Feature Selection Is Important: State-of-the-Art Methods and Application Domains of Feature Selection on High-Dimensional Data

G. Manikandan and S. Abirami

1 Introduction

Since high-dimensional data could provide ample information, it is hard to establish a precise prediction model along with the growing dimensionality and scale of dataset. High-dimensional data are the data that have more number of variables or attributes. With recent advances in data acquisition and storage technology, high-dimensional data extensively occur in nature, finance, industry, biomedicine, and several other fields, which contain complex nonlinear relationship among multiple features [1, 2]. There are three categories of attributes in prediction model: relevant, redundant, and irrelevant. The relevant feature is highly correlated with the target, whereas redundant features are correlated with each other; in case of irrelevant feature, they do not have any significant information on target.

In order to select the relevant features by reducing the redundant and irrelevant feature selection provides the way with increased accuracy. Feature selection varies with supervised learning, unsupervised learning, and semi-supervised learning. In supervised learning feature selection, it uses the label/class data to select the features, that is, it selects the feature based on class information. In unsupervised learning feature selection, it selects the feature without class information, based on the feature–feature relation that eliminates irrelevant features. The main demerits of this method are that it ignores the correlation between the features and class information and also sometimes, it ignores the correlation between features [3].

G. Manikandan (✉)
Department of Computer Science and Engineering, College of Engineering Guindy, Anna University, Chennai, Tamil Nadu, India

S. Abirami
Department of Information Science and Technology, College of Engineering Guindy, Anna University, Chennai, Tamil Nadu, India

R. Kumar, S. Paiva (eds.), *Applications in Ubiquitous Computing*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-030-35280-6_9

Semi-supervised feature selection is the combination of supervised and unsupervised feature selections; it uses and selects the feature with and without class information.

The feature selection process consists of three main stages: in the first stage, it identifies the search direction by taking any one of the randomly selected features or it may be an empty set. After that, it selects a new feature and adds into the feature set with “n” number of iterations till the final optimal set [2]. This type-selecting feature from the empty set to the optimal set is called forward search. In addition to this, the process of selecting feature from full set and by eliminating the irrelevant feature in each iteration in order to get the final subset is called backward elimination. Sometimes, the search direction performs both forward selection and backward elimination methods to select the predominant features in the each of the iteration, which is called bidirectional search method. Based on the properties of the dataset, the search direction has to be selected by the researchers to perform better feature selection. The second stage of the feature selection process is to identify the search strategies. There are three main strategies: exponential, random search, and sequential search. Determining the evaluation criteria is the last stage of feature selection. Filter, wrapper, hybrid, and embedded are the best evaluation criteria of the feature selection process [4].

In the medical domain, gene expression may have more than thousands of genes among the vast amounts of genes and many of the genes will be completely irrelevant for classifying the purpose and also it leads to overfitting. Due to the presence of overfitting, it needs more computational time and processing power to classify the data with high accuracy. Most of gene expression data contain irrelevant, redundant information; here, the feature selection plays an important role to select the highly informative genes from a vast number of genes in order to reduce the computational cost with a high classification accuracy. These types of microarray data analysis help to find out deadly diseases like cancer early. Mostly humans may suffer 200 types of cancer; for this, microarray analysis [3, 4] was used to detect the cancers early. Breast cancer, colon cancer, leukemia, prostate cancer, and lung cancer are some types of cancer which the human beings encounter.

Dimensionality reduction can be done by two ways, namely feature selection and feature extraction, where feature selection selects the predominant feature from the vast data without modification of the originality of the data, that is, it is a subset of the whole data, whereas the feature extraction converts the original data into some form based on the application of linear or nonlinear transformation to reduce the data [3]. The static feature selection (FS) and streaming feature selection are the two categories of FS; the original data do not change over time, whereas the new feature may be added to original data in streaming feature selection. In this chapter, we have focused on only feature selection methods such as filter, wrapper, embedded, ensemble, and hybrid methods which are shown in Fig. 1.

In this chapter, Sect. 1 describes the introduction to feature selection, process of feature selection, and objectives of feature selection, whereas Sect. 2 summarizes the various related works with respect to filter, wrapper, embedded, and hybrid methods. Section 3 provides various application fields such as microarray analysis,

