JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

Review on Machine Learning Feature Selection

Abubakar Atiku Muslim, Muhammad Saidu Aliero
Department of Electrical and Electronic Engineering,
Abudllahi Fodio University of Science and Technology Aliero

ABSTRACT

Research in machine learning application provides significant improvements in human activities and surroundings. Feature selection is one of the important aspects of model development in machine learning in order to lower the computational cost, increases reliability, and makes the data easier for professionals and machine learning models to understand. There are a number of selection alternatives available, though. In this study, a number of datasets, various tactics from various viewpoints, and a three of the most prominent techniques from various feature selection categories are compared based on analytical evaluation. The findings indicate that, wrapper techniques are more dependable but theoretically costly, filter procedures are usually quicker based on various computational strategies aim to get throughput prediction.

ARTICLE INFO
Article History
Received: August, 2025
Received in revised form: October, 2025
Accepted: December, 2025
Published online: January, 2026

KEYWORDS
Machine Learning, Feature Selection, IDS, SVM

INTRODUCTION

Today, machine learning (ML) models are being used more in the big data era to extract knowledge and trends from enormous datasets [1]. However, the calibre of the features offered for training frequently determines how efficient these models are [2]. One of the most important stages of preprocessing is feature selection, which entails removing unnecessary or duplicate features from the dataset while selecting the most pertinent ones [3]. Through this approach, the model's predictive abilities and computational effectiveness are improved in addition to its interpretation. Machine learning algorithms have difficulties when dealing with high-dimensional data, which is frequently found in fields like genomics, image processing, and natural language processing [4]. Overfitting occurs when a model performs well on training data but badly on unknown data due to the presence of excessive or unnecessary features [4].

Previous research have review different feature selection approaches by lowering dimensionality, feature selection [5]. The authors provides recommendation to helps to cut down on the time and statistical expenses needed for model training. Similarly, the review on feature elimination techniques in [6] shed light that improves generalization and forecast accuracy. Furthermore, it reduces the possibility of overfitting and simplifies models, making them simpler to comprehend and implement [6]. Research in compared wrapper and filter techniques for feature selection using Pearson's correlation coefficient which quantifies the straight-line correlation between two independent variables [7]. The result of authors finding in [7] were utilized in formal research to pinpoint features that directly influence model prediction results. Despite several review studies with different approaches of evaluation with significant recommendation. However, it is important to compared most commonly used approaches for feature selection in order to shed light for research with regard to cost, speed and complexity [8]. The research of the paper is organized as follows: Section 2 provides review of related work and section 3 presents research conclusion.

RELATED WORK

Feature selection can also be divided into three categories based on the various pursuit approaches: filter methods, wrapper methods, and embedding approaches. Filter techniques use the

Corresponding author: Abubakar Atiku Muslim
alatiku@gmail.com
Department of Electrical and Electronic Engineering, Abudllahi Fodio University of Science and Technology Aliero.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

characteristics of the data to identify the most distinguishing traits. Filter approaches often fall into a two-step approach and pick features prior to classification and clustering activities. First, a set of parameters is used to rate each characteristic. The features that have the best rankings are then chosen. Wrapper approaches analyse the characteristics using the chosen learning technique itself. Support Vector Machine techniques are used in the work [8] to identify the gene most associated with malignancies. During the model-building process, embedded models carry out feature selection.

Filter Based Approach

Even with increasing knowledge of network security, current approaches are still unable to completely defend networks of machines and online applications from risks posed by constantly evolving cyberattack strategies like malware and denial-of-service attacks. As a result, creating security strategies that are both efficient and flexible is now more important than ever [9]. Conventional security methods, such firewalls, user authentication, and data encryption, which are the first line of defence against intrusions, are not enough to cover the whole network security environment while dealing with constantly changing intrusion strategies and talents [10]. Therefore, an additional line of defence for security, like IDS, is strongly advised. Antivirus applications and an IDS have recently evolved to be crucial components for many businesses security systems. Combining these two lines improves network security and offers a more thorough defence toward such risks.

