

Literature Review on Feature Selection in Artificial Intelligence Applications

¹Erick Odhiambo Omuya

¹School of Engineering and Technology, Department of Computing and IT, Machakos University, Kenya

ABSTRACT

Feature selection aims at removing repetitive and irrelevant feature sets to enable more accurate classification of new instances. It has fundamentally changed the performance of algorithms used in artificial intelligence through machine learning, sentiment analysis and data mining. Classification, clustering and prediction without proper feature selection yields poor performance with machine learning algorithms. **This literature review explores the impact of feature selection in artificial intelligence applications. It is a qualitative study that examines research on feature selection in artificial intelligence through articles from peer reviewed scientific journals, conference proceedings and reports.** Online resources including but not limited to the semantic scholar were used where searches were guided by relevant keywords. **This paper uses content analysis to examine literature to establish the feature selection techniques, the use of feature selection in artificial intelligence (AI), the impact of feature selection in AI applications and challenges experienced.** The paper distills the latest developments in feature selection, highlights issues around its classification and how it has been used in an array of AI applications. It concludes that feature selection plays a pivotal role in artificial intelligence applications. However, in order to realize its full potential, there are significant issues that must be dealt with for instance computational complexity, scalability and stability of feature selection algorithms.

Keywords: Machine Learning, Artificial Intelligence, Feature Selection, curse of dimensionality, feature correlation, computational complexity.

1. INTRODUCTION

The exponential increase in data significantly reduces the quality of data needed for data mining, pattern recognition, classification, clustering and other machine learning applications. This kind of data is said to be high dimensional and has a lot of noise, irrelevance and redundancy. **To improve performance of areas where the data is applied, we need to reduce these dimensions which is part of pre-processing of the data. The two main dimensionality reduction methods that can be used are feature selection and feature extraction.** The research community has engaged in feature selection research in terms of the criteria for selecting features and recommended for further study the aspect of mapping feature selection techniques to suitable applications [1, 2]. Feature selection is the process that involves choosing a subset of attributes from the original feature set so that the feature space is optimally reduced according to a certain criteria. **Feature selection is a critical step in the feature construction process. It significantly reduces the dimensionality of the feature space hence increasing the speed and predictive accuracy of the learning algorithm. This ultimately improves the comprehensibility of the learning results [3].**

Artificial intelligence has been applied in an array of domains that include classification, clustering. **The main objective of this paper is to provide a comprehensive review of the current state of the art of feature selection methods and how they have performed in different application areas.** We look at some of the feature selection methods developed for the purpose of enhancing performance in machine learning and other applications [4,5]. We focus on feature selection methods using supervised methods, unsupervised methods and semi-supervised methods. Supervised techniques select relevant features based on labelled datasets. Unsupervised feature selection methods identify and select relevant features without using class label information while semi-supervised techniques on the other hand use both labelled and unlabeled data to evaluate the relevance of features. This paper brings out the latest developments on feature selection methods, highlights relevant issues

and suggestions for further research. **Therefore the contributions of this paper are as follows: The paper: (a) provides an overview of feature selection and feature selection methods for understanding; (b) compares the performance of feature selection techniques in different application domains; and finally, (c) provides a recommendation of suitable application domains for feature selection techniques.**

The rest of the paper is organized as follows. In Section 2, the process of literature review is presented followed by an overview of applications of feature selection algorithms or techniques in section 3. This section also has recommendations of applications for feature selection techniques for unsupervised, supervised and semi-supervised feature selection. In Section 4, we present the open issues and areas of further research while section 5 concludes the paper.

1.1 Process of Literature Review

This literature review was done using a combination of quantitative and qualitative literature review approach to problem solving. This combination has a number of advantages over other methods namely: It's able to enumerate different spheres covered by existing work while pointing out the gaps to be filled and also access literature and new insights from various perspectives [6]. **This survey of applications of feature selection methods involved the use of online databases and other sources of materials or articles that meet a certain criteria or threshold from which relevant information was extracted and summarized.** This was done through systematic literature review [7] with the following steps:

- **Identifying Databases:** This includes electronic databases such as Google Scholar, Web of Science, Research Gate and Science Direct with a collection of published academic journals.
- **Selecting Keywords:** This is where we selected the keywords used for the searches e.g. "Feature Selection", "Classification", "Clustering", "Machine Learning", "Feature Selection Applications" and "Feature Set".
- **Selecting the Range of Papers:** The research was limited to papers published between 2005 and 2021.
- **Choosing Exclusion Criteria:** This research restricted itself to academic papers published in English as well as news articles, stories and annual reports on feature selection methods and FS applications.
- **Retrieving & Recording of Papers:** The papers retrieved were recorded based on information about authors, the year of publication and the journal in which the study was conducted. Subsequently, each article was classified according to the method used and whether the analysis was quantitative, qualitative or mixed.
- **Identifying Feature Selection and Applications Gaps:** This analysis was done to determine gaps in relation to suitable applications of feature selection and make recommendations for future studies.