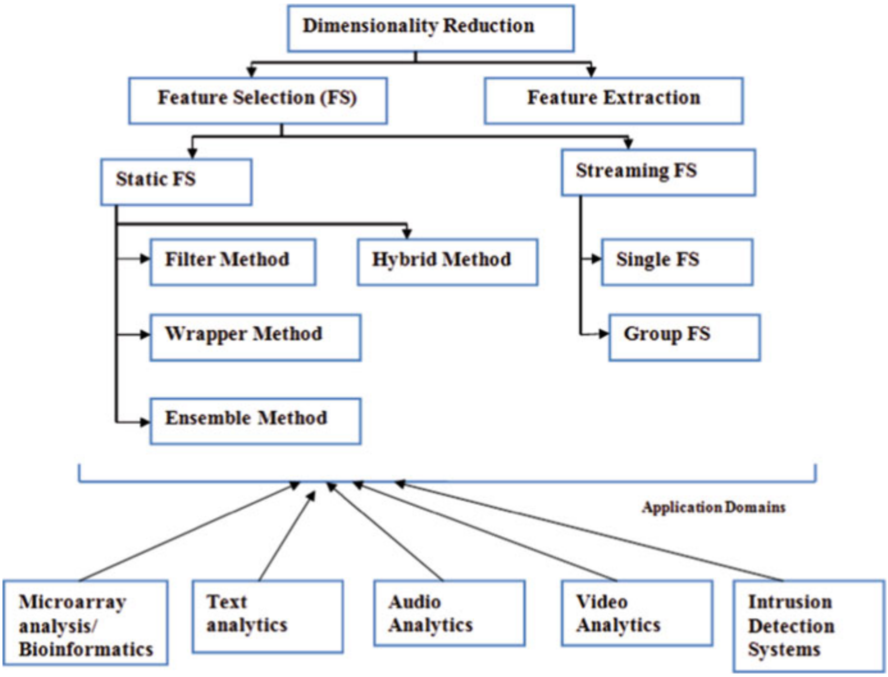

Fig. 1 Feature selection methods and their applications

text analysis, video analytics, audio analytics, and intrusion detection systems; discussion among various feature selection methods is given in Sect. 4. Section 5 concludes the chapter with the issues relating to selecting the optimal feature subset and future directions in dimensionality reduction.

2 Related Works

2.1 Filter Methods

In filter methods, there are various ranking approaches used as the principle criteria for feature selection. Ranking methods assign ranks to each feature based on some intrinsic and statistical properties, and an appropriate ranking method is selected and used to rank the features; after this, based on the threshold value, the feature can be selected. Here, the threshold value is selected by applying their own proposed algorithms or by using trial-and-error method. Filter methods select features based on the internal characteristics of the features. They do not involve any classifiers to select the features. Table 1 presents the recent important techniques based on the filter method.

Table 1 Recent techniques based on filter method

Ref. no.	Dataset	Algorithm	Classifier	Accuracy	Year
[5]	Colon, Central Nervous System (CNS), GLI_85, SMK_CAN_187	Whale Algorithm, Mutual Congestion, Forward Feature Selection Method	SVM, NB, DT	80%	2019
[6]	Emails	Information gain, Latent Dirichlet Allocation, Topic Guessing	NB, SVM, C4.5, Adaboost C4.5, Bagging C4.5, Random Forest, Logistic Regression, Rough Sets	88%	2019
[7]	Ionosphere, Breast Cancer Wisconsin (BCW), Connectionist Bench, Iris, Statlog (Vehicle Silhouettes), Parkinson	Modified-BPSO	K-means Clusters	–	2019
[8]	NIST and RIMES Databases	Symmetrical Uncertainty, Chi-Square, Relief, Information gain, Gain Ratio, Correlation-based Feature Selection, Consistency Criterion	K-NN, Bagging and Random Forest	72.66%	2019
[9]	Simulated and Industrial datasets	Wide Spectrum feature selection for regression (WiSe)	Forward Stepwise regression (FSR), Selector operator (LASSO), Partial Least Squares (PLS) and Least Absolute Shrinkage	–	2019

2.2 *Wrapper Method*

The wrapper method selects the features with the help of machine learning algorithms by knowing the classification accuracy and error rate. This method aims to minimize the classification error and to improve the classification performance. In general, it gives better accuracy than the filter method because it tunes the model and selects the feature based on the learning algorithm. The main disadvantages of this method are that it consumes more computational time, and also the classifier performance may vary across the different classifiers. Mostly, SVM, Naïve Bayes,

Table 2 Recent techniques based on wrapper method

Ref. no.	Dataset	Algorithm	Classifier	Accuracy	Year
[10]	Colon, Central Nervous System (CNS), GLI_85, SMK_CAN_187	Binary Gross hopper Optimization algorithm	k-Nearest Neighbor	77.04%	2019
[6]	Emails	Multi objective Evolutionary Algorithm	Linear Discriminant Analysis, k-Nearest Neighbor	89.9%	2019
[11]	Breast cancer, BreastEW, Exactly, Exactly2, HeartEW, Lymphography, M-of-n, penglungEW, SonarEW, SpectEW, CongressEW and IonosphereEW datasets	Las Vegas Wrapper	Ensemble learner	97.04%	2019
[12]	Real Motory image dataset	Wrapper-based selection	J48, PART, Adaboost, Random Forest, Naïve Bayes	76.33%	2019
[13]	Kidney disease dataset	Chaotic crow search algorithm (CCSA)	k-Nearest Neighbor	83.8%	2019

and Random Forest algorithms are used as classifiers in the wrapper method. Some of the recent wrapper-based methods and applied datasets with accuracy are presented in Table 2.

2.3 Hybrid Method

The hybrid feature selection method performs the feature selection process by merging or joining two different feature selection methods. For example, we can merge the filter method and wrapper method into one and perform the selection process, but the evaluation criterion has to be the same for the two methods. In the hybrid method, it uses the merits of the two methods for selecting the features and combines the selected features into a single subset. Sometimes, this method uses different evaluation criteria and search strategies for gaining accuracy with less computational time. For example, the filter method is used for selecting the initial

Table 3 Recent techniques based on hybrid method

S. no.	Dataset	Algorithm	Classifier	Accuracy (approx)	Year
[14]	Medical datasets	Roughset, BGWO, BCGWO	K-nearest neighbor	72.6%	2019
[15]	Phishing dataset	Hybrid Ensemble Feature Selection, Cumulative Distribution Function Gradient	Naïve Bayes Random Forest, SVM, C4.5, JRip	94.6%	2019
[16]	Breast Cancer, Zoo, Tic-tac-toe, Vote, Waveform, Wine, CongressEW, Lymphography	BGWOPSO	K-nearest neighbor	93%	2019
[17]	Ionosphere, Automobile, BreastDiagnostic, BreastPrognostic, German, SPECTF Heart HillValley, Ozone Level, Parkinsons and Sonar	MPMDIWOA-Maximum Pearson MaximumDistance Improved Whale Optimization Algorithm	SVM	90.4%	2019
[18]	Irvine repository datasets	Grey Wolf Optimization and Crow Search Algorithm	K-nearest neighbor	90.61%	2019

features by assigning weights; later, the same features will be given to the wrapper method in order to get optimized results. Some of the recent hybrid-based methods and applied datasets with accuracy are presented in Table 3.

