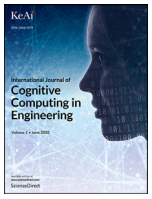
International Journal of Cognitive Computing in Engineering 1 (2020) 78–97

Contents lists available at ScienceDirect

International Journal of Cognitive Computing in Engineering

journal homepage: https://www.keaipublishing.com/en/journals/international-journal-of-cognitive-computing-in-engineering/



Microarray cancer feature selection: Review, challenges and research directions

Moshood A. Hambali [a,*], Tinuke O. Oladele [b], Kayode S. Adewole [b]

[a] Department of Computer Science, Federal University Wukari, Taraba, Nigeria
[b] Department of Computer Science, University of Ilorin, Kwara, Nigeria

ARTICLE INFO

Keywords:
Feature selection
Filter
Wrapper
Embedded
Microarray technology
Microarray data

ABSTRACT

Microarray technology has become an emerging trend in the domain of genetic research in which many researchers employ to study and investigate the levels of genes' expression in a given organism. Microarray experiments have lots of application areas in the health sector such as diseases prediction and diagnosis, cancer study and soon. The enormous quantity of raw gene expression data usually results in analytical and computational complexities which include feature selection and classification of the datasets into the correct class or group. To achieve satisfactory cancer classification accuracy with the complete set of genes remains a great challenge, due to the high dimensions, small sample size, and presence of noise in gene expression data. Feature reduction is critical and sensitive in the classification task. Therefore, this paper presents a comprehensive survey of studies on microarray cancer classification with a focus on feature selection methods. In this paper, the taxonomy of the various feature selection methods used for microarray cancer classification and open research issues have been extensively discussed.

1. Introduction

Cancer seldom referred to as malignant neoplasm, is one of the complex diseases that occur as a result of some cells manifesting certain traits of unrestrained development and perhaps later invades other sections of the body (American Cancer Society, 2017). According to the WHO Fact sheet, about 14 million new incidences of cancer was recorded in 2012, which is one of the prominent causes of morbidity and mortality globally (WHO, 2018). In the next 2 decades, the anticipated new cases are likely to rise to about 70%. Globally, virtually 1 out of 6 deaths reported is as a result of cancer disease. About 8.8 million deaths documented in 2015 which place cancer as the second prominent cause of death in the world (WHO, 2018).

However, the Cancer mortality rate can be cut down if diagnosed and treated early. Medical data mining is one of the branches of data mining that help in analysing, extracting, transforming, interpreting and visualizing medical records stored in repositories. Medical data mining is not only interesting but also challenging, because diagnosis and prediction of diseases are not the issues of accuracy alone but it also a matter of life and death. A wrong classification or prediction can be disastrous to the life of patients and their relatives. Hence, medical data mining is considered sometimes as an expert system that employs machine learning to make decisions that help experts to easily and quickly diagnose and predict diseases (Doreswamy & UmmeSalma, 2016).

Microarray data are mostly employed in cancer researches where an early diagnosis of cancer disease is very important in determining the nature of treatment and its survivability (Selvaraj & Natarajan, 2011). Microarray Technology (MT) assists biologists in examining the activity of ten thousand of genes in a single experiment and gets important information about the cell's functionality. This specific information can be exploited for the diagnosis of many diseases such as Alzheimer's disease (Panigrahi & Singh, 2013), diabetes diseases and cancer diseases (Hira & Gillies, 2015; Alomari et al., 2017). The gene expression data can be produced from this technology, which is highly germane for cancer classification and prediction. Though, gene expression data were characterized by high dimensionality data which are irrelevant, redundant, and noisy genes that are of no importance to the diagnosis of diseases. A large number of genes (features) versus the small size of sample and presence of redundancy in data expression are the major reasons that usually result into poor diagnosis and classification, in machine learning and data mining approaches employed in the field of medicine (Veerabhadrappa & Rangarajan, 2010; Bennet, Ganaprakasam & Kumar, 2015). In recent times, researchers had employed computational intelligence algorithms to investigate the most informative genes that contribute to cancer diagnosis (Alomari et al., 2017).

Numerous studies have revealed that most genes present in DNA (Deoxyribonucleic Acid) Microarray datasets are not pertinent in the accurate diagnosis of different diseases (Alomari et al., 2017;

* Corresponding author.
E-mail addresses: hambali@fuwukari.edu.ng (M.A. Hambali), tinuoladele@gmail.com (T.O. Oladele), adewole.ks@unilorin.edu.ng (K.S. Adewole).

https://doi.org/10.1016/j.ijcce.2020.11.001
Received 14 August 2020; Received in revised form 13 October 2020; Accepted 4 November 2020
Available online 7 November 2020

Saqib, Qamar, Khan & Aslam, 2020; Saeid, Nossair & Saleh, 2020; Hambali & Gbolagade, 2016; Hambali, Saheed, Oladele & Gbolagade, 2019; Liang et al., 2018). To circumvent the problem of the 'curse of dimensionality', feature selection (which is sometimes called gene selection) is the technique employed to seek the most informative genes that can augment diagnosis and predictive accuracy of diseases (Alomari et al., 2017; Hambali & Gbolagade, 2016; Hambali, Saheed, Oladele & Gbolagade, 2019).

Therefore, dimensionality reduction is one of the exciting research areas in the domain of data mining, pattern recognition, machine learning, and statistics. The main drive of dimensionality reduction is to augment the classification accuracy performance of classification algorithm through the removal of redundancy and irrelevant features in the Microarray dataset. Dimensionality reduction can be accomplished in many ways. But, the choice of selecting the appropriate dimensionality reduction method is determined by the domain application and the peculiarity of the dataset. Feature selection methods are categorized into four (4) approaches, namely; filter, wrapper, embedded and hybrid approaches (Veerabhadrappa & Rangarajan, 2010).

Selection of features in filter algorithms is done by using individual characteristic features. Wrapper approaches employ a certain machine learning algorithm or evolutionary approaches to select subset features. The filter methods are characterized with very fast computation and low accuracy while wrapper approaches tend to have better accuracy performance with less computation rate. In a domain with large datasets, filter-based methods have been established to be the best choice than wrapper method due to their speed. A mutual drawback of the two approaches is that they both ignore the interaction between the classifier and the dependence among features, and this result to diverse classification accuracy when features selected are subjected to different classification algorithms. While embedded approaches, make use of specific learning algorithms. Embedded methods have the superiority advantage of interacting with the classification algorithm; while at the same time have lower computational cost compared to wrapper methods (Veerabhadrappa & Rangarajan, 2010). Although, there is a lot of feature selection approaches in the literature aim to select informative and relevance features that will augment classification rate and minimize computation cost but few of these studies actually survey the comprehensive approaches employed. Though the majority of the reviews centred on a particular feature selection approach (Liang et al., 2018; Lazar et al., 2012; Zhang, Nie, Li & Wei, 2019), or application of feature selection in the medical field in general (Remeseiro López & Bolon Canedo, 2019). None of them provided a comprehensive state-of-the-art of all available approaches with their taxonomy in the previous studies, addressed challenges issue related to microarray data and microarray experiments. For instance, Ref. (Lazar et al., 2012) presented a survey on filter techniques for microarray genes analysis. Ref. (Zhang, Nie, Li & Wei, 2019) presented a survey on feature selection and feature-level fusion strategies but the focus is not on microarray data. Ref. (Liang et al., 2018) presented a compressively a review on matched-pairs feature selection for gene expression microarray data. Ref. (Hira & Gillies, 2015) provided a brief review of existing popular feature extraction and feature selection techniques that are widely in used. Also, Ref. (Remeseiro López & Bolon Canedo, 2019) presented a review of the feature selection approaches designed for and applied to medical field problems, which include biomedical signal processing, medical imaging and DNA microarray data analysis. Therefore, the most closely related work to this review is the study presented by Bolón-canedo et al. (2014), however, their review only addressed the classification of feature selection, microarray datasets and their sources, and related challenges. This work is a comprehensive review covering over 150 articles from the literature. The objective of this work is to highlight the challenges and issues related to microarray cancer datasets, the existing feature selection techniques employed for feature selection, a brief description of microarray experiment and identify the limitation of the existing approaches. Also, this paper further identifies the crucial areas for future study in this domain.

The remaining part of this paper is structured as follows: Section 2 briefly discusses the introduction to Microarray technology and its data; followed by Section 3, which focuses on the analysis of Microarray data. Section 4 presents the taxonomy of the methods used for feature selection in this domain. Section 5 briefly reviews recent hybrid approach. Open research issues were discussed in Section 6 and finally, Section 7 concludes the paper.

2. Microarray technology

The advent of Microarray technology has brought about innovative changes into the biological fields. It is regarded as a motivating advent for valuable researches. It has allowed concurrent examination of hundreds to thousands of gene activities at the same time. Though most of biologists and researchers from other fields find hitches while mining and working with this type of data, likewise Microarray experimental results are stored in diverse and multiple databases.

Microarray technology first appeared in the research arena in the late 1980s (Rafii, Hassani & Kbir, 2017). Augenlicht et al. (1987) were the first researchers to describe DNA Microarrays, where about 4000 of complementary DNA (cDNA) sequences were spotted on nitrocellulose (Augenlicht, Taylor, Anderson & Lipkin, 1991). Microarray has provided a robust opportunity for the biologists to examine and measure hundreds of thousands of gene expressions simultaneously (Remeseiro López & Bolon Canedo, 2019; Rafii, Hassani & Kbir, 2017; Chen et al., 2012). The Microarray technology has inspired a new line of research in bioinformatics, medical fields and machine learning (Hira & Gillies, 2015; Ventimiglia & Petralia, 2013). DNA Microarrays are typically referred to as DNA chips or biochips that contain a sequence of microscopic DNA spots mounted on a solid surface. The biologists are making uses of DNA Microarrays as a platform to examine the expression levels of thousands of genes simultaneously, or the numerous parts of the genes genotype (Rafii, Hassani & Kbir, 2017).

2.1. DNA microarray

Cells consist of a core part called a nucleus, and in this nucleus, there is DNA that contains encoded "program" for upcoming generations. DNA comprises of coding and non-coding components. The coding parts known to be genes, stipulate the protein structures which perform the vital job in every organism. Proteins are generated in genes in two phases: the first stage is the transcription of DNA into mRNA and then followed by mRNA being translated into proteins. The development of molecular genetics technologies, for example, DNA Microarrays, has provided a platform to successfully view the functionality of the cells and with potential to examine the expression of tens of thousands of genes simultaneously (Bolón-Canedo et al., 2014). Fig. 1 shows the generic process of obtaining the gene expression profiles from a DNA Microarray. These gene expression data can serve as inputs to large-scale data analysis, such as, to improve understanding and classifying the genes into diseased or normal cells. Gene expression is the process of portraying how genotype metamorphosis into phenotype by examining the total number of transcribed mRNA in a genomic system. There are lots of standardized approaches for recognizing the variation in gene expression which include the differential display, Microarray hybridization, RNA-seq sequencing, Serial Analysis of Gene Expression (SAGE), subtractive hybridization. Detailed information about all these approaches can be found in Acunzo, Romano, Wernicke and Croce (2015), Hammond (2015), de, Monobe and da Silva (2016).

Fig. 1 depicts the procedure for acquiring microarray data which involves a collection of sample tissues from both cancerous and healthy human tissues, isolate mRNA from samples using either a column or a solvent such as phenol-chloroform. After that, the cDNA label is created and then hybridized by combining both tissues on a Microarray plate.

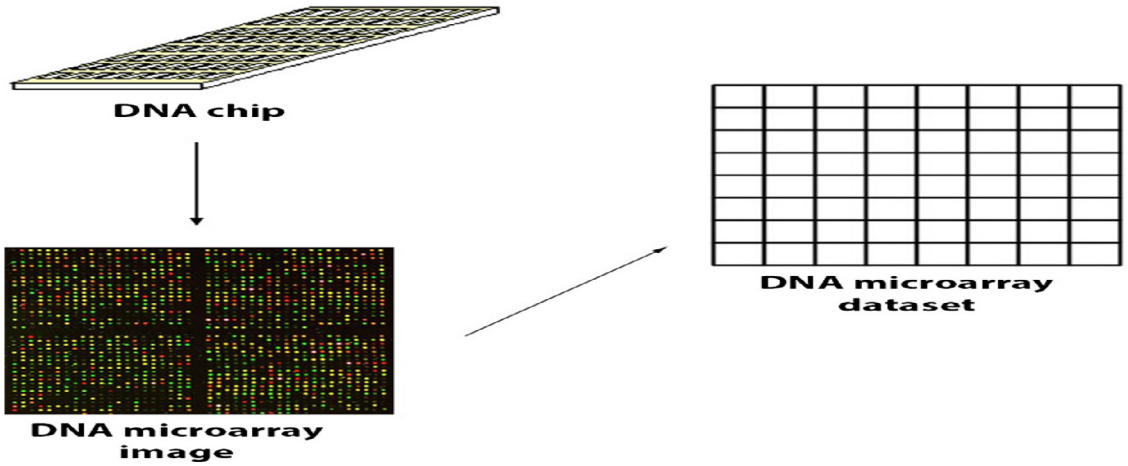


Fig. 1. Generic procedure for acquiring the gene expression data (Bolón-Canedo et al., 2014).