For this paper, all the scientific papers were accumulated from available online resources. Digital databases such as IEEE Xplore, Google Scholar, Science Direct and the ACM Digital Library were used to obtain scientific articles that were used in this review. The chosen literature, as well as the keywords used for the search process were Feature Selection, Classification, Clustering, Machine Learning, Feature Selection Applications and Feature Set.

2. LITERATURE REVIEW

2.1 Feature Selection Overview

Features are distinctive attributes in a data set that can be used to measure a process being observed. Feature selection as a process involves removing irrelevant and repeated features from a data set in order to improve the performance of machine learning techniques and their applications [5,12, 11]. **Feature selection improves the performance of prediction models by removing the burden caused by dimensionality, consequently enhancing generalization performance, speeding up the learning process, and improving interpretability of models** [17, 18]. The author in [9] reiterates that feature selection has been used to handle the curse of dimensionality in order to enhance performance of data mining and machine learning techniques. It can also reduce computational cost and improve the interpretability of the model used for a defined task. **Proper feature selection enhances learning accuracy for classification and other applications like clustering and regression tasks. In cloud systems, selection of features forms part of the pre-processing stage, and enables identification of the main or more useful features that can be used for learning.** The use of features therefore enhances classification accuracy and decreases complexities in computation [14].

The feature set used in constructing a classification model is the main information source for a learning algorithm, and is therefore important to choose an optimal set that best represents the entire data set. Selecting too many features may increase the computational cost of the classifier while very few features may eliminate better features that would have increased

classification accuracy [11]. **Thus feature selection should focus on selecting a set of variables that adequately represent input data while reducing effects of irrelevant and redundant data and providing good classifier performance** [12]. This paper reviews literature of existing techniques that have been used to find sets of features that can improve the performance of classification, as well as similar applications using machine learning algorithms.

Redundancy and relevance are important in optimal feature selection. A feature can be categorized depending on certain characteristics that have been identified. **One of the options is based on strong relevance, which means that the feature will definitely form part of the final subset.** In case such a feature is removed from the relevant set, it affects its composition. Other options include a weak relevance in which features are not necessarily required to form optimal sub sets, and finally a case where redundant features are not needed at all. Redundancy thus checks in during assessment of feature subsets while relevance concentrates on specific attributes in a set [19, 20]. A criterion for choosing acceptable feature sets is required to remove non-relevant options. It is also necessary to determine a procedure for a sub set of required features that are necessary and should also be devised. Some of the methods for selecting features that can be used for classification and other applications are filter, wrapper and embedded techniques [16, 12]. **The filter option ranks features as those with higher values and are part of a set used for prediction [15]. The wrapper technique uses performance as the criteria of the predicting feature. Embedded methods on the other hand focus on selecting variables during training. Various techniques of selecting and extracting data attributes exist and have been used extensively as in [14, 11].**

2.2 Feature Selection Algorithms

Feature selection methods are generally used for pre-processing data in order to ensure that data is efficiently reduced. **This pre-processing may involve cleaning data by removing noise, reducing the size of the data set or adapting the data set to best suit the chosen model. High dimensions of data that contain hundreds of variables have prompted scholars to document a good number of feature selection methods** [16, 13, 11]. Feature selection tasks usually involve using machine learning techniques to select a small group of relevant text that then forms the basis of classifying data or text. It has been applied in broad domains like text categorization, bioinformatics, astronomy, data mining, and pattern recognition. The classification task and other machine learning applications may involve high data sets that are challenging to handle especially when the data has been collected from various techniques, methods and devices. Most of these massive data sets contain a lot of redundant and non-relevant data due to high dimensional features that downgrade the performance of machine learning techniques. **This happens in terms of over-fitting, increasing the time to develop machine learning models and reducing the accuracy of classification** [5, 21]. **Consequently, it is imperative for the number of features to be reduced to levels that would give better performance such that existing challenges can be handled by the process of feature selection.**

Many dimensions coupled with huge amounts of data normally pose a great problem to classification and other machine learning tasks. This fact makes feature selection algorithms very necessary. Feature selection algorithms vary in terms of how the search for optimal features is carried out. These algorithms can be classified into supervised, unsupervised and semi-supervised methods. **Supervised techniques select relevant features based on labelled datasets. Unsupervised feature selection methods identify and select relevant features without using class label information while semi-supervised techniques on the other hand use both labelled and unlabeled data to evaluate the relevance of features** [16, 22]. These techniques are reviewed in details hereunder.