2.4 *Embedded Method*

Embedded method uses the machine algorithms to select the features based on internal optimization of the features. Here, the features are selected based on the predefined function to evaluate the features. Many research studies reveal that the embedded methods are more convenient to select the features with less computational cost. This is because this method does not need continuous evaluation of the feature repetitively and is less prone to overfitting. Ensemble feature selection follows the idea of ensemble learning which uses the aggregated results of various learners. Ensemble feature selection aims to find stable and robust feature subset. In order to find a stable subset, this method employs different types of feature selectors.

Table 4 Recent techniques based on embedded method

S. no.	Dataset	Algorithm	Classifier	Accuracy (approx)	Year
[19]	Mushroom, Thyroid, Diabetes, Liver, Breast Cancer, Heart, CKE, Dermatology, Ionosphere, Tumour data, Audiology, Lymphography, Zoo	Wrapper Method, Bagging	SVM, RF, NB	92%	2019
[20]	Diabetic Electronic Medical Records	GBM – Gradient Boosting Machine with mean rank	–	82%	2019

Finally, results from various feature selectors are aggregated to produce the desired feature subset. Some of the recently proposed embedded-based methods and applied datasets with accuracy are presented in Table 4.

2.5 *Feature Selection Based on Fuzzy Logic*

Currently, researchers concentrate on computational intelligence-based techniques such as fuzzy rough set-based attribute selection, fuzzy support vector machines (FSVM), high-performance feature selection, fuzzy feature selection, supervised neural networks, unsupervised neural networks, etc., for selecting the best features. Despite feature selection, feature relevance, feature redundancy, and anomaly detection are other issues which are present in the classification of high-dimensional dataset. Apart from this, feature grouping or feature clustering is one of the approaches which are being currently used in most of the application domains to reduce the dimensionality of feature before classification. In order to achieve the abovementioned, we have presented some of the recent fuzzy logic feature selection techniques in Table 5.

2.6 *Comparison of Feature Selection Methods*

Filter methods select the features without the help of classifiers; hence, it is faster and gives considerable accuracy. Compared to the filter method, it selects the feature with the interaction of the classifier and also provides higher accuracy than filter because of tuning the parameters with the classifiers, but it takes more computation

Table 5 Recent techniques based on fuzzy logic

S. no.	Dataset	Algorithm	Classifier	Accuracy (approx)	Year
[21]	Flash Flood Dataset	FURIA-GA	C4.5 Decision Trees, JRip	89.03%	2019
[22]	Cancer Dataset	TCGA	Fuzzy Rule-based classification	90%	2019
[23]	ALL AGENTS and INBOUND AGENTS	MultiObjective-EvolutionarySearch method with the multi-objective evolutionary algorithm ENORA	Fuzzy Rule-based classification	73.25%	2019
[24]	Diabetes dataset	Fuzzy principal component analysis (FPCA)	FPCA-SVM	71%	2019
[25]	39 Bus drivers data	Adaptive Neuro-Fuzzy Inference System (ANFIS), Particle Swarm Optimization (PSO)	SVM	98.12%	2019

time. Embedded methods select the feature with classifier and give better accuracy compared to the wrapper method, but it does not consider the overfitting problem. The hybrid method is the combination of feature selection methods and it gives better accuracy than the filter method and also gives better computation cost when compared to the wrapper method.

3 Importance of Feature Selection and Application Domains

This section provides the importance of feature selection in various application domains such as text categorization, video analytics, audio analytics, microarray data analysis, bioinformatics, instruction detection techniques, and streaming data analysis. Also, this section discusses the various recent feature selection techniques with respective application domains.

3.1 Importance of Feature Selection in Text Categorization

Due to the fast growth of the text-based content on WWW, it has led to the problem of categorizing the content based on a particular context. In text mining, identifying the common pattern among the text documents is a difficult task because most of the text data contain irrelevant and redundant data. The application domain includes sentiment analysis, bioinformatics, movie/product recommendation, web spam detection, clinical data analysis, etc. In particular, dimensionality reduction by taking into consideration both feature extraction and feature selection plays an important role in text categorization. Feature selection selects the feature subset from the vast amount of dataset without modifying the original data, whereas feature extraction transforms or combines the features based on certain models for reducing and selecting the predominant features. Alternatively, the feature extraction technique is used to extract the core feature of the text data, whereas feature selection is used to select the predominant features. Many statistical and machine learning-based text categorization techniques have been proposed by researchers in order to extract the core features from the vast amount of text data. However, there are problems and challenges that exist still: to select the feature and in classifying the text document automatically.

Authors from the study cited herein [26] mentioned two major problems which are currently facing the researchers in the field of text categorization: the first one is huge numbers of dimensionality in the data and the second one is the presence of noisy features in the text document while creating the vector space model through the Bag of Words (BoW) approach. Due to the existence of these problems, the computational complexity becomes high and also leads to problems in the classification accuracy. As mentioned in the previous section, the feature selection method can be classified into filter, wrapper, embedded, hybrid, and ensemble approaches.