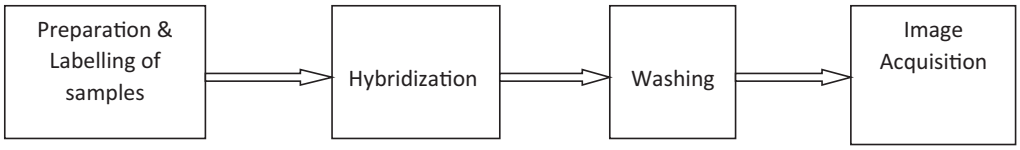


Fig. 2. Steps involved in microarray experiment.

The last stage is the detection of the relative intensities of fluorescence under Microarray scanner to view the results and also save the data in form of an array for future use.

2.2. Microarray experiment

Microarray experiment is the hybridization reaction that involves the comparison of relative cellular mRNA acquired from two tissue samples. The hybridization reaction takes place when RNA molecules or the single-stranded DNA join to integrate into double-stranded complexes. The importance of gene expression Microarray experiments is to investigate and examine the differential gene expression between groups, that is, class comparison, or by classification studies, that is, class prediction and class discovery (Rafii, Hassani & Kbir, 2017).

There are basically four stages that are involved in measuring the gene expressions in Microarray experiment as shown in Fig. 2. The first stage is referred to as preparation and labeling of samples, where RNA extraction is done from a specific tissue, and the labelling depends on the selected technology. The second phase is hybridization. It is the stage where the probes and labels of DNA or RNA targeted for heteroduplexes; they are base-paired on the surface of glass through Watson-Crick. Detection of hybridization can be done optically, electrochemically or by employing mass sensitive devices (Rafii, Hassani & Kbir, 2017). Washing is the third stage where excess solution from hybridization array is eliminated. It is a stage where non-specifically cRNA bound are removed from the microarray surface in order to reduce the effect of the sensitivity and background level of the whole microarray. The final phase is the image acquisition stage where the hybridized image of the array is produced.

2.3. Microarray data

The data generated by Microarray experiments are arranged and kept as large matrices ($M \times N$). Each Microarray data matrix comprises of the samples illustrated in rows and the genes (features) in the columns as shown in Table 1.

Microarray data is of the form of M by N matrix and is very huge. Where M is rows and it indicates the samples, N is columns and it indicates the genes and every single cell has a certain value of gene expression in a sample (Kong, Yu, Minion & Rajan, 2011). x_{ij} denotes the expression level of the gene j and the condition or sample i. Where j ranges between *1* and M, and i from *1* to N.

Table 1
Microarray data matrix format.

	G_1	G_2	...	G_M
S_1	x_{11}	x_{12}		x_{1M}
S_2	x_{21}	X_{22}		x_{2M}
.				
.				
.				
S_N	x_{N1}	x_{N2}		x_{NM}

3. Microarray data analysis

Biomedicine field has gained popularity in machine learning in the last few decades mainly as a result of the enormous amount of data available for extraction from genetic tissues. Especially with the advent of DNA Microarray datasets has paved way to the development of a novel and active area of research in machine learning and bioinformatics. For the purpose of machine learning, Microarray data is usually considered as structured data that are characterized with few samples (most of the time it is less than 100) but has numerous features (in the order of thousands). Dealing with this nature of data that has just a few samples with a large number of features pose serious challenges for researchers in the field of machine learning; because of the likelihood of "false positives", which may occur perchance at the time of building the model for prediction or during the selection of relevant features (genes) (Remeseiro López & Bolon Canedo, 2019). It has been established from the literature that very few genes from these enormous number of genes available in a DNA Microarray are pertinent for classification purpose. In this situation, feature selection is crucial in removing redundant and irrelevant features, and also to assist experts to uncover fundamental connections between gene expression and a particular disease.

Therefore, it is necessary to reduce the dataset of those genes to reduce the cost of identifying the best genes that can distinguish between the cell classes (normal or abnormal cells). There are two famous techniques for in-depth Microarray data analysis: clustering and classification (Mutch et al., 2001). Clustering is an unsupervised technique of classifying data into clusters of genes or samples with related patterns or characteristics. Classification is a supervised technique that learns from examples. Provided with a set of pre-classified samples, the classifier learns to assign unknown sample cases to one of the predefined classes (Selvaraj & Natarajan, 2011).

3.1. Cluster analysis

Clustering is among the famous methods employed in the Microarray data analysis. It is employed to seek co-regulated and functionality of related groups (Svrakic et al., 2003). Clustering is generally useful in the instances when the whole sets of an organism's genes are not available. Cluster techniques can be categorized into three common types such as hierarchical clustering, k-means clustering, and self-organizing maps. Hierarchical clustering is a popularly employed unsupervised technique that models clusters of genes with identical patterns of expression (Eisen, Spellman, Brown & Botstein, 1998). Its process is performed in iteration where genes that are highly correlated in terms of the extent of their expression are grouped together. It is a way of modelling cluster hierarchically. Genes are represented as leaves of tree branches in the dendrogram. In general, hierarchical clustering is usually modeled in two ways; that is agglomerative and divisive. Agglomerative uses a bottom-up approach where each individual that has a similar pattern are clustered together and different pairs of clusters are amalgamated as one moves up the hierarchy. Divisive employs top to down approach that is, all patterns (Datasets) start in a single cluster and splits are done recursively as one traverse down the hierarchy (Eisen, Spellman, Brown & Botstein, 1998).

K-means clustering is a machine learning algorithm that clusters datasets into groups of similar observations without the need of having prior knowledge of the relationships that exist in the data (Tavazoie et al., 1999). It is a clustering technique that is very easy to accomplish and is commonly employed in biometrics and medical imaging applications. Euclidean properties of the vector space are usually applied in the K-means clustering algorithm. The vector space is initially partitioned into K parts, then the algorithm computes the focal positions in each of the subspaces and modifies the partition in order to assign each vector into the cluster that is closest to the midpoint. This iteration is recurring until either the partitioning is steady or the numbers of specified iterations are reached (Brazma & Vilo, 2000).

The third category of the cluster is the Self-Organizing Map (SOM) which is a neural network-based non-hierarchical clustering approach. SOMs work in the same manner as K-means clustering (Tamayo et al., 1999).

3.2. Classification

Classification is a supervised learning approach, sometimes referred to as class prediction or discriminant analysis. When given a set of a pre-classified dataset, for instance, cancer dataset with different classes such as normal and diseased cells or Acute myeloid Leukemia (AML) and Acute Lymphoblastic Leukemia (ALL). A classifier will seek for rules that will enable it to assign unknown samples supply to any of the above classes (or diagnose the diseased cell) (Quackenbush, 2001). In the classification task, there is a need for sufficient numbers of samples to allow the algorithm to be able to use a particular sample as training dataset and other fresh independent samples are used to test the algorithm which will serve as the test set. There are lots of classification algorithms available in the literature, such as Artificial Neural Networks, Decision Trees, K Nearest Neighbors (KNN), Weighted voting, Support vector machines (SVM) and many others. One of the interesting areas of classification approach is in the clinical diagnostic application which predicts and classifies disease types and subtypes. Prevalent examples include diagnosis of leukemia into ALL or AML (Golub et al., 1999), brain tumor diagnosis (Wang et al., 2003) and lymphoma (Kim et al., 2005) and breast cancer diagnosis (Hambali & Gbolagade, 2016).

3.3. Big data and deep learning

The enormous quantity of data continuously produced by Bioinformatics is a vital challenge for the machine learning techniques (Basgall et al., 2019; Rendón et al., 2020). This occurrence is not limited to the amount of information, but at the same time by the swiftness of data transmission and the diversity of data; that is, the big data physiognomies (Fernández, del Río, Chawla & Herrera, 2017; Elshawi, Sakr, Talia & Trunfio, 2018). Deep learning techniques are gaining more attention as alternative approaches, to efficiently analyze a large volume of information obtained from the applications of big data. The approaches usually yield a better analysis result than typical machine learning techniques (Rendón et al., 2020; Guo et al., 2016; Reyes-Nava et al., 2018). Deep learning employs supervised and/or unsupervised learning algorithms to model high-level abstractions in data, to learn from multiple levels of abstractions. It applies hierarchical illustrations of data for classification. Deep learning approaches have been adopted in many domains, such as computer vision, pattern recognition, speech recognition and natural language processing. As a result of the exponential growth of data in these domains, deep learning expedients accurately prediction of the huge amount of data. Recently, researchers have designed efficient and scalable parallel approaches for training deep models (Hinton et al., 2012). Deep learning has been employed in many fields for decision making, semantic indexing and information retrieval (Kashyap et al., 2014).

The major difference between deep learning architectures and machine learning is the number of hidden layers. Usually, machine learning designs consist of 3 layers, that is, input layer, hidden layer and output layer; while deep learning has more than 3 layers, with the presence of more than one single hidden layer. The network architecture is categorized as deep learning when it has more than 3 layers in its design (Shekar & Dagnew, 2020; Viloria, Bonerge, Lezama & Mercado-caruzo, 2020).

In the domain of Bioinformatics, deep learning has started attracting the attention of researchers for diseases classification (LeCun, Bengio & Hinton, 2015); for example, in the classification of microarrays gene-expression data (Hira & Gillies, 2015; Cleofas-Sánchez, Sánchez & García, 2019; Geman et al., 2016). Classical applications problems that required deep neural networks approach usually have a high number of samples and dimensionality (Chen, Qian, Shi & Pan, 2017; Maqlin, Thamburaj, Mammen & Manipadam, 2015). Though in microarray cancer datasets, the number of instances are few, and the dimensionality is very high, which pose challenges for this scenario. In some circumstances, the cancer classes are imbalanced, where one class is greatly underrepresented related to the others (Alejo et al., 2016; Reyes-Nava et al., 2019).

The deep learning architectures are majorly artificial neural networks of multiple non-linear layers and numerous kinds have been proposed base on the input data characteristics and research objectives. Seonwoo et al. (Min, Lee & Yoon, 2017) categorized deep learning architectures into four types: convolutional neural networks (CNNs) (LeCun, Bengio & Hinton, 2015; Lawrence, Giles, Tsoi & Back, 1997; Krizhevsky, Sutskever & Hinton, 2012), emergent architectures (Lena, Nagata & Baldi, 2012; Graves & Schmidhuber, 2009; Hadsell et al., 2009; Masci, Meier, Cireşan & Schmidhuber, 2011), deep neural networks (DNNs) (Svozil, Kvasnicka & Pospichal, 1997; Vincent, Larochelle, Bengio & Manzagol, 2008; Vincent et al., 2010; Hinton, Osindero & Teh, 2006; Hinton & Salakhutdinov, 2006) and recurrent neural networks

(RNNs) (Williams & Zipser, 1989; Bengio, Simard & Frasconi, 1994; Hochreiter & Schmidhuber, 1997; Felix, Schmidhuber & Cummins, 2000). Most of the authors adopted 'DNNs' to have encompasses all deep learning architectures (LeCun, Bengio & Hinton, 2015; Goodfellow, Bengio, Courville & Bengio, 2016); though some researchers used 'DNNs' to refer specifically to a multilayer perceptron (MLP) (Rendón et al., 2020; Svozil, Kvasnicka & Pospichal, 1997; Guillen, Ebalunode & Learning, 2016), deep belief networks (DBNs) (Hinton, Osindero & Teh, 2006; Hinton & Salakhutdinov, 2006), and stacked auto-encoder (SAE) (Vincent, Larochelle, Bengio & Manzagol, 2008; Vincent et al., 2010), which use perceptrons (Minsky & Papert, 1969), restricted Boltzmann machines (RBMs) (Hinton & Sejnowski, 1986; Hinton, 2012) and auto-encoders (AEs) (Fukushima, 1975) as the building components of neural networks, respectively. CNNs are architectures most thrived specifically in image recognition and comprises of convolution layers, pooling layers and non-linear layers. RNNs are developed to apply sequential information of input data with cyclic connections among building components like perceptrons, gated recurrent units (GRUs) (Cho et al., 2014) or long short-term memory units (LSTMs) (Hochreiter & Schmidhuber, 1997; Felix, Schmidhuber & Cummins, 2000). Furthermore, common emergent deep learning architectures include deep Spatio-temporal neural networks (DST-NNs) (Lena, Nagata & Baldi, 2012), convolutional auto-encoders (CAEs) (Hadsell et al., 2009; Masci, Meier, Cireşan & Schmidhuber, 2011) multidimensional recurrent neural networks (MD-RNNs) (Graves & Schmidhuber, 2009). Much sophisticated deep learning largely depends on the technology employed to implement them. The most common libraries used for this approach include Caffe, Theano, Tensorflow (Abadi et al., 2016), PyLearn2, H2O (Aiello et al., 2016) and the Apache Spark working platform (Zaharia et al., 2016).

Ref. (Chen, Qian, Shi & Pan, 2017) used multiple Recurrent Neural Networks (RNN) for the classification of breast cancer patients into benign and malignant. The proposed model comprises of four RNN to excerpt the characteristics of patients and the last RNN for the final classification. Ref. (Abdel-Zaher & Eldeib, 2016) proposed an automatic diagnosis system for breast cancer detection. The system is used deep belief network (DBN) for the training of the system and followed by a back-propagation neuronal network. Ref. (Reyes-Nava et al., 2019) proposed a Deep Learning Multi-Layer Perceptron (DL-MLP) for classification of gene-expression microarray datasets which characterized with the presence of class imbalance. Also, Ref. (Viloria, Bonerge, Lezama & Mercado-caruzo, 2020) studied the behavior of the MLP DL classifier applied on high dimensional, low pattern, and high-class imbalance gene expression microarray datasets.

4. Microarray cancer feature selection techniques

Many classification algorithms were initially not designed to handle huge volumes of irrelevant features; merging them with feature selection (FS) methods have become inevitable in several applications (Guyon & Elisseeff, 2003; Yu & Liu, 2004).