In order to improve network security, a great deal of investigation has been done to create intelligent intrusion detection methods. Among the first attempts to create IDS were [11] and packed boosting based on C5 decision trees. ML approaches like Support Vector Machine (SVM) have been effectively used in methods suggested in [12] to categorize network activity trends that differ from typical network activity. Findings from experiments demonstrate the endurance and efficacy of SVM in IDS. In order to detect intrusions,[12, 13] looked into the possibilities of combining different learning techniques, such as SVMs, and Artificial Neutral Network (ANN). To differentiate between the four main kinds of attacks and regular traffic, they trained five distinct classification systems[5]. After comparing every training technique's effectiveness compared to their model, they discovered that the combination of ANNs and SVMs produced the best accuracy when classifying across all five classes. In their construction of a detection system, [7] integrated several fuzzy classifiers and optimized the architecture of the fuzzy systems utilized in the classifiers using a genetic algorithm. The traffic that came in was detected using the pre-established fuzzy inference system. We have put out an anomaly-based method for DoS attack detection [14].

Using the KDD Cup 99 and ISCX 2012 datasets, the system demonstrated encouraging identification reliability of 99.95% and 90.12%, accordingly. The efficacy of IDSs is challenged by processing massive data, which could slow down the detection process overall and result in subpar classification accuracy because of the computing challenges involved [15]. Large-scale data classification typically results in numerous mathematical challenges, which raises the computational complexity. as a well-known dataset for intrusion evaluation. Over five million examples for training and two million samples from testing make up this dataset, accordingly. Such a vast dataset slows down the classifier's development and testing procedures or prevents it from working properly because of system errors brought on by inadequate memory.

Additionally, noisy, duplicated, or uninformative characteristics are frequently found in huge datasets, which poses significant obstacles to data mining along with information discovery. The process of feature selection involves removing duplicate and unnecessary characteristics and choosing the best subset of features that better characterizes trends across classes. Filter approach offers feature selection techniques fall [16]. While wrapper model use specific learning techniques to assess the value of features, then filter model use different metrics as a criterion for assessing the relationship of a set of

Corresponding author: Abubakar Atiku Muslim
alatiku@gmail.com
Department of Electrical and Electronic Engineering, Abudllahi Fodio University of Science and Technology Aliero.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

features [17]. When working with large-scale or high-dimensional data, wrapper approaches are frequently far more expensive to compute than filter techniques. Therefore, we concentrate on filtering techniques for IDS in this work. Feature selection as a pre-processing step has grown into a crucial component of developing intrusion detection systems because of the ongoing increase in data dimensionality [18].

[19] suggested a novel feature selection approach to narrow down the KDD Cup 99 dataset's feature space from 41 dimensions to 6 dimensions. An SVM-based IDS was then used to assess the six features that were chosen. The findings indicate that employing the chosen features improves the reliability of classification by 1%. [20] examined how well a Markov model performed when used for feature selection, demonstrating that it could cut the KDD Cup 99 feature count from 41 to 12. A proactive feature selection approach that measures the relationship between features using the sharing of data method has been put forward by [18]. The SVM classifier was then trained using the ideal feature set, and the IDS was constructed. suggested an SVM-based IDS that integrates the SVM with clustering [21]. In order to decrease the typical training and testing time and enhance the classifier's ability to classify, the clustering model was utilized to give the classifier fewer, higher-quality training data. The SVM-based IDS achieved an overall accuracy of 95.75% with a false positive rate of 0.7% when tested on the corrected labels KDD Cup 99 dataset, which contains some novel attacks. Table 1 presents summary of advantage and disadvantage of filter approach.

Table 1: Advantage and disadvantage of filter-based approach

Advantage	**Disadvantage**
Fast: majority of filter-based model rapidly complete the given task without compromising the model performance.	Ignoring feature dependency: The potential to overlook feature dependencies, obtain unstable features, and not be able to utilize every feature in a wide feature space are the main obstacles in feature selection.
Scalable: when it comes to scalability, filter approach can handle large volumes of data or traffic without sacrificing model performance.	Redundancy: filter approach may lead to features redundancy which might result in overfitting, redundancy is usually seen as harmful.
Independent: In general, filter-based model proved to be classifier independent indicating that the transition or operation is not dependent on or unique to a certain ML technique.	Interaction with classifier: filter model naturally does not have feedback loop control on classifier after training.
Computational complexity: filter approach offers computational complexity which is essential since it influences the viability and effectiveness of algorithms when used on big datasets.	Discrimination power: The filter approach lacks strong discrimination power which has two goals. At first to generate artificial dataset that is almost identical to genuine data. Essentially, it aims to mislead the discriminator.