Multivariate relative discriminative criterion (MRDC) is one of the effective multivariate filter-based feature selection techniques which has been proposed to select the predominant features from the text data [27]. MRDC is mainly designed for effective feature selection and text classification. This technique consists of three main steps such as preprocessing, feature selection, and evaluation. Preprocessing is common in data mining; the main objective is to perform stop-word removal, stemming, pruning, and term-weighting. As a second step, the authors addressed and proposed the feature selection technique to select the informative feature from the text data with consideration of the two principles' relevancy by employing relative discriminative criterion (RDC) and redundancy by using Pearson correlation. Finally, the subset of the features is selected from the whole feature set. WebKB, Reuters-21578, and 20-Newsgroup datasets are used for evaluating their proposed algorithm, and results are evaluated in terms of precision, recall, and F-Measure. Decision tree and multinomial naïve Bayes and multilayer perceptron were used as the classifier algorithms. Friedman test has been used to prove that their obtained results are statistically significant.

Feature selection in text is classified into four categories: syntactic, semantic, stylistic, and information gain-based methods. The syntactic model focuses on selecting the model automatically by learning the features in the document by separating the subjective expressions from the polarities. This model focuses only on the subjective-based expressions by ignoring the irrelevant features. The semantic-based approach represents the document as the collection of words, where the sentiment of the each word can be predicted based on the linguistic features such as noun, adjectives, and verbs. Stylistic-based text analysis approach focuses on the various semantic functions of the words and phrases based on the usefulness of the features. Feature selection based on information gain selects the features based on the entropy values between the features/variables and class.

3.2 Importance of Feature Selection in Speech Recognition

Speech recognition plays an important role in many application areas such as the medical domain for detecting stress and pain, robot interactions, computer games, cyber forensics, and call centers for predicting the speaker emotions. These emotions can be predicted by applying various pattern recognitions, machine learning, and artificial intelligence algorithms by finding the patterns and classification. In general, the emotion in the speeches varies among different speakers, but it is important to analyze the emotions for classification. Speech emotion recognition is one of the techniques which extract the emotions in the speech signal. Feature extraction, feature selection, and emotion recognitions are the basic steps involved in the speech recognition. Linear Predictor Coefficients (LPC), Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictor Cepstral Coefficients (LPCC) significantly contribute to emotion recognition and are some of the important feature selection methods applied in speech recognition which are mentioned in Kasiprasad Mannepalli et al. [28].

Currently, researchers focus on finding the number of speech features that are used to categorize the emotional content of the speech. For selecting the features, many feature selection algorithms have been developed and used to select the features, as well as to improve the classification performance. Some of the feature selection methods which are used to select the important features are as follows:

(a) Fast correlation-based filter
(b) Forward feature selection (FFS) and backward feature selection (BFS)
(c) Wrapper-based feature selection
(d) Fuzzy-based feature selection
(e) Sequential floating forward selection (SFFS)
(f) Principal component analysis
(g) Least squares (LS) bound
(h) Mutual information (MUTINF)
(i) Minimum redundancy maximum relevance (mRMR)

ReliefF, Symmetrical uncertainty, Fisher score, spectral feature selection (SPEC), Laplacian score, sparse, local feature selection based on scatter separability (LFSBSS), multi-cluster-based feature selection (MCFS), relief, inconsistency criterion, clustering-based feature selection, and ReliefC.

Authors of the study cited herein [29] proposed a novel feature selection technique for speech emotion recognition based on observing the changes in feature subset according to particular emotions. In their experiment analysis, four

Table 6 Feature selection algorithms on speech recognition

Ref. no.	Dataset	Feature selection algorithm	Classifiers used	Accuracy	Year
[29]	EMO-DB, eNTERFACE05, EMOVO SAVEE	A novel feature selection method for speech emotion recognition	SVM, k-NN, MLP	42.60–84.07%	2019
[30]	Tollywood and Bollywood Popular Songs (TBPS).	GAFS – Genetic algorithm-based feature selection	SVM – support vector machines, ANN – Artificial neural networks, RF – random forest	74.69–91.58%	2018
[31]	EMO-DB, RML, eNTERFACE05, and BAUM-1s	DTPM – Discriminant Temporal Pyramid Matching	Deep convolutional neural networks (DCNN) with five layers	EMO-DB (87.31–86.30%) RML (75.34–75.20%) eNTERFACE05 (79.25–79.40%) BAUM-1s (44.61–44.03%)	2019
[32]	eNTERFACE	Emotion recognition using deep learning approach	RBF kernel in the SVM Polynomial kernel in the SVM	99.9%	2019
[33]	SAVEE, Emo-DB, MES and DES	Salient discriminative feature analysis (SDFA)	CNN	71.8% (SAVEE) 57.2% (Emo-DB) 60.4% (DES) 57.8% (MES)	2014

datasets, namely EMO-DB, eNTERFACE05, EMOVO, and SAVEE, were used for the analysis features and for classification, support vector machine, multilayer perceptron, and k-NN classifiers were used. And also, the proposed feature selection methods are compared with the standard methods such as principal component analysis, fast correlation-based feature selection, and sequential forward selection.

Some of the recent speech emotion recognition methods and applied datasets with accuracy are presented in Table 6.

3.3 Importance of Feature Selection in Video Processing

In video analytics area, the feature selection is used in selecting the important features. The application of the video analytics includes surveillance of public data to predict the crime, face recognition, gait recognition, human action recognition, etc. Object tracking and recognition of the robots is one of the trending areas and also vehicle detection from the traffic video avoids the crimes and helps safe driving. Feature selection with Joint $l_{2,1}$*-norm* minimization (FSNM), Minimum-Redundancy–Maximum-Relevance (MRMR), Fisher Score, Fast Correlation-Based Filter (FCBF), $l_{2,1}$*-norm* Manifold (L21), SParse Multinomial Logistic Regression (SBMLR) via Bayesian L1 Regularization [34] are some the standard existing techniques which are used to compare the proposed feature selection technique which is currently being developed.