The major focus of feature selection (FS) is to select a subset of features from the input data, and that subset can successfully describe the input data. Feature selection can also diminish negative impacts from irrelevant or noise features (Zhang, Nie, Li & Wei, 2019; Brown, Pocock, Zhao & Luján, 2012; Bolón-Canedo, Sánchez-Maroño & Alonso-Betanzos, 2015; García, Luengo & Herrera, 2016; Zhou, Hu, Li & Wu, 2017). No extra information can be obtained from redundant features except that it creates noise for the classification algorithm. In a nutshell, few unique features can provide crucial information that can be achieved by using the whole dataset as input, which can serve as class representative. Therefore, by discarding the redundant features, the volume of data can be reduced to enhance the performance of the classification algorithm. In some instances, some features did not have any correlation with the classes at all. They are just pure noise and hence, introduce bias to the classification algorithm and weaken its performance. By employing FS techniques, it provides insight into the feature selection process and therefore improves the classification accuracy (Zhang, Nie, Li & Wei, 2019).

In summary, the goals of FS are multifarious; the vital ones include the following:

(a) It prevents overfitting and therefore, enhance model performance,
(b) It brings about the faster processing and reduce the model cost-effectively and
(c) It helps in gaining deeper knowledge into the core processes that generate the data.

Nevertheless, the benefits of FS methods come with a trade-off. The process of searching for the subset of relevant features brings about an extra layer of complexity in the processing task. There is need to seek the best model parameters for the best feature subset rather than optimizing the model's parameters for the full feature subset since assurance for the optimization of parameters for the complete feature set that can equally lead to the best feature subset is not granted (Daelemans, Hoste, De Meulder & Naudts, 2003).

Quite a lot of statistical approaches have been used to select the genes for medical disease diagnosis, prognosis, and therapeutic targets (Tu, Yu, Guo & Li, 2004). Apart from the statistical methods, in recent times, data mining and machine learning approaches have been widely adopted in genomic data analysis (Khan & Alam, 2012; Zahiri et al., 2013). For instance, Cho et al. (2004) proposed a modified kernel Fisher discriminant analysis (KFDA) for the analysis of the hereditary breast cancer dataset. The KFDA classifier employed the mean-squared-error as the gene selection criterion.

Furthermore, numerous hybrid evolutionary algorithms are available in the literature to improve the accuracy of the classification techniques (Huang & Chang, 2007; Tan, Fu, Zhang & Bourgeois, 2008; Tong & Schierz, 2011). The goal of many evolutionary algorithms is to seek for best features subset by adopting bio-inspired approaches such as particle swarm optimization (PSO), Honey Bee, Firefly algorithms and so on. These types of algorithms show suitable performance improvement over several problems but require experts' involvement to attain the desired performance outcome.

FS methods differ from one another in the mode of integrating their search in the added space of feature subsets, in the model selection. Gene selection is a subclass of larger machine learning class of feature selection. There are two major ways of categorizing FS and the related taxonomy (Zhang, Nie, Li & Wei, 2019) – Label status and Search strategy as illustrated in Figure 3.

4.1. Label status

In this category, methods are categorized based on the samples' feature labeled or not. The labeled samples offer hands-on information, such that relevant and informative features are selected to differentiate samples from different classes through the supervised FS methods (Nie, Huang, Cai & Ding, 2010). In label status category, FS technique can be further divided into three groups, such as supervised methods (Nie, Huang, Cai & Ding, 2010; Li et al., 2007; Zhao & Liu, 2007; Huang, 2015; Zhang, Nie & Li, 2018), semi-supervised methods (Zhao & Liu, 2007; Xu, King, Lyu & Jin, 2010; Wang et al., 2017), and unsupervised methods (Cai, Zhang & He, 2010; Yang et al., 2011; Liu, Rallo & Cohen, 2011; Hu et al., 2017; He et al., 2017). In semi-supervised FS approach, parts of the data are labeled while others are not labeled.

Most of the famous semi-supervised FS methods make use of similarity matrix construction, such that the selected features are based on graph structure (Zhao & Liu, 2007; Cheng, Zhou & Cheng, 2010). Due to the nonappearance of labels that can help to guide the exploration of informative genes, unsupervised FS make use of some tricky approaches (Zhang, Nie, Li & Wei, 2019) to select the relevant features from the unlabeled dataset. In its simplest form, the FS procedure is performed by evaluating individual features and then ranks them using

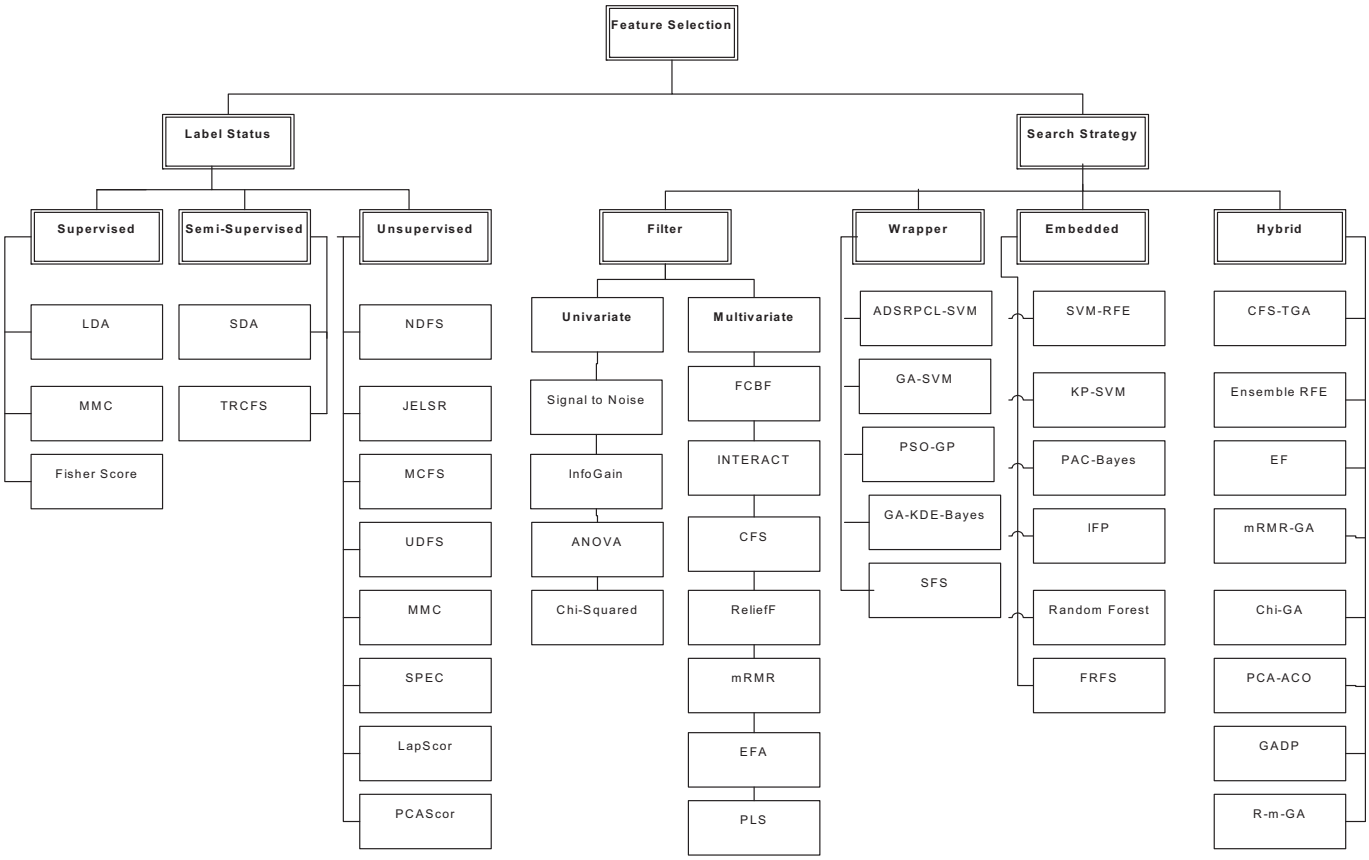


Fig. 3. Taxonomy of feature selection methods for microarray cancer classification.

their correlation with class labels (Zhang, Nie, Li & Wei, 2019). Though, Hall (1999) has revealed that the best features subsets are those subsets whose features did not correlate with one another. Therefore, it is desirable to evaluate and consider the features as a subset rather than the individual. Liu et al. (2010) provided a summary of the subset feature selection procedure into three main phases:

1 **Search:** Producing features subset to be evaluated.
2 **Evaluation:** Evaluating the efficacy of the produced subset.
3 **Stop:** Determining whether to continue the search or to halt the search.

To achieve the aim of feature selection, a lot of criteria have been suggested to estimate the relevance of the feature (Zhang, Nie, Li & Wei, 2019; Zhao & Liu, 2007).

4.1.1. *Supervised method*

Supervised FS methods are more related to the labeled data points. Conservative supervised approaches such as Fisher Score (Duda, Hart & Stork, 2012) does the ranking of the features individually and it did not consider feature correlation. Linear discriminant analysis (LDA) (Fisher, 1936) was employed to select features by maximizing the ratio between the 'between-class scatter and within-class scatter'. However, with the small sample size data, LDA encountered some challenges (Fukunaga, 2013). To solve this problem, maximum margin criterion (MMC) is used in Li, Jiang & Zhang (2004), which converts the computation of ratio of traces between the 'between-class scatter and within-class scatter' into the subtraction form. In summary, all supervised methods need appropriately labeled data to achieve the task of FS. But when the data for training are not sufficiently labeled, the performance of the method will degenerate (Luo et al., 2013).

4.1.2. *Semi-supervised method*

In semi-supervised FS methods, Luo et al. (2018) and Han et al. (2014) worked on both labeled and unlabeled data. Consequently, semi-supervised approaches demonstrated the ability to select features from unlabeled data, when provided with few numbers of labeled data. Graph Laplacian based semi-supervised techniques assumed that most data samples lay on a low-dimensional manifold, such as semi-supervised Discriminant Analysis (SDA) (Cai, He & Han, 2007). In graph Laplacian based approaches; unlabeled samples are harnessed through uses of graph Laplacian matrix. For example, trace ratio criterion for feature selection (TRCFS) is one of the effective semi-supervised FS algorithms to select informative features (Liu, Nie, Wu & Chen, 2013). However, semi-supervised FS techniques are typically less effective in handling huge dataset due to the cost of computation time of the graph (Chang, Nie, Yang & Huang, 2014).

4.1.3. *Unsupervised method*

As a result of the missing label information that can help in guiding the search for informative features in the data and the problem of getting the actual label information, unsupervised FS is regarded as a challenging problem (Zhang, Nie, Li & Wei, 2019; Dy & Brodley, 2004). Multiple criteria were considered for the evaluation of feature relevance. One of the famous criteria used is to select those features that retain the original data manifold structure. Another popularly employed criterion is to find cluster indicators through clustering algorithms and then follow a supervised approach of problem-solving. Unsupervised FS can be categorized into two different approaches. One of the techniques is to find cluster indicators which concurrently execute the supervised feature selection. For instance, Yang et al. (2011) and Li et al. (2014) designed a combined framework that consists of nonnegative spectral clustering and structural learning to select features from unlabeled data. The second one is the hierarchical form, which the first search for cluster indicators, then

executing feature selection, and lastly iterating two previous steps until a certain criterion is satisfied. Zhao et al. (2011) proposed Similarity Preserving Feature Selection (SPFS) for feature selection by conserving sample similarity, which can handle feature redundancy successfully.

Other examples of unsupervised FS methods include Nonnegative Discriminative Feature Selection (NDFS), Embedded Unsupervised Feature Selection (EUFS) (Wang, Tang & Liu, 2015), Multi-cluster feature selection (MCFS) (Cai, Zhang & He, 2010), Joint Embedding Learning and Sparse Regression (JELSR) (Hou et al., 2013) and Unsupervised Discriminative Feature Selection (UDFS) (Yang et al., 2011).

In recent times, some investigations have appeared that leveraged on both the learning mechanism and manifold structure. The typical techniques include; PCA Score (PcaScor) (Krzanowski, 1987), Laplacian Score (LapScor) (He, Cai & Niyogi, 2006), Spectral Feature Selection (SPEC) (Zhao & Liu, 2007), Minimum Redundancy Spectral Feature Selection (MRSF) (Zhao, Wang & Liu, 2010) and MCFS (Cai, Zhang & He, 2010). Generally, these approaches first employ numerous graphs to depict the manifold structure. LapScor and SPEC then utilize other metrics for the ranking of each feature. MCFS and MRSF, however, enhance the process with sparse constraints in multi-output regression (Hou et al., 2013).