2.3 Supervised Feature Selection Methods

Supervised FS techniques select relevant features based on labelled datasets. They are further classified as filter, wrapper and embedded feature selection algorithms. **The filter model works in a way that feature selection and learning of the model are independent. The wrapper model uses a small set of features to train the model. It enables the model and features to interact through feature dependencies. Embedded model on the other hand mainly deals with selecting features that rate highly in terms of accuracy** [24, 25]. The feature search process is embedded into the classification algorithm, and the learning process and the feature selection process can't be separated.

2.3.1 Filter Methods

A filter method is a feature ranking method which evaluates the relevance of features through internal data properties. In this method, features are ranked using suitable criteria and are subsequently either selected or dropped as per the threshold set or established. Here, features are selected independently in relation to the classifier [24]. The filtering process is illustrated by figure 1.

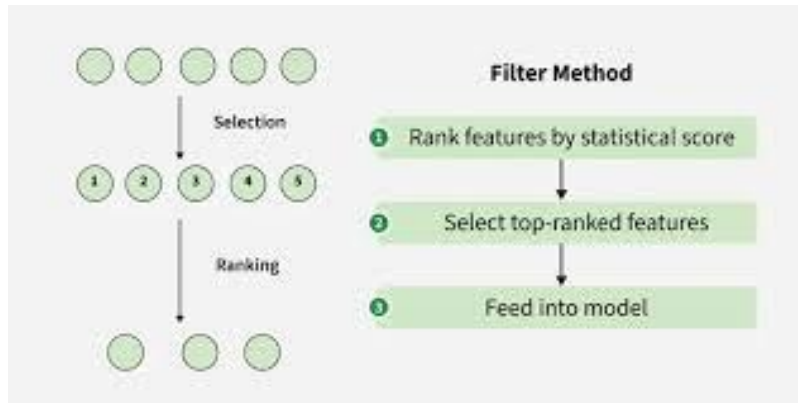


Fig. 1: Filtering Process

Filter methods can handle either single variables or multiple variables. Single-variate techniques do not consider dependencies among features, and examples include Information Gain, Gain Ratio and Symmetric Uncertainty. Multi-variate methodologies model dependencies among features independent of the classifier e.g. Feature Selection based on correlation and Minimum Redundancy Maximum Relevance [24]. In their studies [23, 24] have illustrated that filter methods for minimizing redundancy and maximizing relevance are useful in reducing feature sets by discarding any feature that is redundant or not relevant so as to improve prediction performance.

Filter methods have been applied in bioinformatics and data mining among other areas. The analysis of filter methods' application noted that they are highly accurate, fast and more suitable for processing high dimensional data though they are independent of the classifier. However, they cannot remove features in systematic targets. Filter methods also totally ignore the effects of the selected feature subsets on how an induction algorithm will perform. They do not consider the biases of machine learning algorithms used in their applications and also, not all filter features can be used for all the classes of data mining tasks [22,25]

2.3.2 Wrapper Methods

This is referred to as wrapper because it wraps up or combines the feature selection process with a predefined classifier. The wrapper approach involves the use of a small set of features for training the model. This approach derives its merit from enabling interaction between features and models while considering dependencies of features. Here, the classifier is used to select features by search [26]. The wrapper approach has been implemented using sequential selection algorithms and Heuristic Search Algorithms. Figure 2 illustrates the wrapping process.