As described in in Table 1, filter methods are a common option for feature selection since they provide several benefits. Because they assess features irrespective of any ML technique, they are extensible and computationally fast. They are especially well-suited for high-dimensional datasets because of their efficacy, which allows for quick preprocessing. They are also adaptable in exploring data due to their algorithm nature, which guarantees that they may be used consistently across various ML models. Simple filtering techniques like chi-square testing and correlation analysis make comprehension simple and offer

Corresponding author: Abubakar Atiku Muslim
alatiku@gmail.com
Department of Electrical and Electronic Engineering, Abudllahi Fodio University of Science and Technology Aliero.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

insightful information about the connections between attributes and the target variable.

Furthermore, filter methods lessen the possibility of overfitting, which can happen with more intricate selection processes through minimizing dealing directly with model training. Filtering techniques do have some significant drawbacks, though. They disregard potential connections among variables and assess features individually.

Consequently, they could be unable to recognize feature groupings that work together to improve model performance. Furthermore, a lot of statistical metrics, like correlation, that are employed in filtering techniques are restricted to linear correlations and may miss significant nonlinear dependencies. Additionally, the feature set is needlessly complicated by filter algorithms, which frequently choose duplicate characteristics that have a high correlation with one another. Filter methods may also choose features that don't necessarily improve the computational model's predicted accuracy because they function independently of it. Although filter methods offer a quick and simple way to choose features, their drawbacks show that additional methods are required to get the best results in complicated datasets.

Wrapper Based Approach

The number of intrusions rises annually because of people' increased use of networking devices and the Internet to share content. Researchers have been motivated to develop Intrusion Detection System (IDS) by the significance of preserving security in such a setting [22]. The IDS examines traffic flow characteristics to determine whether or not there has been an intrusion. IDS can be divided into two kinds based on their detection methods: misuse detection and unusual activity detection. A model of typical behaviour is developed for anomaly-based detection. Methods like grouping and machine learning techniques are used to develop a model of regular behaviour. Several areas of information extraction and machine learning explore feature selection using metaheuristic methods [23]. Therefore, the use of metaheuristic the enhancement techniques which can be categorized as random method that aim to obtain a nearly optimal solution through multiple repetitions is one of the practical and efficient approaches in the path closer to resolving the choice of features issue and its associated problems. Because the issues could not be solved optimally by the metaheuristic algorithms alone. Thus, integrating the benefits of algorithms can enhance their efficacy in the hybrid approach. Training the system and the training process is one way to increase the effectiveness of IDS; in this context, choosing more crucial characteristics is necessary for better training.

Several areas of information extraction and machine learning explore feature selection using metaheuristic methods. Therefore, the use of metaheuristic the enhancement techniques which can be categorized as random algorithms that aim to obtain a nearly optimal solution through multiple repetitions is one of the practical and efficient approaches in the path closer to resolving the choice of features issue and its associated problems [1, 2]. because the issues could not be solved optimally by the metaheuristic algorithms alone. Thus, integrating the benefits of algorithms can enhance their efficacy in the hybrid approach. Training the system and the training process is one way to increase the effectiveness of IDS; in this context, choosing more crucial characteristics is necessary for better training.

[24] focused particularly on detecting Application Layer Distributed Denial of Service Attacks (DDoS) on web-servers by proposing a time series prediction model to detect flooding attacks and asymmetric attacks. The proposed algorithms effectively detect both attacks with a high detection rate and low false positives when compared with the work in [25] and [26]. Both papers detect DDoS attacks at the Application layer compared to our work that carries out detection at the network layer covering all endpoints in the network.

Another work that uses the Software Defined Network (SDN) architecture for dynamic attack detection and mitigation in Internet of Things (IoT) networks is in [13] where the authors aim to prevent the attacks at the network level

Corresponding author: Abubakar Atiku Muslim
alatiku@gmail.com
Department of Electrical and Electronic Engineering, Abudllahi Fodio University of Science and Technology Aliero.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

instead of the device level; a framework called SoftThings was proposed to prevent both a DDoS attack against an IoT device or using an IoT device to target another victim or system. SoftThings framework uses a SVM, a machine learning algorithm implemented in the SDN controller to detect anomalous traffic and subsequently mitigate the attack [12]. The algorithm was implemented using different topology scenarios. In the first scenario, an IoT device is under attack. In the second scenario, a compromised IoT device starts sending erratic packets. In the third scenario, two IoT devices are compromised and used to conduct a DDoS attack. The authors claimed SoftThings has a detection accuracy of 98%. The algorithm used is a supervised machine learning algorithm, which requires constant retraining of the detection algorithm and also requires a labeled dataset for normal traffic and malicious traffic.