4.2. Search strategy

Based on the different searching strategies, FS can further be grouped into three approaches, that is, filter techniques, wrapper techniques, and embedded techniques (Zhang, Nie, Li & Wei, 2019; Remeseiro López & Bolon Canedo, 2019; Khan & Alam, 2012; Saeys, Inza & Larrañaga, 2007). Filter techniques accomplish the FS task before classification or clustering of the data and usually execute in a two-step approach. In the first step, all the features are ranked based on defined criteria and the second step involves the retaining of the top-ranking features. Filter techniques evaluate the relevance of the feature genes by searching for the inherent properties of the data. In some of the filter approaches, a feature relevance score is estimated, and low-scoring features are disregarded. Subsequently, the features with high rank are chosen (Tan, Fu, Zhang & Bourgeois, 2008; Pihur, Datta & Datta, 2008; Qi, Sun, Sun & Pan, 2011). The famous Filter methods include ReliefF (Kira & Rendell, 1992; Raileanu & Stoffel, 2004), signal-to-noise ratio (Golub et al., 1999), receiver operating characteristics analysis (Khodarev et al., 2003), information gain (Raileanu & Stoffel, 2004; Peng, Long & Ding, 2005), Laplacian score (He, Cai & Niyogi, 2006), F-statistic (Ding & Peng, 2005), mRMR (Peng, Long & Ding, 2005), fuzzy logic for removing redundant features (Huang & Chang, 2007), T-test feature ranking for gene selection (Zhou & Wang, 2007), consensus independent component analysis that employs gene expression value for cancer classification (Zheng, Huang, Kong & Zhao, 2008), mutual information (Cai, Hao, Yang & Wen, 2009) and maximum-minimum correntropy criterion (Mohammadi, Noghabi, Hodtani & Mashhadi, 2016).

Filter approaches are machine learning algorithm independent. It focuses on the general inherent property of the data. Due to their independence from the learning algorithm, they have less computational cost and have a good generalization capacity. Wrappers and embedded approaches both need a machine-learning algorithm to accomplish feature selection. In wrappers, evaluation of candidate feature subsets is done by a learning algorithm. The interaction of wrappers with the classifier makes them more computationally expensive than filters, though they perform better than filters. Wrapper methods use the learning algorithm to evaluate the FS. Guyon et al. (2002) used SVM based Recursive Feature Elimination (RFE) to choose significant genes associated with cancer diseases. Embedded methods implement FS in relation to optimization model (Zhang, Nie, Li & Wei, 2019). Embedded techniques maintain mid-course between filters and wrappers since FS is part of the training process for the learning algorithm. The exploration of the best features subset is done during the training of the learning algorithm (for instance, during weights optimization in a neural network). Due to this reason, embedded approaches are less computationally expensive compared to wrappers. There are also other approaches to combining algorithms. This is referred to as hybrid approaches, which typically combines two or more FS algorithms of the different search strategy sequentially. For example, a less computationally expensive algorithm such as filter may be first applied to eliminate some features and later employ a more complex and costly algorithm (such as wrapper) for fine-tuning.

4.2.1. Filter techniques

Filter techniques evaluate the gene subsets informative by observing the inherent data characteristics only, that is, statistical measures by which a single gene (a subset of genes) is typically assessed against the class label. Classical filter techniques are commonly employed in Microarray data analysis, such as consistency-based filter, Fast Correlation-Based Filter (FCBF), Correlation Feature Selection (CFS) and ReliefF. In these approaches, individual features are considered separately; hence it did not give room for feature dependencies, as a result of that, it usually gave poor diagnostic accuracy when compared to other types of FS methods. Filter methods have some pros which are not limited to the following: they are computationally simple and easy, efficiently scaled through very high-dimensional datasets, and classifiers independent. FS techniques can also be classified into univariate and multivariate types. Univariate approaches treat each feature individually of each other, a hitch that multivariate methods proffer solution to by incorporating feature dependencies to some certain level, but with an increase in demanding for more computational resources. Signal-to-noise ratio (Golub et al., 1999; Chow, Moler & Mian, 2001) is one of the common examples of univariate approach; it evaluates each feature independently based on certain criteria. Thus, univariate methods are simple, fast and the most available methods in the literature (Bolón-Canedo et al., 2014; Ben-Dor et al., 2000; Xing, Jordan & Karp, 2001; Bolón-Canedo, Sánchez-Maroño & Alonso-Betanzos, 2013). On the other hand, forward feature selection is one example of multivariate techniques. It evaluates relevance and dependency of the features in a group subset (Grate, Bhattacharyya, Jordan & Mian, 2002; Bø & Jonassen, 2002). It starts with a single discriminative feature and continuously adds the new discriminative features using greedy search technique. Another multivariate approach is base-pair selection; the relevance of the features are estimated via pairs of features. Though, several snags are limiting the application of multivariate methods. Mainly, they are characterized by the overtraining of the data and normally are of the high computational cost.

Filter techniques that have been employed in various tasks can be found in the work of Bekkerman et al. (2003), Caruana and Sa (2003) and Davidson and Jala (2010). Javed et al. (2010) and Peng et al. (2005), developed a novel ranking criterion based on the class density of the binary data. The related algorithm applies the filter technique to rank the features and later the wrapper approach to further remove irrelevant features. Caruana and Sa (2003), eliminated features are employed for multi-task learning (MTL). In Stoppiglia, Dreyfus, Dubois and Oussar (2003), a random variable known as a probe is utilized to rank the features in terms of Gram-Schmidt orthogonalization.

The noticeable drawback of the filter approach is the disregard of interaction between the classification algorithm and filter algorithm; that is, the feature subset space search is separated from the search in the hypothesis space. Since filter approaches did not rely on the learning algorithms, the selected features subset may not yield optimal subset, such that a redundant subset could be attained respectively (Zhang, Nie, Li & Wei, 2019). Brief descriptions of common filter techniques are as follows:

(i) **Correlation-based Feature Selection (CFS):** is one of the multivariate filter algorithms that does feature subsets ranking by applying correlation based heuristic evaluation function (Hall & Smith, 1998). The evaluation function usually biases toward the subsets that contain features that are uncorrelated within themselves but

Table 2
Summary of other filter techniques used in microarray data (Bolón-Canedo et al., 2014).

Method	Original Reference	Type of evaluations (R/S)	Type of Data (B/M)
Backward Elimination using Hilbert-Schmidt Independence Criterion (HSIC) (BAHSIC)	(Song et al., 2012)	R	M
Discritizer+filter	(Bolón-Canedo, Sánchez-Maroño & Alonso-Betanzos, 2010; Sánchez-Maroño, Alonso-Betanzos, García-González & Bolón-Canedo, 2010)	S	M
Entropic Filtering Algorithm (EFA)	(Navarro & Muñoz, 2009)	S	B
Monotone dependence	(Bolón-Canedo, Sánchez-Maroño & Alonso-Betanzos, 2010)	R	M
Multi-task feature selection filter (MFS)	(Lan & Vucetic, 2011)	R	M
Matrix of Average Sub-Subset Information for Variable Elimination (MASSIVE)	(Meyer, Schretter & Bontempi, 2008)	R	M
maximum weight and minimum redundancy (MWMR)	(Wang, Song, Xu & Zhou, 2013)	R	B
Partial Least Squares (PLS)	(Student & Fujarewicz, 2012)	R	M
Relevance Feature Selection (RFS)	(Ferreira & Figueiredo, 2012)	R	M
robust feature selection method (RFS)	(Nie, Huang, Cai & Ding, 2010)	R	M
Relevance Redundancy Feature Selection (RRFS)	(Ferreira & Figueiredo, 2012)	R	M

Note: R/S for ranker/subset and B/M for binary/multiclass.

demonstrated a high correlation with the class. Irrelevant features are eliminated as a result of their low correlation with the class and also, redundant features are discarded since they are highly correlated with one or more of the selected features. The choice of a selected feature depends on to what level in which it can predict classes in areas of the instance space not already predicted by other features.

(ii) **The Fast Correlation-Based Filter (FCBF) technique:** is a multivariate algorithm that evaluates feature-class and feature-feature correlation (Yu & Liu, 2003). FCBF begins selection using Symmetrical Uncertainty (SU) to select a set of features that are highly correlated with the class. SU is described as the ratio between the information gain and the entropy of two features. Also, it employs three heuristic approaches to eliminate the redundant features and retain those features which are more pertinent to the class. FCBF is suitable for high-dimensionality data and has been proven to be efficient in eliminating both irrelevant and redundant features. Though, it neglects interaction between features.

(iii) **The INTERACT algorithm:** utilizes the same SU goodness measurement employed by the FCBF, but in addition to that, it includes the consistency contribution, which reveals how significantly the exclusion of a feature will likely affect consistency (Zhao & Liu, 2007). The algorithm entails two major steps. The first step involves the ranking of features in descending order based on their evaluated SU values. While the second step performs the evaluation of features consistency by individually estimating the features consistency contribution of the feature beginning from the last on the ranked feature list and it selects only those features whose consistency contribution exceeds an established threshold and discard those with the less value. The authors established that this technique takes care of feature interaction and effectively selects relevant features.

(iv) **Information Gain:** is one of the famous attribute evaluation approaches that compute mutual information for each class and attribute of the data. It is a univariate filter which provides orderly ranked features with the guide of a threshold (Kira & Rendell, 1992). The selection of feature is done based on those features with a positive information gain value.

(v) **ReliefF:** is one of the famous multivariate filters based on nearest neighbors (Kononenko, 1994). Its original version is known as Relief algorithm developed by Kira and Rendell (1992). The Relief algorithm works by sampling an instance randomly from the data and seeking for nearest neighbors from the same class and other class present in the data. The nearest neighbor values of the selected attributes are compared to the sampled instance and then, utilized it to update the relevance scores for each attribute. The basis is that an informative attribute should be distinguished between instances from different classes and have identical value with instances of the same class. ReliefF has the capability of efficiently solving the problem of multiclass and also more robust and proficient in dealing with missing and noisy data. This algorithm can be employed in all circumstances with low bias; it accommodates features interaction and captures local dependencies which are lacking in other approaches. The weakness of the ReliefF algorithm includes the process of choosing a threshold. In Acuna, Coaquira and Gonzalez (2003), compared ReliefF with other wrapper techniques under different datasets.

(vi) **The minimum Redundancy Maximum Relevance (mRMR) algorithm:** chose features with the highest relevance to the target class and also with minimal redundancy (Peng, Long & Ding, 2005); that is, it chooses features that have maximum dissimilarity to each other. Both criteria employed are based on mutual information that is, maximum-relevance and minimum-redundancy.

(vii) **Consistency-Based Filter:** is a multivariate method which selects subsets of the features, but the selection is done based on the degree of consistency with the class (Dash & Liu, 2003). The acceptability of the selected feature rate is determined using an inconsistency criterion. Summary of other Filter techniques provided in Table 2 based on the evaluation approach used (either rank or subset) and type of data.

4.2.2. *Wrapper techniques*

Evolutionary approaches were normally used to lead their searches in wrapper techniques. It usually commences with a population of solutions that contains a subset of the features. Then each subset would be computed using a learner algorithm to allot fitness to each subset. Generally, the feature subsets selection is improved by an iterating process to come out with optimal solutions. Some of the famous wrapper methods found in the literature are Ant Colony Optimization (Yu et al., 2009), Distance Sensitive Rival Penalized Competitive Learning – Support Vector Machine (ADSRPCL-SVM) (Xiong, Cai & Ma, 2008), Artificial Bee Colony (ABC) algorithm (Garro, Rodríguez & Vázquez, 2016; Hancer, Xue, Karaboga & Zhang, 2015), Genetic algorithm with SVM (Wang et al., 2011), Particle Swarm Optimization (Kar, Das Sharma & Maitra, 2015) and Genetic Programming (employed to predict alternative mRNA splice variants) (Vukusic, Grellscheid & Wiehe, 2007). The wrapper technique's performance is usually better than filter techniques since wrapper techniques allow interactions between the solutions and predictors. While filter methods address the challenges of seeking a better feature subset individually for the model selection phase, wrapper

Table 3
Summary of wrapper techniques used in microarray (Bolón-Canedo et al., 2014).

Method	Original Reference	Type of Evaluation (R/S)	Type of Data (B/M)
GA-KDE-Bayes	(Wanderley, Gardeux, Natowicz & de Pádua Braga, 2013)	S	B
SFS	(Sharma, Imoto & Miyano, 2011)	S	M

Table 4
Summary of embedded techniques used in microarray data (Bolón-Canedo et al., 2014).

Method	Original Reference	Type of Evaluation (R/S)	Type of Data (B/M)
FRFS	(Wang, Song, Xu & Zhou, 2013)	S	M
IFP	(Canul-Reich et al., 2012)	S	B
KP-SVM	(Maldonado, Weber & Basak, 2011)	S	M
PAC-Bayes	(Shah, Marchand & Corbeil, 2011)	R	B
Random Forest	(Anaissi, Kennedy & Goyal, 2011)	S	M

approaches integrate the model assumption search within the feature subset search. In this approach, possible feature subsets space is defined in a search procedure, and many features subsets are generated and weighed. The particular features subset evaluation was attained by training and testing the subset with a specific classification algorithm, making this method designed for a particular learning algorithm. The search algorithm is enfolded around the classification technique to explore all the feature subset space. Though, as the feature space subsets upsurge exponentially with the number of features, heuristic search approaches are employed to achieve the search for the best subset. These search techniques can be categorized into two groups: deterministic and randomized search techniques. Among the benefits of wrapper techniques is the capability to take feature dependencies into consideration and existence of an interaction between feature subset search and model selection. A prominent setback for these approaches is that they are prone to overfitting than filter methods due to the few sample size of Microarray data, and are characterized with high computational expenses, particularly if modelling the classifier requires high computational cost.

Wrapper method did not receive much attention as received by the filter approaches, as a result of its high computational cost. As the feature number increases, the feature searching space of subsets upsurges exponentially. This sometimes becomes critical when the features grow to tens of thousands or more. For this reason, the wrapper method has been mostly avoided in the literature.

