers for the SMS spam dataset, it uses a reduced dataset to achieve almost the same level of performance with no significant difference; there was a very marginal and insignificant increase or decrease in the performances of the classifiers. This implies that while reducing the space complexity by reducing the number of features, the performance of the classifiers is not affected negatively; the feature selection process is advantageous by reducing space complexity.

6 Conclusion

This study investigated the impact of Particle Swarm Optimization as feature selection technique on the performance of classifiers. Three classification algorithms: C4.5 decision tree, k-NN and SVM were applied on both original and reduced text data from two different domains. The evaluation of the impact of PSO in this study is based on the precision, recall, accuracy and F-measure of the classifiers. The study showed that classification performances are not improved at all time when PSO feature selection is applied on datasets. Also, performance is case-based with the different dataset showing different performances.

All the classifiers, except k-NN, showed very marginal and insignificant decrease in performances when PSO was used on the SMS spam dataset while k-NN neither improved nor reduced performance which implies that PSO has no effect on the performance of this classifier. Some performance improvements are observed in the use of PSO on the sentiment analysis dataset. When PSO is applied, SVM showed a consistent and significant performance improvement on the sentiment analysis dataset.

Table 1. Difference in Performance of the Classifiers with and without PSO Feature Selection

Performance Metric	Classifier	Dataset SMS Spam	Sentiment Analysis
Accuracy	C4.5	-0.3	-0.8
	K-NN	0.00	+2.64
	SVM	-0.5	+14.4
Precision	C4.5	-0.3	-0.7
	K-NN	0.00	-4.4
	SVM	-1.0	+13.5
Recall	C4.5	0.00	-0.01
	K-NN	0.00	+0.03
	SVM	-0.07	+14.4

The result of this study does not agree with that of Devi et al. (2016) in which C4.5 performs well with PSO. These results are in tandem with that of Muthusamy et al. (2015); Saini et al. (2014) and Xue, Zhang & Browne (2012). PSO and its variant as used by these researchers with or without other feature selection techniques on different datasets such as emotion speech recognition, human motion tracking shows outstanding results. This study showed that although PSO can improve performance of classifiers, it is not always so thus, the use of PSO for feature selection must be done with caution to ensure that performance is actually improved.

Further studies will focus on using more dataset from differs domains to observe the impact of PSO feature selection on classifiers to yield results with better external validity. Also, variants of PSO, other meta-heuristic optimization and feature selection algorithms such as ant colony and bee colony will be studied. More classifiers such as Naïve Bayes, Logistic Regression and ensemble methods with PSO for feature selection will also be investigated. The effect of the type of dataset on feature selection and classifier performance is worth investigating since this study showed that performance is case-dependent.

REFERENCES

Abu, M., & El-Henawy, I. (2017). A Feature Selection Algorithm based on Mutual Information using Local Non-uniformity Correction Estimator. *International Journal of Advanced Computer Science and Applications, 8*(6), 418–423.

Akintola, A.G., Balogun, A.O., Lafenwa-Balogun, F.B. & Hameed A.M.(2018). Comparative Analysis of Selected Heterogeneous Classifiers for Software Defects Prediction Using Filter-Based Feature Selection Methods. *FUOYE Journal of Engineering and Technology, 3* (1), 133-137.

Al-ab, R. Z., & Al-taani, A. T. (2017). Single Document Text Summarization Using Particle Swarm Optimization Algorithm. *Procedia Computer Science, 117*, 30–37. http://doi.org/10.1016/j.procs.2017.10.091

Ashari, A., Paryudi, I., & Tjoa, M. A. (2013). Performance Comparison between Naïve Bayes , Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool. *International Journal of Advanced Computer Science and Applications, 4*(11), 33–39.

Aurangzeb Khan, Baharum Baharudin, Lam Hong Lee, Khairullah Khan (2010). A Review of Machine Learning Algorithms for Text-Documents Classification. *Journal of Advances in Information Technology 1*(1): 4–20.

Bajeh, A.O., Alabidun, M.O., Sadiku, P.O. (2018). Performance Evaluation of some Select Machine Learning Algorithms for Sentiment Analysis. *Int'l Journal of Information Processing and Communication, 6*(2), 409-419.

Banu, G. R. (2017). A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease. *International Journal of Computer Sciences and Engineering, 4*(11), 111–115.

Cui, X., Beaver, J. M., Charles, J. S., & Potok, T. E. (2008). Dimensionality Reduction Particle Swarm Algorithm for High Dimensional Clustering. *2008 IEEE Swarm Intelligence Symposium* DOI: 10.1109/SIS.2008.4668309

Dahiya, R., & Singh, A. (2014). A Survey on Application of Particle Swarm Optimization in Text Mining. *International Journal of Innovative Research and Development 3*(5), 101–107.

Dang, S., & Ahmad, P. H. (2014). Text Mining: Techniques and its Application. *International Journal of Engineering & Technology Innovations, 1*(4), 866–2348.

Devi, B. R., Rao, K. N., & Setty, S. P. (2016). Towards Better classification using Imporved Particle Swarm Optimization and Decision tree for Dengue Datasets. *International Journal of Soft Computing, 11*(1), 18–25.