Mobile video streaming QoE prediction is one the research areas in video analysis, particularly in the video providing services such as YouTube and Netflix. The main aim of the providers is to provide high quality and at considerable operational cost, but there is a trade-off between these two. In general, the prediction of the QoE has been divided into two types, namely continuous time QoE and retrospective QoE. In retrospective QoE, a single score provided by the subjects will describe the overall QoE for the entire video, whereas continuous-time QoE provides the real-time measurement of each subject with the current QoE, which may lead to trigger the current quality of the video. HAS, which is called the HTTP-based adaptive video streaming, has been proposed [35] which selects the features based on the video quality features, number of staling event, etc., and the experimental analysis is conducted on LIVE-Netflix DB and Waterloo DB.

Some of the recent video processing methods and applied datasets with accuracy are presented in Table 7.

3.4 Importance of Feature Selection in Intrusion Detection Systems

Due to recent advancement in network-based technology, the threat of spammers, criminals, and attackers has also been increasing. The total annual financial loss which is caused due to network intrusion was about US$130 million in 2005; now, it will be more than thrice. Intrusion detection is the technique for detecting the intruder's attacks on the networks; these attacks can be detected by signature-based misuse detection or anomaly-based detection. In the misuse-based detection system, the patterns of attacks are already stored in the network data and database, and if the data are matched with the database, then it is declared as attack. Anomaly detection creates the profile afterward; it analyzes and observes the behavior of the network.

One of the issues in intrusion detection system is accuracy of the classification, since most of the datasets are imbalanced. In this case, feature selection algorithms

Table 7 Feature selection algorithms on video processing

Ref. no.	Dataset	Feature selection algorithm	Classifiers used	Accuracy	Year
[36]	MSR Action3D, MSRDailyActivity3D, Online RGBD Action	Novel method for feature selection based on a Markov blanket combined with the wrapper method	HMM (Hidden Markov Model), DBN (Dynamic Bayesian Network)	91.80% MSRDaily-Activity3D 94.17% Online RGBD Action 97.95% Chalearn LAP 2014 95.23%	2017
[34]	Youtube Kodak HumanEva MIR FLICKR COIL-20 COREL-50	GLocal Structural feature selection with Sparsity (GLSS)	SVM Adaboost KNN	Accuracy for SVM(*NoFS/ Proposed*) Youtube 38.2%/34.3% Kodak 49.3%/45.9% HumanEva 95.7%/83.4% MIR FLICKR 52.7%/46.1% COIL-20 82.2%/68.7% COREL- 44.3%/38.2%	2014
[37]	TRECVID 2012 Open videos (OV) YouTube videos	Video Semantic Analysis-based Kernel Locality-Sensitive Discriminative Sparse Representation (KLSDSR)	–	Recognition Rate TRECVID 2012 – 91.20% Open videos (OV) – 89.20% YouTube videos – 90.17%	2019

are used to select the predominant features in order to classify and improve the efficiency of the algorithm. Filter, wrapper, and embedded-based feature selection techniques were used to reduce and eliminate the number of features which were explained in the previous section. The main types of attacks are DoS attack, replay attacks, selective forwarding attack, Sybil attack, Sinkhole attack, Wormhole attack, black hole attack, Jamming attack, false data attack, etc.

KDD Cup 1991 dataset is one the well-known datasets for the intrusion detection system which consists of five million records, and each record consists of 41 nominal and continuous features with the class label (Normal, DOS, Probe, U2R, R2L) which are available in http://kdd.ics.uci.edu/.

The authors from the study cited herein [38] proposed intelligent fuzzy rough set-based feature selection algorithm and temporal classification for intrusion detection system in WSNs for selecting the important attributes in order to predict the attacks and also a Fuzzy Rough set-based Nearest Neighborhood technique (FRNN) is developed for the effective classification of the multiclass data. Their proposed FRNN approach gives better detection accuracy of about 99.87%.

Correlation and interact feature selection (based on symmetrical uncertainty), Random Forest-Backward Elimination Ranking (RF-BER) FS and Random Forest-Forward Selection Ranking RFFSR), Markova blanket model and decision tree analysis in feature selection, NPGA algorithm, NSGA, NGSA-II, GHSOM-pr, Fuzzy Enhanced Support Vector Intrusion Detection model (Fuzzy ESVDF) using Fusion of chi-square feature selection, Latent Dirichlet Allocation (LDA) and genetic algorithm are some of feature selection algorithms which are used in intrusion detection systems.

3.5 Importance of Feature Selection in Microarray Data Analysis

During the last few decades, the advancement in DNA microarray data analysis has created a direction of the research in machine learning, statistical analysis, and bioinformatics. Generally, these types of DNA microarray data are collected from the tissue samples of the persons and based on the differences in the gene expressions, the person will be distinguished by specific tumors. Particularly, these microarray medical datasets consist of tens and thousands of features with less number of instances since with small sample size with large number of features it suffers with classification accuracy and computational time.