Most of the works that utilized the wrapper method can be found in the early years of investigating Microarray data. In a classical wrapper, a search is conducted in the space of genes and the goodness of each gene subset is determined by evaluating the accuracy achieved by the specific learning algorithm. This learning algorithm is trained only with the selected genes. For instance, Ref. (Inza, Sierra, Blanco & Larrañaga, 2002) employed typical wrapper techniques such as floating selection, sequential forward and backward selection, and best-first search on three Microarray datasets. Another case can be found in Ruiz, Riquelme and Aguilar-Ruiz (2006) that make uses of an incremental wrapper named BIRS (Best Incremental Ranked Subset) for the selection of genes. Though the application of wrappers on Microarray data is not well embraced compared to the other feature selection approaches. Other examples were found in recent years.

Ref. (Sharma, Imoto & Miyano, 2011) presented an approach called Successive Feature Selection (SFS). It is a famous conventional feature selection algorithms that trying to overcome the weakness of individual ranking and forward selection schemes. The proposed SFS firstly arrange the features into smaller block partitions. From each of the blocks, peak features were obtained based on their classification accuracy values and then compared to select the optimal feature subset. This technique yields high classification accuracy on many DNA Microarray datasets.

Ref. (Wanderley, Gardeux, Natowicz & de Pádua Braga, 2013), presented an evolutionary wrapper technique called Genetic Algorithm-Kernel Density Estimation (GA-KDE-Bayes). It employs a Bayesian classification algorithm and a non-parametric density estimation technique. The authors explained that non-parametric techniques are good alternatives for scarce and sparse data, such as the bioinformatics analysis because assumptions about its structure are not predefined and all the information originates from the data itself. The results of their works on six Microarray datasets revealed their approach performed better. Table 3 shows the summary of the wrapper technique.

4.2.3. *Embedded techniques*

Despite its lower computational time, the major drawback of the filter method is that it lacks interaction with the classifier, mostly leading to poor performance of classifier when compared with those obtained with wrappers. Though, wrapper technique is also characterized by a huge computational cost with Microarray data. A middle course solution for researchers is the application of embedded approaches, which employ the core of the learning algorithm for ranking of feature criteria. In this approach, optimization techniques are employed (Guyon, Weston, Barnhill & Vapnik, 2002) to rank the features embedded in the projection subspace. The goal of embedded technique (Zhang, Nie, Li & Wei, 2019; Guyon & Elisseeff, 2003) is to minimize the computational time for reclassifying different feature subsets. The integration of feature selection as part of the learning process is the key purpose of embedded techniques. In Battiti (1994), a greedy search algorithm is utilized to evaluate the feature subsets. Support Vector Machine based on Recursive Feature Elimination (SVM-RFE), is one of the most popular embedded technique proposed by Guyon, Weston, Barnhill and Vapnik (2002). It was specifically designed to select genes for cancer classification. Feature selection was done in this approach by repeatedly training SVM classifier with the presented set of features and eliminating the less significant feature as shown by the classifier. Table 4 shows a description of a few embedded methods used for Microarray data analysis.

Ref. (Maldonado, Weber & Basak, 2011) introduced a novel embedded technique that concurrently selects significant features during the classifier modelling by penalizing individual features used in the dual formulation of SVM. This technique is referred to as kernel penalized SVM (KP-SVM) and it enhances an anisotropic radial basis function (RBF) Kernel shape by removing features that have low significance for the learning algorithm. In their experiment, two benchmark Microarray datasets and two real-world datasets were used and it revealed that KP-SVM performed better than the alternative approaches with steadily fewer relevant features.

To solve the problem of data imbalance in some Microarray datasets, Ref. (Anaissi, Kennedy & Goyal, 2011) proposed a new embedded ap-

Table 5
Summary of other hybrid techniques for feature selection used in microarray data (Bolón-Canedo et al., 2014).

Method	Original Reference	Type of Evaluation (R/S)	Type of Data (B/M)
Correlation-based Feature Selection (CFS) and the Taguchi-Genetic Algorithm (CFS-TGA)	(Chuang, Yang, Wu & Yang, 2011)	S	M
Ensemble- Recursive Feature Elimination (RFE)	(Abeel et al., 2010)	S	B
Ensemble of Filters (EF)	(Bolón-Canedo, Sánchez-Maroño & Alonso-Betanzos, 2012)	S	M
Fast clustering-based feature selection algorithm (FAST)	(Song, Ni & Wang, 2011)	S	M
Genetic Algorithm with Dynamic Parameter setting (GADP)	(Lee & Leu, 2011)	S	M
Counting Grid (CG)	(Lovato et al., 2012)	R	B
Multi-Criterion Fusion-based Recursive Feature Elimination (MCF-RFE)	(Yang & Mao, 2010)	S	B
multiple-filter-multiple-wrapper (MFMW)	(Leung & Hung, 2008)	S	B
ReliefF, mRMR and GA (R-m-GA)	(Shreem, Abdullah, Nazri & Alzaqebah, 2012)	S	B
Stratified Random Forest (SRF)	(Ye et al., 2013)	S	M
SVM-RFE with MRMR	(Mundra & Rajapakse, 2009)	R	B

proach based on the random forest algorithm. The approach used different strategies and algorithms to overcome the different problem of complex gene expression of Leukemia dataset. Firstly, a technique is used to seek for the best training error cost for an individual class, and also, to treat data imbalance. Then, the random forest is employed to select the significant features and lastly, a strategy to prevent overfitting is also applied. The experimental outcome proved to be a very acceptable result.

Ref. (Shah, Marchand & Corbeil, 2011) proposed a Probably Approximately Correct (PAC) Bayes feature selection embedded in decision stumps and three other formulations based on different learning principles. PAC-Bayes yields a viable classification performance and at the same time using fewer significant features.

Ref. (Canul-Reich et al., 2012) presented an embedded gene selector called iterative perturbation method (IFP). They applied it on four different Microarray datasets. The least significant features were determined by employing a backward elimination approach and a criterion, which rely on classifier performance of each feature when perturbed by noise. A feature is considered significant if when adding noise produces a great change in the classifier performance. The IFP algorithm results showed similar or superior average class accuracy when compared to the SVM-RFE on three out of the four datasets used.

Ref. (Wang, Song, Xu & Zhou, 2013) presented a First Order Inductive Learner (FOIL) Rule-based feature subset selection algorithm (FRFS). This technique in the first instance produces the FOIL classification rules using a modified propositional implementation of the FOIL algorithm. It then merges the subset features that was obtained in the antecedents of rules and achieves a candidate feature subset that has eliminated redundant features and preserves the interactive and informative ones. In the last stage, it evaluates the significance of the features in the selected feature subset by their new metric called Cover Ratio and detects and eliminates the irrelevant features.

4.2.4. Hybrid techniques

Today, the trend has moved away from using only typical FS techniques (such as filters, wrappers and embedded) but now focusing on combinations of typical FS techniques such as hybrid or ensemble methods. Hybrid approaches typically combine two or more FS algorithms of diverse conceptual origin serially (linear combination).

The hybrid technique is a novel approach which tries to seek the benefits of both filter and wrapper techniques. A hybrid method uses the combination of independent test and performance estimation function of the feature subset. Hybrid approaches are appropriate for high-dimensional data (such as Microarray dataset) in order to reduce the time complexity of selecting relevant features from the dataset. The notion behind the hybrid approach is that the filter approach is used to remove irrelevant features (dimensionality reduction) from the original dataset and aimed to attain a trade-off between the time complexity and feature space size. Then, the wrapper technique is employed to find the best features subset from the selected feature pool. This approach speeds up feature selection because the filter technique swiftly reduces the irrelevant features from the dataset. Advocates of hybrid techniques claim that the chance of removing fantastic predictors by filter approaches is very low if the filter threshold point for a ranked list of features is set low. Some common hybrid techniques include: Chi-Squared statistics with Genetic Algorithm (GA) (Lee & Leu, 2011), information gain with a novel memetic algorithm (Zibakhsh & Abadeh, 2013), a novel similarity scheme with ABC (Hancer, Xue, Karaboga & Zhang, 2015), MRMR with GA (El Akadi, Amine, El Ouardighi & Aboutajdine, 2011), hybridization of GA and SVM (Li et al., 2005), correlation-based feature selection algorithm and genetic algorithm (Nguyen & Rocke, 2002), t-statistics and GA (Mitchell, 1997), principal component analysis (PCA) and Ant Colony Optimization (ACO) algorithm (Nguyen & Rocke, 2002), mutual information and GA (Narayanan, Keedwell & Olsson, 2002), chi-square approach and multi-objective optimization algorithm (Nguyen & Rocke, 2002). Recently, discrete wavelet transform (DWT) and modified genetic algorithm had been used by Saeid, Nossair and Saleh (2020), Entropy and Signal-to-Noise Ratio (EnSNR) by Hengpraprohm and Jungjit (2020). Also, (Saqib, Qamar, Khan & Aslam, 2020) proposed an hybrid of multiple filters and GA wrapper based approach (MF-GARF) which employed Random forest to compute fitness function of the features. Table 5 depicts a summary of other hybrid methods available in the literature.

5. Review of hybrid approach

In recent times, many researchers proposed hybrid approaches taping the advantages of both filter and wrapper methods.

Ref. (Wang, Chu & Xie, 2007) presented a technique that involves two main steps of Microarray data analysis. In the first step, T-test and Class Separability (CS) scoring were used for genes selection and the second stage involves classification of genes using Fine classifier. Divide and conquer technique were employed to achieve better accuracy performance. They used two different datasets in this research, Lymphoma Data and SRBCT Data. They employed a KNN algorithm to handle missing values in Microarray data. The best marker genes selected were passed one after the other to the classifier until better accuracy was attained. Also, they compared the result they obtained with a KNN and SVM classifier.

Ref. (Rangasamy, 2009) developed a model that ranked the genes based on a classical statistical approach and two different learning algorithms. The researchers utilized the datasets that contained two classes like Liver and Leukemia datasets, and the one that had more than two classes in the dataset such as Lymphoma. They used Analysis of Variance (ANOVA) and Linear Discriminant Analysis (LDA) for genes selection and SVM-one-against-all (SVM-OAA) and Radial Basis Function (RBF)

Kernel-based on suitability of dataset type. The learning algorithm was evaluated utilizing all possible gene combinations and the optimal gene combination was reported. Ref. (Martín-Merino & De Las Rivas, 2009) employed a Kernel Alignment KNN for cancer classification using gene expression profiles. The Kernel alignment KNN achieves better performance when compared with other machine learning strategies and enhances the performance of classical KNN.

Ref. (Revathy & Amalraj, 2011) proposed a method that hybridizes the enrichment score and SVM classifier for cancer classification in Microarray data. The dataset is randomly partitioned into train and test dataset. The relevant genes were selected using a ranking approach and then the selected genes were passed into the classifier one after the other. But, if better accuracy was not achieved, the gene combination is performed on the ranked dataset. The report of their work shows that the combination of SVM and enrichment score performed better than SVM with T-Score.

Ref. (Ghorai, Mukherjee, Sengupta & Dutta, 2010) developed a hybrid computer-aided diagnosis (CAD) framework based on filters and wrapper techniques for cancer classification using Microarray gene expression profiles. Minimum redundancy maximum relevance (MRMR) ranking algorithm was utilized for feature selection and wrapper technique based on nonparallel plane proximal classifier (NPPC) was applied on selected genes to reduce the computational burden.

Ref. (El Akadi, Amine, El Ouardighi & Aboutajdine, 2011) presented a framework that comprises of a two-stage feature selection approach for Microarray data. The first stage involves the filtering of genes using Minimum Redundancy Maximum Relevance (MRMR). While GA is used in the second stage to generate the gene subsets and both Naïve Bayes (NB) and support vector machine (SVM) were employed for the classification and assessment analysis.

Ref. (Rajeswari & Reena, 2011) proposed an approach for tumor cells identification. They used Support Vector Machine (SVM) and Fuzzy Neural Network (FNN) to solve the classification problem. Liver cancer dataset was utilized to analyses the proposed method. Their results revealed that the proposed method could diagnose cancer diseases considerably when it was compared with the conservative procedures of a cancer diagnosis. Fuzzy Neural network recorded performance accuracy between 92-96%.

Ref. (Sahu & Mishra, 2012) presented a novel hybrid feature selection method for high dimensional cancer Microarray dataset before performing classification. The authors ranked the genes using the filtering method of signal-to-noise ratio (SNR) score and optimization technique was carried out using Particle Swarm Optimization (PSO) for dimensionality reduction. K-nearest neighbor (KNN), Probabilistic Neural Network (PNN) and Support vector machine (SVM) were utilized as classifiers. PNN recorded 96% classification accuracy. Ref. (Swathi, Babu, Sendhilkumar & Bhukya, 2012) designed default Adaptive Resonance Theory (ART1) network model for breast cancer detection. This model consisted of three stages which included recognition, comparison and search phases. In their work, the Backpropagation algorithm was applied to minimize the error and training of network architecture. The experimental results revealed that ART1 was an efficient network in classifying cancer dataset. The model achieved better results for unsupervised machine learning breast cancer datasets with 92% accuracy. Ref. (Dev, Dash, Dash & Swain, 2012) designed three different algorithms for cancer tumor diagnosis. In this research work, Backpropagation network (BPN), Functional Link Artificial Neural Network (FLANN) and PSO-FLANN were employed for breast cancer classification. In this work, the outcomes of the three classifiers were examined and compared. They reported 56.12% Classification rate for BPN and 63.34% FLANN, while PSOFLANN gave the best classification rate of 92.36%.