Doshi, M. &, & Chaturvedi, S. K. (2014). Correlation Based Feature Selection (CFS) Technique to Predict Student Performance. *International Journal of Computer Network & Communicxation, 6*(3), 197–206.

Durga, M. L., Lalitha, P., & Application, C. (2015). Data Mining By Adopting Particle Swarm Optimization. *International Journal for Research in Science Engineering and Technology, 2*(12), 32–36.

Durgabai, R. P. L. (2014). Feature Selection using ReliefF Algorithm. *International Journal of Advanced Research in Computer and Communication Engineering, 3*(10), 8215–8218.

Fathima, A. S., & Manimeglai, D. (2012). Predictive Analysis for the Arbovirus-Dengue using SVM Classification. *International Journal of Engineering and Technology, 2*(3), 521–527.

Gareth J., Daniela W., Trevor H. & Robert Tibshirani (2013). *An Introduction to Statistical Learning with Applications in R.* Springer Text in Statistics, Springer Science+Business Media, New York USA

Gorade, S. M., & Deo, P. A. (2017). A Study of Some Data Mining Classification Techniques. *International Research Journal of Engineering and Technology, 4*(4), 3112–3115.

Heide, B., Gerhard, K., & Marc-andré, M. (2002). Document Classification Methods for Organizing Explicit Knowledge. In *Proceedings of the Third European Conference on Organizational Knowledge, Learning and Capabilities*, pp. 1–26, Athens, Greece: University of Bern.

Imandoust, S. B., & Bolandraftar, M. (2013). Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events : Theoretical Background. *Int. Journal of Engineering Research and Applications, 3*(5), 605–610.

Jindal, R. (2015). Techniques for text classification : Literature review and current trends. *Webology, 12*(2), 1–28.

Jodas, D. S., Marranghello, N., Pereira, A. S., & Guido, R. C. (2013). Comparing support vector machines and artificial neural networks in the recognition of steering angle for driving of mobile robots through paths in plantations. *Procedia Computer Science, 18*, 240–249.

Kaur, S., & Grewal, A. K. (2016). A Review Paper on Data Mining Classification Techniques For Detection of Lung Cancer. *International Research Journal of Engineering and Technology, 3*(11), 1334–1338.

Korde, V., & Mahender, C. N. (2012). Text Classification and Classifiers: A Survey. *International Journal of Artificial Intelligence & Applications, 3*(2), 85–99.

Lavanya, B., & Divya, B. (2017). Big Data Analysis Using SVM AND K-NN Data Mining Techniques. *International Journal of Computer Science and Mobile Computing, 6*(1), 84–91.

Muthusamy, H., Polat, K., & Yaacob, S. (2015). Particle Swarm Optimization Based Feature Enhancement and Feature Selection for Improved Emotion Recognition in Speech and Glottal Signals. *PloS one, 10*(3), e0120344.

Novaković, J., Strbac, P., & Bulatović, D. (2011). Toward optimal feature selection using ranking methods and classification algorithms. *Yugoslav Journal of Operations Research, 21*(1), 119–135.

Pradesh, A. (2013). Efficient clustering of Dataset based on Particle Swarm optimization. *International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR), 3*(1), 2249-6831.

Saini, S., Rohaya, D., Awang, B., Zakaria, M. N. B., & Sulaiman, S. B. (2014). A Review on Particle Swarm Optimization Algorithm and Its Variants to Human Motion Tracking, 2014, 13–15.

Sarkar, S., & Roy, A. (2013). Application of Particle Swarm Optimization in Data Clustering : A Survey. *International Journal of Computer Applications, 65*(25), 38–46.

Sharma, M. (2017). Literature Review and challenges of Data Mining techniques for Social Network Analysis. *Advances in Computational Science and Technology, 10*(5), pp. 1337-1354.

Singh, D. A., Leavline, E. J., Valliyappan, K., & Srinivasan, M. (2015). Enhancing the Performance of Classifier Using Particle Swarm Optimization (PSO) - based Dimensionality Reduction. *International Journal of Energy, Information and Communications, 6*(5), 19–26.

Singh, T. (2016). A Comprehensive Review of Text Mining. *Internation Journal of Computer Science and information Technology7*(1), 167–169.

Sutha, K., & Tamilselvi, J. J. (2015). A Review of Feature Selection Algorithms for Data Mining Techniques. *International Journal on Computer Science and Engineering (IJCSE), 7*(6), 63–67.

Vashishtha, J. (2016). Particle Swarm Optimization based Feature Selection. *International Journal of Computing Applications, 146(*6), 11–17.

Xue, B., Zhang, M., & Browne, W. N. (2012). Particle Swarm Optimization for Feature Selection in Classification : A Multi-Objective Approach. *IEEE Transactions on Cybernetics, 43*(6), 1656–1671.

Zelaia, A., Alegria, I., Arregi, O., & Sierra, B. (2011). A multiclass/multilabel document categorization system: Combining multiple classifiers in a reduced dimension. *Applied Soft Computing Journal*, 1–10.