Prediction and classification of the cancer-infected genes and normal healthy genes from the microarray data are always a problem in modern society. Most important thing in the prediction is how well the generated or proposed algorithm differentiates those genes because gene dataset usually consists of a lot of noise values and more number of features. From those features, not all the features are useful in classifying the gene as cancer or regular genes. So, an efficient feature selection algorithm is needed to select and extract only informative genes from the vast number of other genes. For solving this issue, many feature selection techniques such as filter, embedded, wrapper, and hybrid methods have been proposed to select the subset of informative features from the high-dimensional dataset.

Feature selection is the preprocessing step to overcome the issues of selecting features from the microarray data. One of the applications of the microarray data analysis is to predict the cancer early since many types of cancer are caused due to the epidemic or genetic changes. Microarray data analysis is one of the established standard tools which are used to identify and analyze the gene data. One of the main functionalities of the microarray data analysis is to monitor the gene expression level from the genome scale, and after the process of genome scale, the experiment results form a matrix called gene expression matrix. The gene expression matrix consists of genes along with persons, where each row represents the persons' sample instances and each column represents the gene values.

Some of the important feature selection algorithms which are currently being used in the field of the microarray data analysis [39] are the following: new robust feature selection method, entropic filtering algorithm (EFA), MASSIVE, maximum weight and minimum redundancy (MWMR), minimum redundancy maximum relevance (mRMR), INTERACT, Information Gain, ReliefF, Correlation-based Feature Selection, Fast Correlation-Based Filter, new hybrid filter-based FS based on the combination of clustering and modified Binary Ant System (FSCBAS), particle swarm optimization, ant colony optimization, genetic algorithm, artificial bee colony, unsupervised feature selection approach based on ACO (UFACO), relevance–redundancy FS using ACO, a binary ant colony optimization (BACO), novel hybrid feature selection called R-m-GA, Multi-Filter Multi-Wrapper (MFMR), SVM-RFE, iterative perturbation method (IFP), First Order Inductive Learner, rule-based feature subset selection algorithm, kernel penalized SVM (KP-SVM), Adaptive Genetic Algorithm, and Mutual Information Maximization (MIM).

3.6 Importance of Feature Selection in Streaming Data Analysis

Streaming feature selection is an emerging research area which is being currently focused on by researchers to reduce the features and select the most informative features. In streaming feature selection, the candidate features arrive in a sequential manner and also the size of the features will be unknown. This type of streaming feature selection has been used in many application areas such as weather forecasting, stock market prediction, and clinical record analysis. The streaming features are defined as features which flow one by one based on the time variations, but the number of instances is fixed. Because features flow one by one, decisions have to be made whether the feature has to be kept or discarded. Noura AlNuaim et al. [40] enumerated the differentiation between the streaming data and the streaming features. In streaming data, the number of features is fixed and the instances of the streaming data will be generated automatically over time, and hence, the size of instances is unknown. In streaming features, the number of instances is fixed where the number of features will be changed over the time.

Peng Zhou et al. [41] proposed a novel neighborhood rough set-based feature selection with adapted neighbors called gap relation and new online feature selection method called OFS-A3M. The main novelty of this paper is that the proposed OFS-A3M does not require any of the domain knowledge in advance. There are three evaluation criteria, namely maximal-relevance, maximal-dependency, and maximal significance that were considered and used to select the optimal features

Table 8 Feature selection algorithms on video processing

Ref. no.	Dataset	Feature selection algorithm	Classifiers used	Accuracy	Year
[42]	LUNG2, IONOSPHERE, ARCENE WDBC, SRBCT, LYMPHOMA, SONAR, HILL, COLON, GLIOMA, MLL, PROSTATE, DLBCL, LEU,	OFS-Density based on neighborhood rough set	KNN, SVM, and CART	Average KNN – 85.56% SVM – 84.18% CART – 80.80%	2019
[43]	ALLAML, GLIOMA, Prostate GE, Breast, SRBCT, CNS + 20 datasets	OSFSMI and OSFSMI-k	Naïve Bayes KNN Decision Tree	Naïve Bayes – 58.64% KNN – 68.97% Decision Tree – 64.09% Average	2018
[39]	Dorothea, arcene, dexter, and madelon, nova, sylva, and hiva, arrhythmia and mf, tm1, tm2, and tm3	OS-NRRSAR-SA	J48, SVM, Naive Bayes	Naive Bayes – 46.03–98.90% SVM – 67.32–67.32% J48 – 51.72–98.20%	2016

and also, it selects the predominant features based on high dependency, high correlation, and low redundancy. In their studies, fifteen different types of datasets were used to compare the results of the proposed algorithm. Some the standard streaming feature selection algorithms are as follows: information-investing and alpha-investing based on streamwise regression for online feature selection, online streaming feature selection framework with two algorithms called fast Online Streaming Feature Selection (OSFS) and OSFS, streamwise feature selection from the Rough Set perspective, OS-NRRSARASA- Rough Set-based method for online streaming feature selection, OSFSMI and OSFSMI-k, novel Neighborhood Rough Set classifier (NRSC), Scalable and Accurate Online FS Approach (SAOLA), Grafting algorithm based on a stagewise gradient descent approach for online feature selection, Drift Detection Method and the Early Drift Detection Method, and ADaptive sliding WINdow (ADWIN).

Some of the recent video processing methods and applied datasets with accuracy are presented in Table 8.