Ref. (Abeer, Basma, El-Sayed & Abdel-Badeeh, 2013) developed differentially expressed genes (DEGs) in Microarray data with the aim of building a reliable and robust classifier. A t-test was used for feature selection and KNN for machine learning classifier. The dataset used was Lymphoma dataset to achieve the DEGs and to analyze the influence of these on genes learning accuracy performance respectively.

Ref. (Shreem, Abdullah & Nazri, 2014) applied a new approach of hybridization that combines Harmony Search Algorithm (HSA) and the Markov Blanket (MB), termed HSA-MB for gene selection in classification tasks. HSA employs naive Baye's classifier in its wrapper approach. Ten different Microarray datasets were employed to perform the experiments. The HSA-MB performance showed similar results with available results in the literature.

Ref. (Abeer & Basma, 2014) proposed a new fusion machine learning (ML) reduction method to improve the cancer classification accuracy of Microarray data using two different gene ranking methods: *t*-test and Class Separability (CS). They combined the genes ranking methods with two ML classifiers - KNN and SVM; to evaluate Microarray gene expression profiles. Four public cancer Microarray (Lymphoma, Leukemia, SRBCT, and Lung) datasets were utilized to measure the proposed approaches, and they succeeded in accomplishing their aim. The approach of selecting genes only used for training samples and completely excluding the testing samples from the process of classifying was employed to obtain better performance and validate the results. The combination of t-test and SVM was reported to have produced higher classification accuracy compared to other combined approaches of (t-test + KNN, CS + SVM and CS + KNN), where they reported classifier accuracy of 100% using peak ranked value of 2 and 3 genes from Lymphoma and SRBCT data respectively. They also obtained 94.12% from Leukemia using 6 genes as peak ranked value. Their results revealed that the proposed reduction approach reached favorable results of the number of genes complemented to the learner classifiers at the same time with the accuracy of the classifier.

Ref. (Alshamlan, Badr & Alohali, 2015) presented a new hybrid gene selection method called Genetic Bee Colony (GBC) algorithm, where a hybrid of both Genetic Algorithm and Artificial Bee Colony (ABC) algorithms were used. The GBC algorithm demonstrates the sign of having a better performance as it attains very high classifier accuracy with the small size number of genes.

Ref. (Doreswamy & UmmeSalma, 2016) designed a Binary Bat Algorithm (BBA) based Feedforward Neural Network (FNN) hybrid model. The authors exploited the advantages of BBA and efficiency of FNN to classify three standard breast cancer datasets into malignant and benign cases. In this work, BBA was utilized to produce a V-shaped hyperbolic tangent function for training the network and a fitness function was employed for error minimization. FNNBBA based classification yielded 92.61% accuracy rate for training data and 89.95% for testing data.

Ref. (Alomari et al., 2017) presented a hybrid gene selection of filter technique - Minimum Redundancy Maximum Relevancy (MRMR) and a wrapper technique - Bat algorithm (BA) using Microarray dataset. MRMR was employed to discover the most significant genes from the entire gene in gene expression dataset, and BA was used to search for the most informative gene subset from the reduced set produced by MRMR that aided in detecting cancers. The classifier method used was a support vector machine (SVM) with 10-fold cross-validation which helped in evaluating the BA. To reveal the performance accuracy of the proposed method, three different Microarray datasets were used, which include: colon, Breast, and Ovarian cancer dataset. The same procedure was carried out with Genetic algorithm (GA) to perform the comparison with MRMR-BA and the results revealed that their proposed approach was capable of finding the small size of gene subset with high classification rate. Table 6 presents a summary of other recent related approaches.

6. Open research issues

In analyzing microarray dataset, researchers are posed with serious challenges of computational techniques, due to dimensionality nature of data (few-samples with many features in thousands). Many classical classification techniques usually have hitches in handling such hefty

Table 6
Summary of related approach.

No.	Author	Year	Title of Work	Objective	Method Used	Conclusion	Limitation
1	Mao et al. (Mao, Cai & Shao, 2013)	2013	Selecting Significant Genes by Randomization Test for Cancer Classification Using Gene Expression Data	Performance evaluation of different gene selection methods to determine the best approach among them.	In this work, the authors employed Randomization test (RT) method to select the relevant genes and deal with gene expression data. A statistic derived of the regression coefficients in a series - partial least squares discriminant analysis (PLSDA) model is employed to evaluate the significance of the genes.	The rationality of the results is validated by multiple linear regression (MLR) modelling and PCA by comparing the results obtained. The superiority of the method was proved by the author and therefore, suggested the method as an alternative tool for classification using the expression data.	All datasets used were binary class data, the approach may not cope with multi-class dataset.
2	Santosa et al. (Santos, Datia & Pato, 2014)	2014	Ensemble feature ranking applied to medical data	The study reported an ensemble FS method that can cope with large datasets.	The authors used ensemble FS algorithms that consist of the following: InfoGain, Gain Ratio, Symmetrical Uncertainty and Chi-square. They evaluated selected features with SVM, Bagging using the RPART function (BAG), Random Forest (RF) and Na¨ıve Bayes (NB) learning algorithms on KDD Cup 2008 -Breast Cancer dataset.	They were able to developed ensemble FS by combining 5 different Feature Ranking algorithms to select relevant feature for high dimensional data and their results showed a better classification models.	Only breast dataset was used.
3	Cao et al. (Cao et al., 2015)	2015	A Fast Gene Selection Method for Multi-Cancer Classification Using Multiple Support Vector Data Description	The focus of the study was to use SVM as feature selection	This paper presented a new fast feature selection approach based on multiple SVDD on the multi-class microarray data. Recursive feature elimination (RFE) algorithm was repeatedly used to eliminate irrelevant features. Their proposed approach was tagged multiple SVDD-RFE (MSVDD-RFE). They employed KNN and SVM classifiers to evaluate the performance of their approach.	Based on the authors' submission SVDD cannot withstand the challenges of multiple class dataset, since it only focuses on the target class. Also, it is consuming more time when it is used for gene selection.	The result of the Lung Cancer dataset was unsatisfying and SVDD is time-consuming for gene selection.
4	Bolón-Canedo et al. (Bolón-Canedo, Sánchez-Maroño & Alonso-Betanzos, 2015)	2015	Distributed feature selection: An application to microarray data classification	Deal with the subsets to yield a more balanced feature/sample ratio and avoid overfitting problems and also improve the performance of original feature selection techniques and reduce the computation time.	Distributed filter approach was adopted with 5 different filter algorithms such as InfoGain, ReliefF, INTERACT, Consistency-based Filter, and CFS. They employed 4 classifiers (C4.5, K-NN, SVM and Naive Bayes) to evaluate the performance of their approaches.	The most important advantage of the distributed approach utilized was that it greatly minimizes the execution time while still maintaining accuracy at reasonable levels, and sometimes even improving it.	Test and Train datasets were selected randomly and the number is not uniform for all the datasets used. Also, Binary class datasets were used for the experiment.
5	Mohammadi et al. (Mohammadi, Noghabi, Hodtani & Mashhadi, 2016)	2016	Robust and stable gene selection via Maximum–Minimum Correntropy Criterion	Maximum–Minimum Correntropy Criterion (MMCC) used for selection of informative genes from microarray datasets.	Filter-based feature selection via Maximum–Minimum Correntropy Criterion and SVM with linear kernel as classifier was used with 25 different datasets. Time, Accuracy and stability of feature selected were the metrics used for performance evaluation of the approach	The results revealed that less than 10 genes among thousands of genes can possibly use to achieve significantly higher accuracy in cancer classification.	The Author trade-off accuracy performance with time and memory space.

(*continued on next page*)

Table 6 (*continued*)

No.	Author	Year	Title of Work	Objective	Method Used	Conclusion	Limitation
6	Sharbaf et al. (Sharbaf, Mosafer & Moattar, 2016)	2016	A Hybrid Gene Selection Approach for Microarray Data Classification Using Cellular Learning Automata and Ant Colony Optimization	This study was designed to seek the smallest subset of biomarkers signify the present of disease efficiently.	They employed Hybrid of Fisher criterion as filter approach and wrapper approach which is based on cellular learning automata (CLA) optimized with ant colony method (ACO). The performance was evaluated with K-nearest neighbour; support vector machine and naïve Bayes on four microarray data.	The evaluations of the approach confirmed that the proposed approach select the smallest subset of genes that yielded maximum accuracy.	Their experiment is time-consuming and only used ROC and Accuracy as the performance metrics.
7	Gao et al. (Gao, Ye, Lu & Huang, 2017)	2017	Hybrid Method Based on information Gain and Support Vector Machine for Gene Selection in Cancer Classification	This study was designed to employed hybrid of InfoGain-SVM as feature selection technique.	They proposed a hybrid gene selection algorithm of Information Gain-Support Vector Machine (InfoGain-SVM). InfoGain was applied first, to filter irrelevant and redundant genes from the dataset. Then, further removal of redundant genes was done using SVM to remove the noise in the datasets in more effective way. Finally, LIBSVM classifier was used to evaluate the performance of the approach.	IG-SVM achieved a classification accuracy of 90.32% for colon cancer, which is difficult to be accurately classified, only based on three genes including CSRP1, MYL9, and GUCA2B	Only Binary class dataset was used and the approach may not likely cope with the multi-class dataset.
8	Passi et al. (Passi, Nour & Jain, 2017)	2017	Markov Blanket: Efficient Strategy for Feature Subset Selection Method for High Dimensional Microarray Cancer Datasets	Comparison of Markov blanket model with other feature subset selection methods.	In this work, two techniques were used. The first one is dependent on the Memetic Algorithm (MA), while the second one is dependent on MRMR for feature subset selection methods. Both techniques were hybridized with GA. Also, four common classifiers were employed (Lazy-nearest neighbour, linear SVM, Naïve Bayes, and AdaBoost ensemble) to evaluate both approaches. They compared the performance of both the techniques with Bayesian Network-based Markov Blanket for feature selection method and KNN classifier.	The main contribution of this paper was to show the capability of the Bayesian network and Markov Blanket for feature subset selection and learning classification models. The outcomes of the experiment shown that the approach can build high quality and effective learning classification algorithms.	Only Running time and Average test Error were used as performance metrics and the approach perform better only in binary class datasets.
9	Dashtban and Balafar (Dashtban & Balafar, 2017)	2017	Gene Selection for Microarray Cancer Classification Using a New Evolutionary Method Employing Artificial Intelligence Concepts	The purpose of this study is to develop hybrid of GA for feature selection.	They proposed a new evolutionary approach based on genetic algorithms and artificial intelligence to select informative genes for cancer classification. A filter technique was applied first, to perform dimensionality reduction of feature space then, followed by applying an integer-coded genetic algorithm with dynamic-length genotype, intelligent parameter settings, and modified operators. Three classifiers -SVM, Naïve Bayes and KNN were employed to evaluate the performance of their proposed algorithm.	Their approach practically explored the concept of random restart hill climbing to prevent biasness or wrong initializations by initiating the population of few individuals with randomly generated length. The proposed technique attained the best results when combine with Fisher score and SVM.	Most of the dataset used have features that are less than 10,000 and there are lot of dissimilarity among the genes selected (stability is in question).

(*continued on next page*)

Table 6 (*continued*)

No.	Author	Year	Title of Work	Objective	Method Used	Conclusion	Limitation
10	Dash (Dash, 2017)	2017	A Two-Stage Grading Approach for Feature Selection and Classification of Microarray Data Using Pareto Based Feature Ranking Techniques: A Case Study	Identify of a model that may not perform better in all the cases but must yield a better result consistently when comparing to majority of the models	The author proposed a bi-objective ranked based Pareto front technique. Performances of the models are evaluated using three classifiers, such as ANN, NB and KNN.	The model performance changes with the datasets and classifier. Thus a two-stage grading approach is used to obtain a superior optimization model.	All the datasets used were binary Class and most of the datasets have features that are less than ten thousand.
11	Ebrahimpour et al. (Ebrahimpour, Nezamabadi-Pour & Eftekhari, 2018)	2018	CCFS: A Cooperating Coevolution Technique for Large Scale Feature Selection on Microarray Datasets	Proposing a Cooperative Coevolving (CC) version of Binary Gravitational Search Algorithm (BGSA).	The authors employed Cooperative Coevolving feature selection (CCFS) for feature selection and C4.5 and Naïve Baye as classifiers with six different microarray datasets	The proposed approach can be applied on hefty datasets without any pre-processing and It converges to near-global optimum since Cooperative Coevolving approach to do the searches.	All the datasets used were binary Class and redundancy data is not taken care of.
12	Maldonado and L´opez (Maldonado & López, 2018)	2018	Dealing with high-dimensional class-imbalanced datasets: embedded feature selection for SVM classification	The proposed method was designed to deal with class-imbalance and high dimensionality issues in machine learning.	They presented embedded strategy penalizes cardinality of the feature set via the scaling factors technique, and used the approach with two SVM formulations (Cost-Sensitive SVM, and Support Vector Data Description) designed to solve the problem of class-imbalanced,. Their proposed concave formulations are solved through a Quasi-Newton update and Armijo line search.	They suggested that KP-CSSVM is an outstanding alternative for feature selection and class-imbalance classification. Since the technique was designed to solve the class-imbalance problem and penalizes the use of features by fine-tuning the kernel width in repetitive process.	AUC was the only metric used and datasets used were all binary Class data.
13	Dashtbana et al. (Dashtbana, Mohammadali & Prashanth, 2018)	2018	Gene Selection for Tumor Classification Using a Novel Bio-inspired Multiobjective Approach	Design of multi-objective version of bat algorithm for binary variable selection, that incorporating social learning concepts in local search strategies	They presented a hybrid approach using the Fisher criterion and the traditional Bat Algorithm with sophisticated formulations, effective multi-objective operators, and new local search strategies employing social learning ideas in manipulating random walks. The accuracy of selected subset was appraised using four different classifiers which include SVM, KNN, NB and Decision Tree (DT)	Experimental results revealed that new combinations of predictive biomarkers have an association with other studies. Though, in prostate cancer dataset, the proposed approach attained a very good accuracy with a few number of genes.	The performance of the SRBCT dataset is not too good because it is only multiclass dataset used for the experiment.
14	Dash (Dash, 2018)	2018	An Adaptive Harmony Search Approach for Gene Selection and Classification of High Dimensional Medical Data	Design a hybrid of harmony search and Pareto Optimization to select the most relevant genes from the datasets.	The author proposed a hybridized harmony search and Pareto optimization method for feature selection in high dimensional data classification problem. In the first stage, an adaptive harmony search algorithm is used for gene selection with probability distribution factor for optimal gene ranking. The selection approach is further refined and applying a bi-objective Pareto based feature selection technique to select an optimal minimum number of top-ranked genes. Four classifiers (ANN, KNN, NB and SVM) are used to evaluate the approach	As the presence of relevant genes in selected features the classification error is minimizes, therefore, it is employed as a fitness function for feature evaluation. The classification results show that the performance of the proposed model with SVM classifier is better than that of the results obtained by using other classifiers.	All datasets used were binary class.

data. Also, there are other experimental impediments that make the microarray data analysis a fascinating domain.