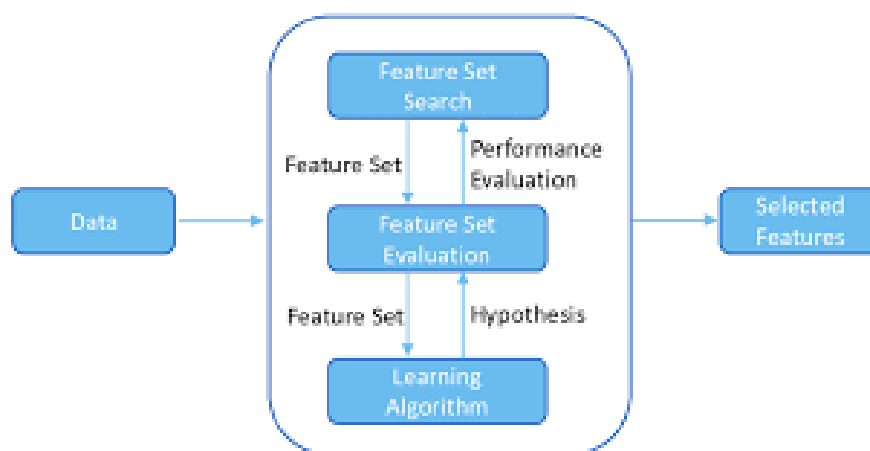


Fig. 2: Wrapping Process

Wrapper technique has been applied widely in areas like data mining, computer vision and industrial applications. It gives better results than filter method except that it is more computationally expensive especially when the number of features is large. [16] Developed a model in relation to this approach that focused on searching for an acceptable lean subset of attributes to be applied in a specific area. Cross validation is preferred as an accuracy estimation technique for the evaluation technique of this model. This study established that using wrapper technique to select features to a greater extent improved classification performance of different algorithms on all data sets that they used. On real data sets the wrapper options outperformed filter options on comparison. Performance of Naïve Bayes algorithm for instance is very outstanding on real data sets once discretization and feature subset selection has been done. [26] Reviewed a wrapper approach for feature selection, and used genetic algorithms to search for and generate subsets of features. This was tested on different classifiers e.g. Naïve Bayes and Decision Trees and experienced a good performance. [20] Explored a wrapper model for handling large scale attribute selection which involved modifying a search method which they named forward selection to create a model for high dimensional data. This proved to be very effective when its computational capabilities were tested. Wrapper methods are however generally slow compared to filter techniques.

2.3.3 Embedded Methods

The Embedded method mainly deals with selecting features that rate highly in terms of accuracy and sets apart a number of feature sets that are subsequently analyzed according to a given criteria. It incorporates feature selection process inside a machine learning model. **Feature selection is actually done when the model is being trained. The method aims at reducing the amount of time used to compare and select data sets by incorporating selection of features.** Embedded methods are faster and more accurate compared to filter methods and less prone to over fitting [24, 27]. Some embedded methods perform feature selection using regularization models which minimize fitting errors. Examples of embedded methods are Classification and Regression Tree (CART), C4.5 Decision Tree, Random Forest Algorithm, Extra Tree, Lasso and Elastic Net.

The embedded process is shown by figure 3

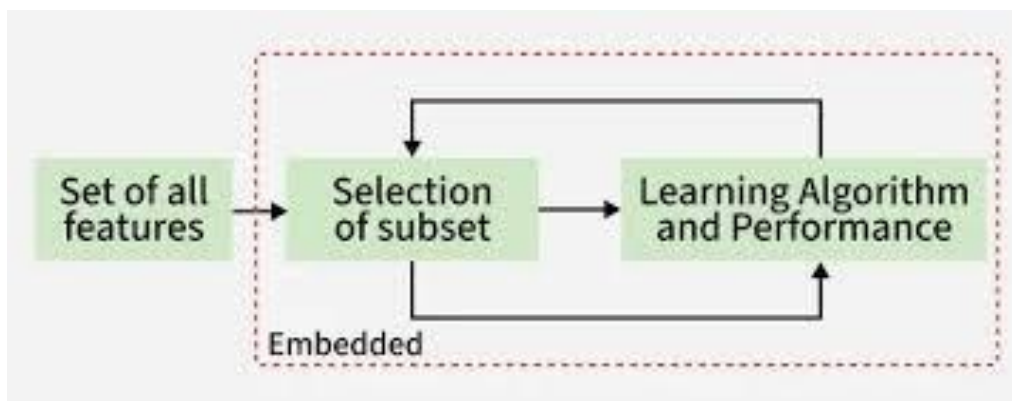


Fig. 3: Embedded Process

Embedded methods have also been used in different areas namely clustering, image recognition, text categorization and systems monitoring. [24] Proposed an embedded model named Multiple Criteria Linear Programming (MCLP) algorithm, which was backed by embedded backward feature selection. On analysis, the model gave good results. Similarly [6] developed a model called Embedded Sequential Feature Selector (ESFS), which allows for simultaneous selection of the most relevant features and their classification without an extra classifier. The method also exhibited better performance in comparison with other existing methods. Yumeng et al. [10], used an ensemble embedded feature selection method to classify multi-label clinical data. This was done by adequately utilizing label correlations by multi-label classifiers and evaluation measures. This method exhibited significant superiority over a number of state of the art algorithms.

2.4 Unsupervised Feature Selection Method

Unsupervised feature selection methods identify and select relevant features without using class label information. They score data dimensions based on different criteria namely data variance, entropy and the ability to separate data.

They have been broadly used to remove irrelevant attributes by reducing dimensions. Examples of unsupervised feature selection methods include Principal Component Analysis and Feature Similarity. Principal Component Analysis (PCA) is where the principal components from the data set are generated. It is done through determining correlation between features to identify the most significant principal components. It is a conversion technique which makes it possible to reduce the size of data sets which include a large number of interrelated features, so that the current data can be expressed with a fewer number of variables [28].