On the other hand, [9] proposed an IoT specific network DDoS detection using machine learning algorithms such as K-nearest neighbors, support vector ma- chines, random forests, decision trees, and neural networks was built. Using some standard IoT devices, the authors set up an IoT network to capture benign traffic while also conducting flood attack, to capture malicious traffic; this will be used for the classifier training. The machine learning algorithms listed above were implemented using the Scikit-learn python library [18] using the captured dataset ; only the neural network algorithm was implemented using the [27]. The results from the experiments show that all the machine learning algorithms can effectively detect DoS attacks in IoT networks with accuracy ranging from 0.91 to 0.99. There is no evaluation and comparison of the proposed system with standard datasets provided in the study.

In [19], a clever method was applied to feature selection throughout the IDS process. The purpose of this study was to use SVM to optimize the feature selection process in order to improve the accuracy of earlier classifications. This objective was accomplished by employing the SVM to reduce characteristics in the KDD Cup99. The suggested method's results were contrasted to those of SVM, and hybrid form of SVM and Linear Regression (LR) methods, which showed substantial improvements. In the information extraction security network lab, more significant features were chosen using genetic algorithms. SVM-based classifiers were employed to analyse the feature selection procedure. According to the findings of assessing the effectiveness of LR and SVM classifiers without feature selection with a scenario including the usage of the dataset's feature count has decreased from 41 to 11, respectively. This resulted in a significant decrease in running time, while simultaneously increasing precision and detection rate and reducing the incidence of false alarms. It was also demonstrated that SVM performed better than the other methods tested.

The assessment criteria were the F-measure and the results showed that detection efficiency is improved, and computational expenses are decreased in relation to using all features [28]. Regretfully, no evaluation has been made between the outcomes of various metaheuristic algorithms and the outcome findings of the dragonfly metaheuristic method and the improved classifiers and Bayesian networks as well. For the goal of IDS, a new classification strategy was suggested in [29] that combined the SVM and RF (random forest) methods. The learning dataset was divided, and extraneous information was removed using fuzzy c-means methods of clustering and feature selection based on cohesion. Then, using the tree classifier method based on the chosen features, "if … then" rules were created to identify both typical and abnormal data.

Simulation was done on the KDD, a broad dataset appropriate for identifying more recent attacks, in order to assess the suggested approach. When the findings of this technique were contrasted to those of other approaches, it was demonstrated that the suggested approach performed better in terms of SVM and RF. The opposite strategy is used by Backward Selection, which begins with every trait and eliminates the least important one at each stage. This procedure keeps going until the model's performance noticeably declines after any extra features are

Corresponding author: Abubakar Atiku Muslim
alatiku@gmail.com
Department of Electrical and Electronic Engineering, Abudllahi Fodio University of Science and Technology Aliero.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

eliminated. Although backward selection offers a thorough assessment of features, it is costly, especially for feature-rich datasets [1]. Table 2 provides summary of advantages and disadvantages of wrapper-based approach.

Table 2: Advantage and disadvantage of wrapper-based approach

Advantage	Disadvantage
Interaction with classifier: unlike filter model, wrapper model continuously interacts with classifier during and after training.	Ignoring feature dependency: The potential to overlook feature dependencies, obtain unstable features, and not be able to utilize every feature in a wide feature space are the main obstacles in feature selection.
Underfitting: ensuring quality of training dataset in a model can improve predictive efficiency by optimizing model training by identifying underfitting.	limitation: restricted to applications that are linear and unsuitable for a lot of real-life issues.
Prone to local optima: Particularly in multimodal surroundings, achieving the local ideal usually takes less time and computer resources than finding the global optimum.	Computational cost: A major obstacle in wrapper model is high computational cost, which affects utilization of resources, real-world practicality, and efficiency of models. Large datasets, intricate techniques, and the requirement for strong hardware.
Computational complexity: filter approach offers computational complexity which is essential since it influences the viability and effectiveness of algorithms when used on big datasets	Poor feature selection linear model: wrapper model need the data points to be separated by a straight line in order to classify the data efficiently. These models struggle with more complex, interaction patterns and are only applicable to data that can be separated linearly.

Embedded Based Approach

In contrast to wrapper approaches, embedded methods [1, 2] aim to minimize the computing time required for redefining various subsets. The primary strategy is to include feature selection in the training procedure. Although we noted in prevision study[4] during the ranking process that used to produced excellent results because it only considered among the feature and the class output. In [25], the subsets are evaluated using a blind search technique. When a model is created, embedded methods choose features.