4 Discussion

When concerning with the high-dimensional spaces of data, the learning phase suffers to observe or conceive the features since all the features do not have the same amount of discriminative power and consist of both redundant and irrelevant data. In this context, filter methods, wrapper methods, and embedded methods are optimally used to select the features with many learning algorithms and contribute in different domains. Filter-based methods extract features in preprocessing step so as to estimate each subset and properties of data. The main merit of this technique is that it is computationally fast and scalable to large-dimensional data and the disadvantage of this technique is that it has no connection with classifier to achieve better performance. The wrapper-based method uses the learning model as the black box to assess the features based on predictive ability. The advantage of this method is that it interacts with the feature subset as well as the learning model well, but the disadvantage is the existence of larger risk of overfitting problem. Embedded methods use both feature selection model and learning model for the purpose of classification, and also it produces good performance results. The advantage of this method is that it interacts with the feature subset as well as the learning model well, but the disadvantage is the existence of larger risk of overfitting problem. Embedded methods use both feature selection model and learning model for the purpose of classification, and also it produces good performance results.

As mentioned earlier, feature selection is used in various domains. In the medical domain, it plays a major role in predicting cancers in an early stage based on the epidemic changes of the genes by selecting the highly informative genes in a timely manner. In the perspective of analytics in video, it is used in predicting the crime, gait recognition, etc., and also in text mining, it is used in categorizing the texts in various domains and aspects. Feature selection plays an important role in streaming data selection for predicting stock marketing, weather forecasting, etc., and also used in the intrusion detection system for detecting intruders in the networks.

5 Conclusion

In this chapter, we have provided a detailed introduction to feature selection with state-of-the-art feature selection techniques based on filter, wrapper, embedded, and hybrid models. Moreover, we have provided the taxonomy of the dimensionality reduction techniques and fuzzy logic-based feature selection techniques in a detailed manner. Further, we have discussed the importance of feature selection among various application domains such as text analytics, video analytics, audio analytics, microarray data analysis, intrusion detection systems, and feature selection in stream data analysis.

References

1. Bolon-Canedo, V., Sanchez-Marono, N., Alonso-Betanzos, A., Benitez, J. M., & Herrera, F. (2014). A review of microarray datasets and applied feature selection methods. *Information Sciences, 282*, 111–135.
2. Wang, H., Tan, L., & Niu, B. (2019). Feature selection for classification of microarray gene expression cancers using bacterial colony optimization with multi-dimensional population. *Swarm and Evolutionary Computation, 48*, 172–181.
3. Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A., Benítez, J. M., & Herrera, F. (2019). A review of microarray datasets and applied feature selection methods. *Information Sciences, 282*, 111–135.
4. Ang, J. C., Mirzal, A., Haron, H., & Hamed, H. N. A. (2019). Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 13*, 971–989.
5. Nematzadeh, H., Enayatifar, R., Mahmud, M., & Akbari, E. (2019, January 17). Frequency based feature selection method using whale algorithm. *Genomics, 111*, 1946–1955.
6. González, J., Ortega, J., Damas, M., Martín-Smith, P., & Gan, J. Q. (2019). A new multi-objective wrapper method for feature selection–accuracy and stability analysis for BCI. *Neurocomputing, 333*, 407–418.
7. Kumar, L., & Bharti, K. K. (2019). An improved BPSO algorithm for feature selection. In *Recent trends in communication, computing, and electronics* (pp. 505–513). Singapore: Springer.
8. Cilia, N. D., De Stefano, C., Fontanella, F., & di Freca, A. S. (2019). A ranking-based feature selection approach for handwritten character recognition. *Pattern Recognition Letters, 121*, 77–86.
9. Rendall, R., Castillo, I., Schmidt, A., Chin, S. T., Chiang, L. H., & Reis, M. (2019). Wide spectrum feature selection (WiSe) for regression model building. *Computers & Chemical Engineering, 121*, 99–110.
10. Mafarja, M., Aljarah, I., Faris, H., Hammouri, A. I., Ala'M, A. Z., & Mirjalili, S. (2019). Binary grasshopper optimisation algorithm approaches for feature selection problems. *Expert Systems with Applications, 117*, 267–286.
11. Xiong, C. Z., Su, M., Jiang, Z., & Jiang, W. (2019). Prediction of hemodialysis timing based on LVW feature selection and ensemble learning. *Journal of Medical Systems, 43*(1), 18.
12. Singh, A., & Jain, A. (2019). Adaptive credit card fraud detection techniques based on feature selection method. In *Advances in computer communication and computational sciences* (pp. 167–178). Singapore: Springer.
13. Sayed, G. I., Hassanien, A. E., & Azar, A. T. (2019). Feature selection via a novel chaotic crow search algorithm. *Neural Computing and Applications, 31*(1), 171–188.
14. Anter, A. M., Azar, A. T., & Fouad, K. M. (2019, March). Intelligent hybrid approach for feature selection. In *International conference on Advanced Machine Learning Technologies and Applications* (pp. 71–79). Cham: Springer.
15. Chiew, K. L., Tan, C. L., Wong, K., Yong, K. S., & Tiong, W. K. (2019). A new hybrid ensemble feature selection framework for machine learning-based phishing detection system. *Information Sciences, 484*, 153–166.
16. Al-Tashi, Q., Kadir, S. J. A., Rais, H. M., Mirjalili, S., & Alhussian, H. (2019). Binary optimization using hybrid grey wolf optimization for feature selection. *IEEE Access, 7*, 39496–39508.
17. Zheng, Y., Li, Y., Wang, G., Chen, Y., Xu, Q., Fan, J., & Cui, X. (2019). A novel hybrid algorithm for feature selection based on whale optimization algorithm. *IEEE Access, 7*, 14908–14923.
18. Arora, S., Singh, H., Sharma, M., Sharma, S., & Anand, P. (2019). A new hybrid algorithm based on grey wolf optimization and crow search algorithm for unconstrained function optimization and feature selection. *IEEE Access, 7*, 26343–26361.