Using variable correlation here generates features that improve performance of the algorithm by reducing time and over fitting of the model. Laplacian linear discriminant analysis (LLDA) has also been used in unsupervised feature selection cases. LLDA recursive feature elimination (RFE) was applied to public data sets of cancer microarrays and its performance compared with that of Laplacian score and singular value decomposition entropy. From the results, LLDA-RFE performed very well compared to singular value decomposition entropy and fisher score method. **Unsupervised feature selection methods have been applied in other different areas namely data clustering, data mining and pattern recognition [28,29].**

3. FEATURE SELECTION APPLICATIONS

3.1 Current Applications of Feature Selection

The choice of a feature selection method can be informed by the application domain. Feature Selection has been applied broadly in Classification and Clustering. The aim of classification is to select a subset of relevant features used to build effective predictive models. This is achieved by removing non-relevant and redundant features [1]. A typical classification task is to distinguish between healthy and cancer patients based on their gene expression profile. Feature selectors are used, along with some initial filtering to drastically reduce the size of these datasets which would otherwise be unsuitable for further processing. Such feature selection has been applied in many other classification areas for instance text mining, computer vision, spam filtering, fault diagnosis and bioinformatics. The work in [10] has an elaborate study of filter feature selection methods for text classification, where various feature selection metrics were evaluated on two hundred and twenty nine (229) text classification problem cases. In the study, feature vectors were formed not as word counts, but as Boolean representations of whether a certain word occurred or not. A linear Support Vector Machine (SVM) classifier with un-tuned parameters was used to evaluate performance. The results in the study were analyzed with respect to precision, recall, F-measure and accuracy. Information gain was shown to perform best with respect to precision, while the author-introduced method of bi-normal separation performed best for recall, F-measure and accuracy.

Text clustering is the other broad application area for feature selection, and it groups similar documents represented as a bag of words which introduces high dimensionality of features and data scarcity. This in turn may affect the performance of clustering algorithms hence the need for feature set reduction. Feature selection has been used in various clustering applications namely web page and book mark categorization, market segmentation, social network analysis, medical imaging, image segmentation and anomaly detection [9,30]. [23] Investigated the use of feature selection in the problem of text clustering, showing that feature selection can improve its performance and efficiency. Five filter feature selection methods were tested on three document datasets. Unsupervised feature selection methods were shown to improve clustering performance, achieving about 2% entropy reduction and 1% precision improvement on average, while removing 90% of the features. The authors further proposed an iterative feature selection method inspired by expectation maximization that combines supervised feature selection methods with clustering in a bootstrap setting. The proposed method reduced the entropy by 13.5% and increased precision by 14.6%, hence coming closest to the established baseline, obtained by using a supervised approach.

Many scholars have attempted to handle selection of features in different applications. For instance, filter methods have been used extensively, and they generally work very quickly because they don't learn and can be scaled to different applications. These methods however do not foster dependency among attributes as well as how features directly associate with classifiers where they are used. Wrapper methods are comparatively slow because at each stage they need to accomplish some evaluation as they select feature sets. The performance of embedded methods depends on the rate of evaluation of sets, and if the evaluation is done at a faster rate, then they would be faster than wrapper methods and slower otherwise. These embedded methods can perform better than the filter method depending on the frequency trainings that they are taken through [22,24,26,27]. A summary of current feature selection applications is given in table 1.

Table 1: Current Feature Selection Applications

S/No.	Application Area	Feature Selection Algorithm	Feature Selection Approach
1.	Microarray data	Genetic Algorithms	Hybrid
		Iterated local search	Embedded
		Particle Swam Optimization and Tabu search	Wrapper
		Eigenvector Centrality Feature Selection	Filter
		Elastic Net Penalty	Semi-Supervised FS
2.	Handwritten digits/ Handwriting Recognition	Robust Unsupervised Feature Selection via Matrix Factorization	Unsupervised FS
		Minimum- Redundancy Maximum Relevance	Wrapper
		Class dependent features	Unsupervised FS
		Harmony Search Algorithm	Unsupervised FS
3.	Texture classification	Ant Colony Optimization	Feature subset selection
		k-Nearest Neighbor	Supervised FS
4.	Hyper spectral image classification	Symmetrical Uncertainty and Approximate Markov Blanket	Filter
		Support Vector Machines	Supervised FS
		Maximum Likelihood Classifier	Supervised FS
		K-means classifier	Unsupervised FS
5.	Image recognition	Locally Linear Embedding	Filter
6.	Spam & Intrusion detection	Mutation and Binary Particle Swarm Optimization	Wrapper
		Mutual Information	Filter
7.	Computer vision	Infinite FS	Filter
8.	Brain-computer interface	Linear discriminant analysis	Supervised FS
9.	Fault detection and diagnosis	Information Greedy Feature Filter	Filter
10.	Image annotation	Structured Multi-view Hessian sparse Feature Selection	Semi-Supervised FS
11.	Binary & Multi-category data	Relevance, Redundancy, Incremental Search	Semi-Supervised FS