Study in [29] use unregulated feature selection utilizing autonomous learning to feature selection unsupervised learning alone, feature selection using unsupervised learning can yield a more accurate description and reliability of the data [30]. There are numerous articles in [8, 31] that try to use unsupervised learning to tackle the feature selection problem. The work in [11] suggested an ML method that modifies a hypothesis derived from labelled data alone by utilising both labelled and unlabelled data. The authors of [24] score a collection of features using a grouping flag construction. The authors of [13] employ the SVM for regularization. A relatively recent method for obtaining a stable feature subset is issue improvement, which is comparable to feature selection [32]. Several subsets of data samples derived from the bootstrapping approach are subjected to a single feature selection process.

A comprehensive feature set is created by combining the findings. To achieve the final feature subset, the authors in [33] employ a variety of aggregation techniques, including ensemble-mean, linear aggregation, and weighted aggregation approaches, after ranking the genes and features using filter techniques. Several methods for picking features can be used for a specific application, and the most effective one that satisfies the necessary requirements can be

Corresponding author: Abubakar Atiku Muslim
alatiku@gmail.com
Department of Electrical and Electronic Engineering, Abudllahi Fodio University of Science and Technology Aliero.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

chosen. The reliability of the feature selection techniques is an issue that is often disregarded. The capacity of a method for choosing features to consistently generate a feature subset upon the addition of fresh training samples or the removal of certain training samples can be considered to be reliable [34]. The method loses its reliability for feature selection if it generates distinct subsets for every disturbance in the training data. By altering the training set and repeating the procedure, it is possible to confirm that the study in [35] showed strong data extraction operations. Their stability is examined using wrapper methodologies, and durability measures are presented along with potential fixes to mitigate the issue.

To assess various subsets acquired for a specific number of runs, several metrics are developed. For various datasets, a stronger subset can be identified via these metrics. The development of the multicriterial fusing technique ranks and scores features using a variety of feature selection algorithms, resulting in a resilient subset that is based on integrating many classifiers to increase reliability. Additionally, the author recommends splitting the input features to create various classifiers and combining the predictions to arrive at a final judgment. Supervised feature selection techniques are used for labelled data. The association between many aspects cannot be taken into account by traditional supervised methods like [36], which score features individually based on the criterion.

By optimizing the ratio among the class scatter and the total class scatter, linear discriminant analysis. [14] was suggested as a way to improve features. Since LDA must compute the inverse matrix of within-class scatter, which is singular when the sum of training samples is less than the complexity of the data, it regrettably encounters a minor sample size issue [25]. In essence, feature selection finds, locates, and processes regions that are important using the object's features. Either the classification method or clustering approach are used to create algorithms in this process [37]. Multiple investigations in the literature have demonstrated the different strategies used by writers to highlight and take into account the key points of feature selection methodologies. Data retrieval and text analysis both greatly benefit from text feature extraction. [18] was the first to establish text mining. It is a method of extracting details from text that is valuable. When data is needed concerning patient records, medical coverage data, networks of friends, and news sources, text mining can be used in data mining and other learning techniques. Moreover, it is employed in image analysis and vision-related applications including the detection and interpretation of license plate numbers [3]. A review of the many text mining tasks and approaches was provided in [34].

In order to circumvent this issue, an approach based on the upper limit criteria is put forth in [38]. It employs a limitation of orthogonal weight matrix and a linear blend of traces inside and outside class scatter in the objective function. However, the demand for a substantial amount of labelled data, which is highly costly to gather in procedure, is a typical restriction shared by all supervised algorithms. However, when there is a lack of labelled training data, these supervised approaches typically perform far worse [39].

Conversely, supervised feature selections make use of both trained and untrained data. precisely therefore, if there is a shortage of trained data, supervised approaches might choose features by using untrained data. This implies the study assume that the majority of data examples fall on a low-dimensional manifold [40]. To utilize the untrained samples, the matrices are utilized in graph techniques. The process of selecting, describing, and identifying speech data from a person or an instrument for use in identification is known as speech recognition. Speech recognition is used in many different applications, including as speaker identification, synthetic speech programming, recognizing emotions, and the classification of different musical genres.

The ability to identify styles of music is an intriguing feature of using speech recognition. Three feature sets representing rhythmic satisfied, timbral appearance, and pitch composition were put forward by the researchers in [41] in order to automatically classify sound signals into their systematic musical genres. Ten musical genres

Corresponding author: Abubakar Atiku Muslim
alatiku@gmail.com
Department of Electrical and Electronic Engineering, Abudllahi Fodio University of Science and Technology Aliero.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

were classified with a 61% accuracy rate using their suggested feature selection method. A frequency-based strategy was put forth in [42]. This method, which was based on an algorithm for onset detection at particular frequencies, was utilized to lower computing complexity.