6.1. Small sample size

The most famous and focused problem that many researchers trying to deal with in microarray data is centred on a few sample size (most of the time less than 100). The main issue in this scenario is that error estimation is greatly affected by small samples of data (Dougherty, 2001). Without proper estimation of error (selecting a robust algorithm), the training classification algorithms tend to perform poorly which has resulted in a large publication of uncorroborated scientific hypotheses (Braga-Neto, 2007). For instance, it was reported by Ref. (Michiels, Koscielny & Hill, 2005) that when re-run experiment was conducted with the seven famous published microarray data in order to predict cancer patients, the studies showed that, five out of the seven datasets did not classify patients better. To solve this challenge, it is crucial to choose a correct validation method for computing the classification error.

6.2. Class imbalance

Class imbalance is another common phenomenon in microarray data. This situation arises when a dataset is prevailed by a major class as a result of having more instances than the other minority classes in the data (He & Garcia, 2009; Sun, Wong & Kamel, 2009; López et al., 2013). Normally, most of the researchers usually have more curiosity for learning rare classes. For example, in the cancer domain, cancer class usually has fewer numbers than the non-cancer class because ordinarily, the population of healthy patients are more than cancer patients. In this situation, typical classification algorithms may tend to be biased toward the classes with the highest number of samples, since the classification rules for predicting those high instances are favored positively in performance accuracy, while ignoring the predicting rules that predict minority class samples and sometimes consider them as noise, since, classification algorithms prefer more general rules. Thus, instances from minority class are often misclassified compared to those instances from the other classes (Galar, Fernández, Barrenechea & Herrera, 2013).

Though the hindrance in learning task is not caused by class imbalance itself, it is due to some complications associated with this problem, such as a few sample size, as in the case of microarray data. Prominent examples of unbalanced microarray datasets are multiclass microarray datasets. This challenge is very prominent when the imbalance in the test set is more obvious than in the training set.

The typical preprocessing approaches usually employed to overcome this challenge are undersampling and oversampling techniques. The former approach generates a subset from the original dataset by eliminating some instances belonging to major class; while the latter approach creates a superset by replicating some instances of the original dataset or generating new instances from original data. In some scenarios, a hybrid of both techniques can be employed. One of the famously used resampling methods is Synthetic Minority Oversampling Technique (SMOTE) (Chawla, Bowyer, Hall & Kegelmeyer, 2002). It works by over-sampling the minority class by introducing synthetic examples into each minority class sample along the segment lines that join all (or any) of the k minority class nearest neighbors. This technique was applied in Hambali and Gbolagade (2016), Lusa (2012) on microarray data, although Ref. (Lusa, 2012) identified that it does not mitigate the bias towards predicting the majority class for most classifiers. Recently, an ensemble of classifiers has been in consideration for a promising solution to the class imbalance problem, enticing great attention among researchers (Galar, Fernández, Barrenechea & Herrera, 2013; Galar et al., 2011), in many cases joined with preprocessing methods such as SMOTE. Ensemble-based approaches have been established for enhancing the results obtained by using data preprocessing methods and training with a single classification algorithm (Galar, Fernández, Barrenechea & Herrera, 2013).

6.3. Data complexity

Data complexity are current issues that measure the characteristics of the data, considered to be problematic in classification tasks, such as overlapping amidst the classes, the linearity of the decision margins (that is, their separability) (SáEz, Luengo & Herrera, 2013; Ho & Basu, 2002). These measures were applied to gene expression analysis in Saeys, Inza and Larrañaga (2007), Lorena, Costa, Spolaôr and De Souto (2012). It was revealed that low complexity associates to small prediction error. Precisely, the class overlapping measurement, such as F1 (maximum Fisher's discriminant ratio) (SáEz, Luengo & Herrera, 2013), measures the efficacy of an individual feature dimension in distinguishing between the classes. They investigated the weighted values' range and spread of the dataset within each class and seek for overlapping among different classes.

6.4. Dataset shift

Dataset shift takes place as a result of dividing original datasets into training and test sets. This arises when the holdout (test set) data experience a situation where the class distribution of inputs and outputs varies between training and test phases (Moreno-Torres, Sáez & Herrera, 2012). As a result, the common notion that the training and test data contain identical distributions is usually violated in the real-world scenarios, which could hamper the process of gene selection and classification. For instance, Leukemia (Golub et al., 1999), Lung (Gordon et al., 2002) and Prostate (Singh et al., 2002) datasets come with different training and test sets. In Lung dataset, there is a vast disparity in the distribution of classes, 50–50% in the training set and 90–10% in the test set. The machine learning algorithms usually encounter a lot of challenges in dealing with Prostate dataset because the test and training datasets were extracted from a different experiment; the test data has an approximately 10-fold dissimilarity in overall microarray intensity from the training data (Bolón-Canedo, Sánchez-Maroño & Alonso-Betanzos, 2010). The test distribution is 26–74% which is significantly different from the distribution of train data which is 49–51% and with unsuitable feature selection algorithms, certain classifiers just allot all the samples to one of the classes (Bolón-Canedo et al., 2014; Bolón-Canedo, Sánchez-Maroño & Alonso-Betanzos, 2010). Dataset shift may also occur when applying a cross-validation approach, which splits the entire training set into different subsets of data (Moreno-Torres, Sáez & Herrera, 2012).

6.5. Outliers

One of the neglected vital aspects that has not gained much attention in the literature is the detection of outliers (Hardin & Rocke, 2004) in microarray data. In microarray, there are some datasets whose samples are wrongly labeled or recognized as probable considered to be contaminated, which in fact should be tagged as outliers. Since, they have a negative effect, on the selection of informative genes from the sample. Ref. (Kadota, Tominaga, Akiyama & Takahashi, 2003) developed a method that detected some outlying samples in Colon dataset. Hence, analysis of samples for detection of outliers ought to be considered as a pre-processing phase in the microarray datasets classification, since they may have an undesirable effect on the feature selection and as well as on the final classification (Navarro & Muñoz, 2009).

6.6. Difference characteristic with the same name

There are lots of microarray data available in a different repository with the same name but with a different number of classes, features and instances (samples). For example, Lung dataset in Garber et al. (2001) has 52 instances with 918 features, in Gordon et al. (2002), their Lung data contains 181 samples with 12,533 features while the same Lung data in Shedden et al. (2008) has 410 samples and 2,428 features. Also, the same dataset in Zhu, Ong and Dash (2007), contains

203 samples with 12,600 features. All the first three datasets are binary class while the last data has 5 different classes. This situation makes it difficult to be able to compare or have a standard measure for different microarray experiment analysis. If the information about the composition of the dataset is not carefully looked into, it may result into publication of uncorroborated scientific hypotheses as a result of comparing different datasets with the assumption that they are the same.

7. Conclusion

This study reviews recent and up-to-date contributions to feature selection techniques in cancer Microarray data. The importance of feature selection as a vital pre-processing tool which is not only limited to the reduction in the input number of features but also reduces the computational time and memory; help in augmenting classification accuracy and lot more was presented. The advent of DNA microarray data has posed a lot of challenges to machine learning research, due to its nature of large dimensional feature with few sample size. Apart from the noticeable drawback of having too many features for a number of small samples, researchers also have to cater for unbalanced classes that characterize the data, testing and training datasets extracted under different situations, the presence of outliers (dataset shift). With all these reasons, new techniques keep on emerging every year, not only limited to the intention of improving previous approaches classification accuracy results but also attempt to aid biologists to detect and understand the fundamental mechanism that connects the gene expression to diseases.

Copious and successful efforts have been employed in the past few years to utilize feature selection to solve these problems, which primarily can be divided into three main techniques; filter, wrapper and embedded approaches. Filter approaches are the most employed methods because of the great computational resources required by these datasets; wrapper and embedded approaches have tactically been avoided. The directions of researches on feature selection are going toward hybrid feature selection techniques. These methods have demonstrated improvement in robustness of the selected genes and subsequently enhance better performance accuracy of the cancer classification model. Furthermore, it also indicates an intensive drive toward applying feature selection in the combination of heterogeneous data source such as integration of genomic and proteomic or microarray and clinical data, instead of analyzing cancer classification in a separate experiment, which can expose the significance of biological information.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research has been financially supported in part by Tertiary Education Trust Fund (TETFUND) with Reference FUW/REG/T.5/VOL.1/T11.

References

Abadi, M., et al. (2016). Tensorflow: A system for large-scale machine learning. In *Proceedings of the twelfth {USENIX} symposium on operating systems design and implementation ({OSDI} 16)* (pp. 265–283).

Abdel-Zaher, A. M., & Eldeib, A. M. (2016). Breast cancer classification using deep belief networks. *Expert Systems with Applications, 46*, 139–144.

Abeel, T., Helleputte, T., Van de Peer, Y., Dupont, P., & Saeys, Y. (2010). Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics, 26*(3), 392–398.

Abeer, M. M., & Basma, A. M. (2014). A hybrid reduction approach for enhancing cancer classification of microarray data. *International Journal of Advance Research in Artificial Intelligence, 3*(10).

Abeer, M. M., Basma, A. M., El-Sayed, M. E., & Abdel-Badeeh, M. S. (2013). Applying a statistical technique for the discovery of differentially expressed genes in microarray data, pp. 220–227.

Acuna, E., Coaquira, F., & Gonzalez, M. (2003). A comparison of feature selection procedures for classifiers based on kernel density estimation. In *Proceedings of the international conference on computer, communication and control technologies, CCCT: 3* (pp. 468–472).

Acunzo, M., Romano, G., Wernicke, D., & Croce, C. M. (2015). MicroRNA and cancer – A brief overview. *Advances in Biological Regulation, 57*, 1–9.

Aiello, S., Click, C., Roark, H., Rehak, L., & Stetsenko, P. (2016). Machine learning with python and H2O, Ed. Lanford, J., Publ. by H, vol. 20, p. 2016.

Alejo, R., Monroy-de-jesús, J., Pacheco-sánchez, J. H., López-gonzález, E., & Antonio-velázquez, J. A. (2016). A selective dynamic sampling back-propagation approach for handling the two-class imbalance problem. *Applied Bioinformatics, 6*(200), 1–17. 10.3390/app6070200.

Alomari, O. A., et al. (2017). Mrmr ba: A hybrid gene selection algorithm for cancer classification. *Journal of Theoretical and Applied Information Technology, 95*(12), 15.

Alshamlan, H. M., Badr, G. H., & Alohali, Y. A. (2015). Genetic Bee Colony (GBC) algorithm: A new gene selection method for microarray cancer classification. *Computational Biology and Chemistry, 56*, 49–60.

American Cancer Society. (2017). *Cancer facts and figures 2017* Atlanta.

Anaissi, A., Kennedy, P. J., & Goyal, M. (2011). Feature selection of imbalanced gene expression microarray data. In *Proceedings of the 2011 twelfth ACIS international conference on software engineering, artificial intelligence, networking and parallel/distributed computing* (pp. 73–78).

Augenlicht, L. H., Taylor, J., Anderson, L., & Lipkin, M. (1991). Patterns of gene expression that characterize the colonic mucosa in patients at genetic risk for colonic cancer. *Proceedings of the National Academy of Sciences of the United States of America, 88*(8), 3286–3289.

Augenlicht, L. H., Wahrman, M. Z., Halsey, H., Anderson, L., Taylor, J., & Lipkin, M. (1987). Expression of cloned sequences in biopsies of human colonic tissue and in colonic carcinoma cells induced to differentiate in vitro. *Cancer Research, 47*(22), 6017–6021.

Basgall, M. J., Hasperué, W., Naiouf, M., Fernández, A., & Herrera, F. (2019). An analysis of local and global solutions to address big data imbalanced classification: A case study with SMOTE preprocessing. In *Proceedings of the conference on cloud computing and big data* (pp. 75–85).

Battiti, R. (1994). Using mutual information for selecting features in supervised neural net learning. *IEEE Transaction on Neural Networks, 5*(4), 537–550.

Bekkerman, R., El-Yaniv, R., Tishby, N., & Winter, Y. (2003). Distributional word clusters vs. words for text categorization. *Journal of Machine Learning Research, 3*(Mar), 1183–1208.