3.2 Recommendation of Feature Selection Applications

Feature Selection has been applied in different domains as highlighted in this work. **Performance of these methods is based on different parameters like scalability, computational cost, stability and data**

representation. These have been applied across different studies and can be used as a basis of selecting suitable applications of FS methods [17,19]. Other measures have also been used to evaluate methods for selecting and classifying features, and how they perform. These include accuracy, recall, precision and F measure. **These metrics are normally selected depending on the nature of the task or tasks in question and the kind of data sets to be used.** For example data sets in medicine concentrate on accuracy, but generally accuracy and other measures can be derailed by the use of features that are repetitive and others that are not relevant. It is therefore important to cut down dimensions of data to encourage better performance. Based on the performance with respect to these metrics, feature selection techniques can be recommended for various applications [21,24,25].

The concept of carefully selecting a subset of features is very important, but again how this is done determines the kind of results obtained. **In particular the number and quality of features selected must be weighed because they solely inform and direct the algorithms or models used in various applications.** For instance if a very lean set of attributes is selected, chances are that its information may not be sufficient for the algorithm to work well. **Conversely, selecting a huge set of attributes may be noisy to the algorithm thus leading to worse performance.** In this section, we recommend application areas that best suit various feature selection methods. Table 2 gives a summary of these applications.

Table 2: Recommended applications for FS techniques

S/No	FS Approach	Recommended Applications
1.	Supervised Feature Selection	Image Recognition/ Image Classification
		Spam Detection
		Intrusion Detection
		Computer Vision
		Brain Computer Interface
		Fault detection and diagnosis
		Data Mining
		Pattern Recognition
2.	Semi-Supervised FS	Micro-array data
		Image annotation
		Binary & Multi-category data
		Medical Diagnosis
		Fraud Detection
		Forensic Science
		Text Processing
		Facial Recognition
3.	Unsupervised FS	Bioinformatics
		Data Clustering
		Handwriting Recognition/ Hand written Digits

4. OPEN ISSUES AND FUTURE RESEARCH DIRECTIONS

From this survey, a number of areas that confound researchers in feature selection were identified. These need to be addressed for successful feature selection applications. The issues include scalability of feature selection, stability of feature selection algorithms, optimal number of features for particular FS algorithms and secure feature selection.

1. **Scalability of feature selection algorithms:** From this review, it has been noted that the scalability of most feature selection algorithms is affected by the spontaneous increase in data sizes especially in scientific and business applications. **These huge data sets usually cannot be directly loaded in memory which ends up impacting negatively on the usability of feature selection methods.** This also applies to extra-high dimensional data (big data) from text mining and information retrieval applications. This directly affects the performance of feature selection algorithms in terms of efficiency and computational cost. **Researchers therefore still need to come up with ways of addressing scalability with respect to big data.** Distributed programming framework has been used to try to solve this through parallel feature selection on large data sets which has worked only to some extent. Infinite data or feature streams that cannot be pushed into the memory also affect the scalability of FS algorithms. **In order to handle this, these features need to be kept in memory or an iterative process for handling multiple data instances should be used. This still proves to be a hard nut to crack hence needs to be researched further.**
2. **Stability of feature selection algorithms:** Stability of supervised feature selection algorithms is another issue of importance in relation to application that needs further research. A good number of FS algorithms that are used regularly depict low stability when slight perturbation is introduced in the data set, something that should be sorted. **Generally, we have seen that feature selection stability can be affected by some level of data perturbation and underlying characteristics of data for instance, dimensionality of data and the number of data instances.** When data samples are added/deleted or noisy/outlier samples are added, slightly different sets of features are generated from new data samples on the slight data perturbation which affects the stability. **Stability is important because it boosts the confidence in the features selected for a particular application. More work therefore needs to be done on how to ensure stability of FS algorithms for better feature sets and better performance.** For unsupervised feature selection, it's challenging to study stability because there is no prior knowledge on the cluster structure of data hence creating uncertainty on whether instability is caused by existing clusters or newly introduced data. Unsupervised feature selection is more sensitive to noise which directly impacts on the stability of these algorithms.
3. **Secure feature selection:** The other open issue and possible area of further research is the aspect of secure feature selection. From the review, there a number of concerns relating to security of feature selection that still need to be addressed. Firstly, security of data used for feature selection should be ensured. Secondly, feature selection needs to be done in a way that preserves privacy of the whole process. Thirdly, some organizations that may require feature selection applications are so sensitive and so may be strict on matters security of both data and the process. **Researchers therefore need to dig deeper into how security of both the process of feature selection and data involved in the process should be ensured.**
4. **Optimal number of features:** The last open and challenging issue is the optimal number of features to use for a particular FS algorithm. For some FS algorithms, the approximate number of features to be selected is normally indicated. However, determining the optimal number of features that would probably give the best feature set required still remains a challenge. **In all the papers reviewed, the number of features selected actually depends on the FS algorithm and approach used.** There is no methodology proposed or used to determine the optimal features making this an open area of research. Generally, selecting too many features may increase the computational cost of the classifier while very few features may eliminate better features that would have increased classification accuracy. **Thus feature selection should focus on selecting a set of variables that adequately represent input data while reducing effects of irrelevant and redundant data and providing good classifier performance.** A heuristic way to grid search a number of selected features and choose the number of features that would supposedly give the best performance in classification or clustering has been used. The challenge with this approach is that it is computationally expensive.