Recognition of emotions is one of feature selection topics. In recent years, this feature has been applied to robotic study where affective interactions between humans and machines and affective computing are required [34]. The three main steps of recognizing emotions computing are featuring selection, and classification. Common databases are essential for comparing research findings in this field and testing recently created techniques. It has been challenging to assess investigations in this field due to the shortcomings in prevalent datasets. The ability to extract pertinent elements that are independent of the speaker, language, and content is crucial to the efficacy of emotion identification.

However, due to the lengthy graph construction, they are typically less effective when managing big amounts of data [40]. Consequently, research on unsupervised feature selection is essential and crucial. Unsupervised feature selection is thought to be a considerably more difficult task because it lacks the label information that directs the search for discriminative features[11]. A number of criteria have been put forth by numerous scholars to define feature importance. Selecting features that can best maintain the original data's manifold structure is one often utilized criterion. Using methods for clustering to find cluster characteristics and then converting the unsupervised feature selection into a controlled framework is another popular technique.

This strategy can be applied in two ways. Finding cluster indicators, which are regarded as pseudo labels, and carrying out the supervised feature selection concurrently within a single, cohesive framework is one method. Nonnegative frequency cluster and structural learning were combined into a single paradigm in the experiments [43]. lately an integrated search strategy with the benefits of both the filter/extraction and the wrapper methods has been utilized for dimension reduction. Huang et al. presented this method in [44]. Two steps make up this mixed reduction of size method. The first is to find the most pertinent aspects in the data sets, filtering and extraction techniques are employed. The second stage, which is a wrapper method, confirms that the important feature subsets that were previously found are confirmed using a technique that yields greater precise classification rates [21]. In order to increase effectiveness and precise classification with improved computing performance, it employs various evaluation standards in various search levels [45]. Table 3 highlight the advantage and disadvantage of embedded based filter approach.

Table 3: advantage and disadvantage of embedded filter-based approach

Advantage	Disadvantage
Fast: majority of embedded-based models are naturally faster than wrapper approach which rapidly complete the given task without compromising the model performance or prediction process, particularly the rate at which it can deliver predictions after been trained.	Ignoring feature dependency: The potential to overlook feature dependencies, obtain unstable features, and not be able to utilize every feature in a wide feature space are the main obstacles in feature selection.
Interaction with classifier: unlike filter model, embedded model continuously interacts with classifier during and after training.	Redundancy: filter approach may lead to features redundancy which might result in overfitting, redundancy is usually seen as harmful.
Independent: In general, filter-based model proved to be classifier independent indicating that the transition or operation is not dependent on or unique to a certain ML technique.	Interaction with classifier: filter model naturally does not have feedback loop control as opposed to merely its labelling accuracy, is known as the "lack of communication with classifiers" in machine learning.

Corresponding author: Abubakar Atiku Muslim
alatiku@gmail.com
Department of Electrical and Electronic Engineering, Abudllahi Fodio University of Science and Technology Aliero.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

Advantage	Disadvantage
Computational complexity: filter approach offers computational complexity which is essential since it influences the viability and effectiveness of algorithms when used on big datasets.	Discrimination power: The filter approach lack strong discrimination power which has two goals. At first to generate artificial dataset that is almost identical to genuine data.

CONCLUSION

Filter model is far quicker and more efficient in computing. It is technically faster than the Wrapper and Embedded techniques, irrespective of the learning technique, and appropriate for low-dimensional data. However, the relationship among classifiers is not taken into account. The association across the attributes is not taken into account. Inadequately identifying trends throughout the training phase.

Wrapper model takes into account the relationship among the class labels and the features. takes into account the features' interdependencies as well. More precise than the filter approach. However, more complicated to compute. Repeatedly assess the chosen feature subset. When existing features are removed from the beginning, some features might not be taken into account for assessment which might lead to overfitting. Embedded model is more efficient in terms of execution than the Wrapper technique. More precise than the Filter and Wrapper approach. However, embedded model is more expensive to compute than the filter approach. Poor generality and unsuitability for high-dimensional data.