19. Mohan, C., & Nagarajan, S. (2019). An improved tree model based on ensemble feature selection for classification. *Turkish Journal of Electrical Engineering and Computer Sciences, 27*(2), 1290–1307.
20. Song, X., Waitman, L. R., Hu, Y., Yu, A. S., Robins, D., & Liu, M. (2019). Robust clinical marker identification for diabetic kidney disease with ensemble feature selection. *Journal of the American Medical Informatics Association, 26*(3), 242–253.
21. Bui, D. T., Tsangaratos, P., Ngo, P. T. T., Pham, T. D., & Pham, B. T. (2019). Flash flood susceptibility modeling using an optimized fuzzy rule based feature selection technique and tree based ensemble methods. *Science of the Total Environment, 668*, 1038–1054.
22. Fan, S., Tang, J., Tian, Q., & Wu, C. (2019). A robust fuzzy rule based integrative feature selection strategy for gene expression data in TCGA. *BMC Medical Genomics, 12*(1), 14.
23. Jiménez, F., Martínez, C., Marzano, E., Palma, J., Sánchez, G., & Sciavicco, G. (2019). Multi-objective evolutionary feature selection for fuzzy classification. *IEEE Transactions on Fuzzy Systems, 27*, 1085–1099.
24. Dzulkalnine, M. F., & Sallehuddin, R. (2019). Missing data imputation with fuzzy feature selection for diabetes dataset. *SN Applied Sciences, 1*(4), 362.
25. Arefnezhad, S., Samiee, S., Eichberger, A., & Nahvi, A. (2019). Driver drowsiness detection based on steering wheel data applying adaptive neuro-fuzzy feature selection. *Sensors, 19*(4), 943.
26. Guru, D. S., Suhil, M., Raju, L. N., & Kumar, N. V. (2018). An alternative framework for univariate filter based feature selection for text categorization. *Pattern Recognition Letters, 103*, 23–31.
27. Labani, M., Moradi, P., Ahmadizar, F., & Jalili, M. (2018). A novel multivariate filter method for feature selection in text classification. *Engineering Applications of Artificial Intelligence, 70*, 25–37.
28. Mannepalli, K., Sastry, P. N., & Suman, M. (2018). Emotion recognition in speech signals using optimization based multi-SVNN classifier. *Journal of King Saud University – Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2018.11.012
29. Özseven, T. (2019). A novel feature selection method for speech emotion recognition. *Applied Acoustics, 146*, 320–326.
30. Srinivasa Murthy, Y. V., & Koolagudi, S. G. (2018). Classification of vocal and non-vocal segments in audio clips using genetic algorithm based feature selection (GAFS). *Expert Systems with Applications, 106*, 77–91.
31. Zhang, S., Zhang, S., & Huang, T. (2019). Speech emotion recognition using deep convolutional neural network and discriminant temporal pyramid matching. *IEEE Transactions on Multimedia, 20*(6), 1576–1590.
32. Shamim Hossaina, M., & Muhammad, G. (2019). Emotion recognition using deep learning approach from audio–visual emotional big data. *Information Fusion, 49*, 69–78.
33. Mao, Q., Dong, M., Huang, Z., & Zhan, Y. (2014). Learning salient features for speech emotion recognition using convolutional neural networks. *IEEE Transactions on Multimedia, 16*(8), 2203–2213.
34. Yan, Y., Shen, H., Liu, G., Ma, Z., Gao, C., & Sebe, N. (2014). GLocal tells you more: Coupling GLocal structural for feature selection with sparsity for image and video classification. *Computer Vision and Image Understanding, 124*, 99–109.
35. Bampis, C. G., & Bovik, A. C. (2018). Feature-based prediction of streaming video QoE: Distortions, stalling and memory. *Signal Processing: Image Communication, 68*, 218–228.
36. Zhou, H., You, M., Liu, L., & Zhuang, C. (2017). Sequential data feature selection for human motion recognition viaMarkov blanket. *Pattern Recognition Letters, 86*, 18–25.
37. Benuwaa, B.-B., Zhana, Y., Monney, A., Ghansah, B., & Ansah, E. K. (2019). Video semantic analysis based kernel locality-sensitive discriminative sparse representation. *Expert Systems with Applications, 119*, 429–440.
38. Selvakumar, K., Karuppiah, M., SaiRamesh, L., Islac, S. K. H., Hassan, M. M., Fortino, G., & Choo, K.-K. R. (2019). Intelligent temporal classification and fuzzy rough set-based feature selection algorithm for intrusion detection system in WSNs. *Information Sciences, 497*, 77–90.

39. Eskandari, S., & Javidi, M. M. (2016). Online streaming feature selection using rough sets. *International Journal of Approximate Reasoning, 69*, 35–57.
40. AlNuaimi, N., Masud, M. M., Serhani, M. A., & Zaki, N. (2019). Streaming feature selection algorithms for big data: A survey. *Applied Computing and Informatics*. https://doi.org/10.1016/j.aci.2019.01.001
41. Zhoua, P., Hua, X., Li, P., & Wu, X. (2019). Online streaming feature selection using adapted neighborhood rough set. *Information Sciences, 481*, 258–279.
42. Zhou, P., Hu, X., Li, P., & Wu, X. (2019). OFS-density: A novel online streaming feature selection method. *Pattern Recognition, 86*, 48–61.
43. Rahmaninia, M., & Moradi, P. (2019). OSFSMI: Online stream feature selection method based on mutual information. *Applied Soft Computing, 68*, 733–746.