5. CONCLUSION

This study aimed at providing a critical review of feature selection techniques with specific focus on application domains. **We provided an overview of feature selection methods for purposes of understanding, compared the performance of feature selection techniques in different application domains, and provided a recommendation of suitable application domains for feature selection techniques.** We have summarized techniques of selecting features by exploring existing literature and studies of various researchers and also analyzed and classified feature selection algorithms. **The performance of feature selection algorithms can be compared using common data sets because they may behave differently on varied data sets.** Research has shown, for instance, that feature ranking based methods is better than subset based methods when analyzed according to memory space and computational complexity.

A notable observation is the current research advancement in feature selection which has had a tendency of leaning towards the use of hybrid feature selection methods. It has been proved that when feature selection is done prior to classification, the performance in terms of accuracy and other measures greatly improves. This would even be better when hybrid methods of feature selection are used. **Future research and applications should therefore focus on ensuring that strategies for selecting features are more efficient and accurate by using hybrid options on wrapper and filter approaches. Based on the performance of feature selection techniques, we recommended suitable applications for these techniques. From the literature survey, we also suggested open issues or opportunities and the future research directions.**

REFERENCES

- [1] V. Bachu and J. Anuradha, "A Review of Feature Selection and Its Methods. Cybernetics and Information Technologies", 2020, 3. 10. 2478/cait-0001.
- [2] Y. Liang, J. Gan, Y. Chen, P. Zhou and L. Du, "Unsupervised Feature Selection Algorithm Based on Graph Filtering and Self-representation", 2024, 10.134. <https://doi.org/10.48550/arXiv.2411.00270>
- [3] S. Jonas and M. Dorn, "Enhancing classification with hybrid feature selection: A multi-objective genetic algorithm for high-dimensional data", Expert Systems with Applications, Volume 255, Part A, 2024, 124518, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2024.124518>.
- [4] T. Zhang, F. Wang and L. Zeng, "Global-local spatially aware preserving projection for dimensionality reduction of hyperspectral images," Infrared Physics and Technology, vol. 150, Nov. 2025, Art. no. 106014.
- [5] L. Ren, D. Wang, L. Gao, M. Wang and M. Huang, "Nonlocal and deep priors for hyperspectral anomaly detection," IEEE Transactions on Geoscience and Remote Sensing, vol. 63, 2025, Art. no. 5520415, doi: 10.1109/TGRS.2025.3593019.
- [6] A. Amina, P. Lucas and J. Kimberly, "Mixed Methods Research: Combining both qualitative and quantitative approaches", Mixed Methods, 2024.
- [7] W. Marc, "Systematic Literature Reviews: Reflections, Recommendations, and Robustness Check", Journal of Consumer Behaviour, vol. 24, Issue 3, 2025, <https://doi.org/10.1002/cb.2479>
- [8] L. Wenjing and W. Xiaofei, "Overview of Hyper-spectral Image Processing", Journal of Sensors. Hindawi, 2020.
- [9] H. Nirali, "A survey on Feature Selection Techniques", Geographical Information Systems science journal. 2020.
- [10] W. Yi-Min and S. Liu, "Semi-Supervised Classification of Data Streams by BIRCH Ensemble and Local Structure Mapping", Journal of Computer Science and Technology, 2020, 35(2): 295-304.
- [11] S. Solorio-Fernandez, J. Carrasco-Ochoa and J. Martinez-Trinidad, "A review of unsupervised feature selection methods Artificial Intelligence Review", 2020, 53, 907-948. <https://doi.org/10.1007/s10462-019-09682-y>