REFERENCES

1. Abiodun, E.O., et al., A systematic review of emerging feature selection optimization methods for optimal text classification: the present state and prospective opportunities. Neural Comput Appl, 2021. **33**(22): p. 15091-15118.
2. Acharya, N. and S. Singh, An IWD-based feature selection method for intrusion detection system. Soft Computing, 2017. **22**(13): p. 4407-4416.
3. Agrawal, P., et al., Metaheuristic Algorithms on Feature Selection: A Survey of One Decade of Research (2009-2019). IEEE Access, 2021. **9**: p. 26766-26791.
4. Albulayhi, K., et al., IoT Intrusion Detection Using Machine Learning with a Novel High Performing Feature Selection Method. Applied Sciences, 2022. **12**(10).
5. Ali, M. and T. Aittokallio, Machine learning and feature selection for drug response prediction in precision oncology applications. Biophys Rev, 2019. **11**(1): p. 31-39.
6. Aljehane, N.O., et al., Optimizing intrusion detection using intelligent feature selection with machine learning model. Alexandria Engineering Journal, 2024. **91**: p. 39-49.
7. Alkahla, L., M. Khalaf Hussein, and A. Alqassab, Feature Selection Techniques in Intrusion Detection: A Comprehensive Review. Iraqi Journal for Computers and Informatics, 2024. **50**(1): p. 46-53.
8. Rahman, M.M., et al., A Review of Machine Learning Methods of Feature Selection and Classification for Autism Spectrum Disorder. Brain Sci, 2020. **10**(12).
9. Rong, M., D. Gong, and X. Gao, Feature Selection and Its Use in Big Data: Challenges, Methods, and Trends. IEEE Access, 2019. **7**: p. 19709-19725.
10. Rudnicki, W.R., M. Wrzesień, and W. Paja, All Relevant Feature Selection Methods and Applications, in Feature Selection for Data and Pattern Recognition. 2015. p. 11-28.
11. Saeys, Y., I. Inza, and P. Larranaga, A review of feature selection techniques in bioinformatics. Bioinformatics, 2007. **23**(19): p. 2507-17.
12. Shyaa, M.A., et al., Evolving cybersecurity frontiers: A comprehensive survey on concept drift and feature dynamics aware machine and deep learning in intrusion detection systems. Engineering Applications of Artificial Intelligence, 2024. **137**.
13. Slimani, K., et al., Feature Selection: A Review and Comparative Study. E3S Web of Conferences, 2022. **351**.

Corresponding author: Abubakar Atiku Muslim
alatiku@gmail.com
Department of Electrical and Electronic Engineering, Abudllahi Fodio University of Science and Technology Aliero.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 14(1), JANUARY, 2026
E-ISSN: 3093-0898, PRINT ISSN: 2277-0011; Journal homepage: www.atbufstejoste.com

14. Ayad, A.G., N.A. Sakr, and N.A. Hikal, A hybrid approach for efficient feature selection in anomaly intrusion detection for IoT networks. The Journal of Supercomputing, 2024. **80**(19): p. 26942-26984.
15. Al-Tashi, Q., et al., Approaches to Multi-Objective Feature Selection: A Systematic Literature Review. IEEE Access, 2020. **8**: p. 125076-125096.
16. Solorio-Fernández, S., J.A. Carrasco-Ochoa, and J.F. Martínez-Trinidad, A review of unsupervised feature selection methods. Artificial Intelligence Review, 2019. **53**(2): p. 907-948.
17. Ang, J.C., et al., Supervised, Unsupervised, and Semi-Supervised Feature Selection: A Review on Gene Selection. IEEE/ACM Trans Comput Biol Bioinform, 2016. **13**(5): p. 971-989.
18. Brezočnik, L., I. Fister, and V. Podgorelec, Swarm Intelligence Algorithms for Feature Selection: A Review. Applied Sciences, 2018. **8**(9).
19. Dhal, P. and C. Azad, A comprehensive survey on feature selection in the various fields of machine learning. Applied Intelligence, 2021. **52**(4): p. 4543-4581.
20. Hira, Z.M. and D.F. Gillies, A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data. Adv Bioinformatics, 2015. **2015**: p. 198363.
21. Deng, X., et al., Feature selection for text classification: A review. Multimedia Tools and Applications, 2018. **78**(3): p. 3797-3816.
22. Zouhri, H., A. Idri, and A. Ratnani, Evaluating the impact of filter-based feature selection in intrusion detection systems. International Journal of Information Security, 2023. **23**(2): p. 759-785.
23. Yan, Z., et al., EXPRS: An extended pagerank method for product feature extraction from online consumer reviews. Information & Management, 2015. **52**(7): p. 850-858.
24. Wang, L., Y. Wang, and Q. Chang, Feature selection methods for big data bioinformatics: A survey from the search perspective. Methods, 2016. **111**: p. 21-31.
25. Bolón-Canedo, V., N. Sánchez-Maroño, and A. Alonso-Betanzos, Feature selection for high-dimensional data. Progress in Artificial Intelligence, 2016. **5**(2): p. 65-75.
26. Gheni, H.Q., et al., Optimizing Feature Selection for Intrusion Detection: A Hybrid Approach Using Cuckoo Search and Particle Swarm Optimization. International Journal of Safety and Security Engineering, 2024. **14**(6): p. 1907-1912.
27. Alsaffar, A.M., M. Nouri-Baygi, and H.M. Zolbanin, Shielding networks: enhancing intrusion detection with hybrid feature selection and stack ensemble learning. Journal of Big Data, 2024. **11**(1).
28. Kuzudisli, C., et al., Review of feature selection approaches based on grouping of features. PeerJ, 2023. **11**: p. e15666.
29. Iqbal, M., et al., Review of feature selection methods for text classification. International Journal of Advanced Computer Research, 2020. **10**(49): p. 138-152.
30. Li, J., et al., A critical review of feature selection methods for machine learning in IoT security. International Journal of Communication Networks and Distributed Systems, 2024. **30**(3): p. 264-312.
31. Nazir, A. and R.A. Khan, A novel combinatorial optimization based feature selection method for network intrusion detection. Computers & Security, 2021. **102**.
32. N, A.S., et al., Enhancing machine learning-based sentiment analysis through feature extraction techniques. PLoS One, 2024. **19**(2): p. e0294968.
33. Kumar, H., et al., A Review on the Classification of Partial Discharges in Medium-Voltage Cables: Detection, Feature Extraction, Artificial Intelligence-Based Classification, and Optimization Techniques. Energies, 2024. **17**(5).
34. Hancer, E., B. Xue, and M. Zhang, A survey on feature selection approaches for clustering. Artificial Intelligence Review, 2020. **53**(6): p. 4519-4545.
35. Almomani, O., A Feature Selection Model for Network Intrusion Detection System Based

Corresponding author: Abubakar Atiku Muslim
alatiku@gmail.com
Department of Electrical and Electronic Engineering, Abudllahi Fodio University of Science and Technology Aliero.

on PSO, GWO, FFA and GA Algorithms. Symmetry, 2020. **12**(6).
36. Subbiah, S.S. and J. Chinnappan, Opportunities and Challenges of Feature Selection Methods for High Dimensional Data: A Review. Ingénierie des systèmes d information, 2021. **26**(1): p. 67-77.
37. Pham, T.H. and B. Raahemi, Bio-Inspired Feature Selection Algorithms With Their Applications: A Systematic Literature Review. IEEE Access, 2023. **11**: p. 43733-43758.
38. Hoque, M.E. and K. Kipli, Deep Learning in Retinal Image Segmentation and Feature Extraction: A Review. International Journal of Online and Biomedical Engineering (IJOE), 2021. **17**(14): p. 103-118.
39. Huang, X., et al., Feature extraction of search product based on multi-feature fusion-oriented to Chinese online reviews. Data Science and Management, 2022. **5**(2): p. 57-65.
40. Logeswari, G., et al., A Comprehensive Approach to Intrusion Detection in IoT Environments Using Hybrid Feature Selection and Multi-Stage Classification Techniques. IEEE Access, 2025. **13**: p. 24970-24987.
41. Lualdi, M. and M. Fasano, Statistical analysis of proteomics data: A review on feature selection. J Proteomics, 2019. **198**: p. 18-26.
42. Kalimuthan, C. and J. Arokia Renjit, Review on intrusion detection using feature selection with machine learning techniques. Materials Today: Proceedings, 2020. **33**: p. 3794-3802.
43. Hasan, M.A.M., et al., Feature Selection for Intrusion Detection Using Random Forest. Journal of Information Security, 2016. **07**(03): p. 129-140.
44. Jaw, E. and X. Wang, Feature Selection and Ensemble-Based Intrusion Detection System: An Efficient and Comprehensive Approach. Symmetry, 2021. **13**(10).
45. Kasongo, S.M. and Y. Sun, Performance Analysis of Intrusion Detection Systems Using a Feature Selection Method on the UNSW-NB15 Dataset. Journal of Big Data, 2020. **7**(1).

Corresponding author: Abubakar Atiku Muslim
alatiku@gmail.com
Department of Electrical and Electronic Engineering, Abudllahi Fodio University of Science and Technology Aliero.