- [12] Z. Shu-Zheng, Z. Zhen-Yu, F. Chao-Chao and W. Lei, "A Machine Learning Framework with Feature Selection for Floorplan Acceleration in IC Physical Design", Journal of Computer Science and Technology, 2020, 35(2): 468-474.
- [13] K. Niedziewski, E. Maciej, R. Marchwiany, M. Piliszek and W. Michalewicz, "Multidimensional feature selection and High performance", SN Computer Science, 2020, 1, 40.
- [14] J. Nayak, K. Vakula, P. Dinesh, B. Naik and D. Pelusi, "Intelligent food processing: Journey from artificial neural network to deep learning", Computer Science Review, Volume 38, 2020, 100297, ISSN 1574-0137, <https://doi.org/10.1016/j.cosrev.2020.100297>.
- [15] B. Andrea, et al., "Benchmark for Filter methods for feature selection in high dimensional classification data", 2020.
- [16] P. Nikita, and I. Smetannikov, "Feature selection algorithms as one of the Python Analytical Tools", Future Internet, 2020.
- [17] K. Manisha, and A. Khaparde, "Automatic hand gesture recognition using hybrid meta-heuristic-based feature selection and classification with Dynamic Time Warping", Computer Science Review, Volume 39, 2021, 100320, ISSN 1574-0137, <https://doi.org/10.1016/j.cosrev.2020.100320>.
- [18] F. E. Bezerra et al., "Impacts of Feature Selection on Predicting Machine Failures by Machine Learning Algorithms". Applied Sciences, 2024. 14(8), 3337. <https://doi.org/10.3390/app14083337>
- [19] F. Kamalov, H. Sulievan, A. Alzaatreh, M. Emarly, H. Chamlal, and M. Safaraliev, "Mathematical Methods in Feature Selection: A Review". Mathematics, 2025. 13(6), 996. <https://doi.org/10.3390/math13060996>
- [20] Y. Lei and L. Huan, "Efficient Feature Selection Via Analysis of Relevance and Redundancy", Journal of Machine Learning Research. 2004. 5. 1205-1224.
- [21] M. Süpürülü, A. Hatipoğlu, and E. Yılmaz, "An Analytical Benchmark of Feature Selection Techniques for Industrial Fault Classification Leveraging Time-Domain Features". Applied Sciences, 2025. 15(3), 1457. <https://doi.org/10.3390/app15031457>
- [22] C. Xueyi, "A Comprehensive Study of Feature Selection Techniques in Machine Learning Models", Insights in Computer, Signals and Systems. 2024. 1. 65-78. [10.70088/xpf2b276](https://doi.org/10.70088/xpf2b276).
- [23] Z. Chen, W. Jiang, J. Tan, Z. Li, and N. Gui, N. "Supervised Feature Selection Method Using Stackable Attention Networks", Mathematics, 2025. 13(22), 3703. <https://doi.org/10.3390/math13223703>
- [24] C. Zhong et. Al., "Supervised Feature Selection with Class Self-representation". 2025. 10.1007/978-3-032-04558-4_51.
- [25] K. Kotlarz, D. Słomian, W. Zawadzka and J. Szyda, "Feature Selection Strategies for Deep Learning-Based Classification in Ultra-High-Dimensional Genomic Data", International Journal of Molecular Sciences, 2025. 26(16), 7961. <https://doi.org/10.3390/ijms26167961>
- [26] M. Salimi and B. Nouri-Moghaddam, "A Review of Wrapper Feature Selection Methods Based on Metaheuristic Algorithms for Improving Classification Accuracy", Future Generation Computer Systems. 2023. 1. 9.
- [27] A. Biernacki, "Evaluating Filter, Wrapper, and Embedded Feature Selection Approaches for Encrypted Video Traffic Classification". Electronics, 2025. 14(18), 3587. <https://doi.org/10.3390/electronics14183587>
- [28] F. Beiranvand, V. Mehrdad and M. Dowlatshahi, "Unsupervised feature selection based on the hidden knowledge of the Two-Dimensional Principal Component Analysis feature extraction method. 2024. 10.21203/rs.3.rs-4298823/v1.
- [29] L. Zhiqiang, Z. Wenzh, Z. Hongyu, M. Yuanqian and R. Jie, "The impact of unsupervised feature selection techniques on the performance and interpretation of defect prediction models", Automated Software Engineering. 2025. 32. 10.1007/s10515-025-00510-y.
- [30] Z. Lim, L. Ong and M. Leow, "A Review on Clustering Techniques: Creating Better User Experience for Online Roadshow". Future Internet, 2021. 13(9), 233. <https://doi.org/10.3390/fi13090233>