Ben-Dor, A., Bruhn, L., Friedman, N., Nachman, I., Schummer, M., & Yakhini, Z. (2000). Tissue classification with gene expression profiles. In *Proceedings of the fourth annual international conference on computational molecular biology* (pp. 54–64).

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transaction on Neural Networks, 5*(2), 157–166.

Bennet, J., Ganaprakasam, C., & Kumar, N. (2015). A hybrid approach for gene selection and classification using support vector machine. *International Arab Journal of Information Technology, 12*.

Bø, T. H., & Jonassen, I. (2002). New feature subset selection procedures for classification of expression profiles. *Genome Biology, 3*(4) research0017-1.

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2010). On the effectiveness of discretization on gene selection of microarray data. In *Proceedings of the 2010 international joint conference on neural networks (IJCNN)* (pp. 1–8).

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2012). An ensemble of filters and classifiers for microarray data classification. *Pattern Recognition, 45*(1), 531–539.

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge and Information Systems, 34*(3), 483–519.

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2015). Distributed feature selection: An application to microarray data classification. *Applied Soft Computing, 30*, 136–150.

Bolón-Canedo, V., Sánchez-Marono, N., Alonso-Betanzos, A., Benítez, J. M., & Herrera, F. (2014). A review of microarray datasets and applied feature selection methods. *Inf. Science (Ny)., 282*, 111–135.

Bolón-canedo, V., Sánchez-maroño, N., Alonso-betanzos, A., Benítez, J. M., & Herrera, F. (2014). A review of microarray datasets and applied feature selection methods, vol. 282, pp. 111–135, doi:10.1016/j.ins.2014.05.042.

Braga-Neto, U. (2007). Fads and fallacies in the name of small-sample microarray classification. *IEEE Signal Processing Magazine, 24*(1), 91.

Brazma, A., & Vilo, J. (2000). Gene expression data analysis. *FEBS Letters, 480*(1), 17–24.

Brown, G., Pocock, A., Zhao, M.-J., & Luján, M. (2012). Conditional likelihood maximisation: A unifying framework for information theoretic feature selection. *Journal of Machine Learning Research, 13*(Jan), 27–66.

Cai, D., He, X., & Han, J. (2007). Semi-supervised discriminant analysis. In *Proceedings of the eleventh international conference on ICCV* (pp. 1–7).

Cai, D., Zhang, C., & He, X. (2010). Unsupervised feature selection for multi-cluster data. In *Proceedings of the sixteenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 333–342).

Cai, R., Hao, Z., Yang, X., & Wen, W. (2009). An efficient gene selection algorithm based on mutual information. *Neurocomputing, 72*(4–6), 991–999.

Canul-Reich, J., Hall, L. O., Goldgof, D. B., Korecki, J. N., & Eschrich, S. (2012). Iterative feature perturbation as a gene selector for microarray data. *International Journal of Pattern Recognition and Artificial Intelligence, 26*(05), Article 1260003.

Cao, J., Zhang, L., Wang, B., Li, F., & Yang, J. (2015). A fast gene selection method for multi-cancer classification using multiple support vector data description. *Journal of Biomedical Informatics, 53*, 381–389.

Caruana, R., & de Sa, V. R. (2003). Benefitting from the variables that variable selection discards. *Journal of Machine Learning Research, 3*(Mar), 1245–1264.
Chang, X., Nie, F., Yang, Y., & Huang, H. (2014). A convex formulation for semi-supervised multi-label feature selection. In *Proceedings of the twenty-eighth AAAI conference on artificial intelligence*.
Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357.
Chen, C.-R., Shu, W.-Y., Tsai, M.-L., Cheng, W.-C., & Hsu, I. C. (2012). THEME: A web tool for loop-design microarray data analysis. *Computers in Biology and Medicine, 42*(2), 228–234.
Chen, D., Qian, G., Shi, C., & Pan, Q. (2017). Breast cancer malignancy prediction using incremental combination of multiple recurrent neural networks. In *Proceedings of international conference on neural information processing* (pp. 43–52).
Cheng, Q., Zhou, H., & Cheng, J. (2010). The fisher-markov selector: Fast selecting maximally separable feature subset for multiclass classification with applications to high--dimensional data. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33*(6), 1217–1233.
Cho, J.-H., Lee, D., Park, J. H., & Lee, I.-B. (2004). Gene selection and classification from microarray data using kernel machine. *FEBS Letters, 571*(1–3), 93–98.
Cho, K. et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation, arXiv Prepr. arXiv1406.1078.
Chow, M. L., Moler, E. J., & Mian, I. S. (2001). Identifying marker genes in transcription profiling data using a mixture of feature relevance experts. *Physiological Genomics, 5*(2), 99–111.
Chuang, L.-Y., Yang, C.-H., Wu, K.-C., & Yang, C.-H. (2011). A hybrid feature selection method for DNA microarray data. *Computers in Biology and Medicine, 41*(4), 228–237.
Cleofas-Sánchez, L., Sánchez, J. S., & García, V. (2019). Gene selection and disease prediction from gene expression data using a two-stage hetero-associative memory. *Progress in Artificial Intelligence, 8*(1), 63–71.
Daelemans, W., Hoste, V., De Meulder, F., & Naudts, B. (2003). Combined optimization of feature selection and algorithm parameters in machine learning of language. In *Proceedings of the European conference on machine learning* (pp. 84–95).
Dash, M., & Liu, H. (2003). Consistency-based search in feature selection. *Artificial Intelligence, 151*(1–2), 155–176.
Dash, R. (2017). A two stage grading approach for feature selection and classification of microarray data using Pareto based feature ranking techniques: A case study. *Journal of King Saud University Computer and Information Sciences*.
Dash, R. (2018). An adaptive harmony search approach for gene selection and classification of high dimensional medical data. *Journal of King Saud University Computer and Information Sciences*.
Dashtban, M., & Balafar, M. (2017). Gene selection for microarray cancer classification using a new evolutionary method employing artificial intelligence concepts. *Genomics, 109*(2), 91–107.
Dashtbana, M., Mohammadali, B., & Prashanth, S. (2018). Gene selection for tumor classification using a novel bio-inspired multiobjective approach. *Genomics, 110*, 10–17.
Davidson, J. L., & Jalan, J. (2010). Feature selection for steganalysis using the Mahalanobis distance. *Media forensics and security II, 7541*, Article 754104.
de, M. M., Monobe, S., & da Silva, R. C. (2016). Gene expression: An overview of methods and applications for cancer research. *Veterinária e Zootecnia, 23*(4), 532–546.
Dev, J., Dash, S., Dash, S., & Swain, M. (2012). A classification technique for microarray gene expression data using PSO-FLANN. *International Journal on Computer Science and Engineering, 4*(9), 1534–1535.
Ding, C., & Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology, 3*(02), 185–205.
Doreswamy, & UmmeSalma, M. (2016). A binary bat inspired algorithm for the classification of breast cancer data. *International Journal on Soft Computing Intelligence and Applications, 5*(2/3), 1–21.
Dougherty, E. R. (2001). Small sample issues for microarray-based classification. *International Journal of Genomics, 2*(1), 28–34.
Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification*. John Wiley and Sons.
Dy, J. G., & Brodley, C. E. (2004). Feature selection for unsupervised learning. *Journal of Machine Learning Research, 5*(Aug), 845–889.
Ebrahimpour, M. K., Nezamabadi-Pour, H., & Eftekhari, M. (2018). CCFS: A cooperating coevolution technique for large scale feature selection on microarray datasets. *Computational Biology and Chemistry, 73*, 171–178.
Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America, 95*(25), 14863–14868.
El Akadi, A., Amine, A., El Ouardighi, A., & Aboutajdine, D. (2011). A two-stage gene selection scheme utilizing MRMR filter and GA wrapper. *Knowledge and Information System, 26*(3), 487–500.
Elshawi, R., Sakr, S., Talia, D., & Trunfio, P. (2018). Big data systems meet machine learning challenges: Towards big data science as a service. *Big Data Research, 14*, 1–11.
Felix, A. G., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation, 12*(10), 2451–2471.
Fernández, A., del Río, S., Chawla, N. V, & Herrera, F. (2017). An insight into imbalanced big data classification: Outcomes and challenges. *Complex & Intelligent Systems, 3*(2), 105–120.
Ferreira, A. J., & Figueiredo, M. A. T. (2012). An unsupervised approach to feature discretization and selection. *Pattern Recognition, 45*(9), 3048–3060.
Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*(2), 179–188.
Fukunaga, K. (2013). *Introduction to statistical pattern recognition*. Elsevier.
Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics, 20*(3–4), 121–136.
Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2011). A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews, 42*(4), 463–484.
Galar, M., Fernández, A., Barrenechea, E., & Herrera, F. (2013). EUSBoost: Enhancing ensembles for highly imbalanced data-sets by evolutionary undersampling. *Pattern Recognition, 46*(12), 3460–3471.
Gao, L., Ye, M., Lu, X., & Huang, D. (2017). Hybrid method based on information gain and support vector machine for gene selection in cancer classification. *Genomics. Proteomics Bioinformatics, 15*(6), 389–395.
Garber, M. E., et al. (2001). Diversity of gene expression in adenocarcinoma of the lung. *Proceedings of the National Academy of Sciences of the United States of America, 98*(24), 13784–13789.
García, S., Luengo, J., & Herrera, F. (2016). Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. *Knowledge-Based Systems, 98*, 1–29.
Garro, B. A., Rodríguez, K., & Vázquez, R. A. (2016). Classification of DNA microarrays using artificial neural networks and ABC algorithm. *Applied Soft Computing, 38*, 548–560.
Geman, O., Chiuchisan, I., Covasa, M., Doloc, C., Milici, M.-R., & Milici, L.-D. (2016). Deep learning tools for human microbiome big data. In *Proceedings of international workshop soft computing applications* (pp. 265–275).
Ghorai, S., Mukherjee, A., Sengupta, S., & Dutta, P. K. (2010). Cancer classification from gene expression data by NPPC ensemble. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 8*(3), 659–671.
Golub, T. R., et al. (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science (80-.), 286*(5439), 531–537.
Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning*: 1. Cambridge: MIT press.
Gordon, G. J., et al. (2002). Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. *Cancer Research, 62*(17), 4963–4967.
Grate, L. R., Bhattacharyya, C., Jordan, M. I., & Mian, I. S. (2002). Simultaneous relevant feature identification and classification in high-dimensional spaces. In *Proceedings of the international workshop on algorithms in bioinformatics* (pp. 1–9).
Graves, A., & Schmidhuber, J. (2009). Offline handwriting recognition with multidimensional recurrent neural networks. *Advances in neural information processing systems* (pp. 545–552).
Guillen, P., Ebalunode, J., & Learning, A. D. (2016). Cancer classification based on microarray gene expression data using deep learning, pp. 1403–1405, doi:10.1109/CSCI.2016.269.
Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing, 187*, 27–48.
Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*(Mar), 1157–1182.
Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning, 46*(1–3), 389–422.
Hadsell, R., et al. (2009). Learning long-range vision for autonomous off-road driving. *Journal of Future Robotics, 26*(2), 120–144.
Hall, M. A. (1999). Correlation-based feature selection for machine learning.
Hall, M. A., & Smith, L. A. (1998). Practical feature subset selection for machine learning.
Hambali, M. A., & Gbolagade, M. D. (2016). Ovarian cancer classification using hybrid synthetic minority over-sampling technique and neural network. *Journal of Advance Computer Research, 7*(4), 109–124.
Hambali, M., Saheed, Y., Oladele, T., & Gbolagade, M. (2019). ADABOOST ensemble algorithms for breast cancer classification. *Journal of Advance Computer Research, 10*(2), 31–52. OnlineAvailable http://jacr.iausari.ac.ir/article_663924.html .
Hammond, S. M. (2015). An overview of MicroRNAs. *Advanced Drug Delivery Reviews*. 10.1016/j.addr.2015.05.001.
Han, Y., Yang, Y., Yan, Y., Ma, Z., Sebe, N., & Zhou, X. (2014). Semisupervised feature selection via spline regression for video semantic recognition. *IEEE Transactions on Neural Networks and Learning Systems, 26*(2), 252–264.
Hancer, E., Xue, B., Karaboga, D., & Zhang, M. (2015). A binary ABC algorithm based on advanced similarity scheme for feature selection. *Applied Soft Computing, 36*, 334–348.
Hardin, J., & Rocke, D. M. (2004). Outlier detection in the multiple cluster setting using the minimum covariance determinant estimator. *Computational Statistics & Data Analysis, 44*(4), 625–638.
He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering, 21*(9), 1263–1284.
He, W., Cheng, X., Hu, R., Zhu, Y., & Wen, G. (2017). Feature self-representation based hypergraph unsupervised feature selection via low-rank representation. *Neurocomputing, 253*, 127–134.
He, X., Cai, D., & Niyogi, P. (2006). Laplacian score for feature selection. *Advances in neural information processing systems* (pp. 507–514).
Hengpraprohm, S., & Jungjit, S. (2020). Ensemble feature selection for breast cancer classification using microarray data. *Inteligencia Artificial, 23*(65), 100–114.
Hinton, G. E. (2012). A practical guide to training restricted Boltzmann machines. *Neural networks: Tricks of the trade* (pp. 599–619). Springer.
Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation, 18*(7), 1527–1554.
Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science (80-.)., 313*(5786), 504–507.
Hinton, G. E., & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann machines. *Parallel Distributed Processing – Explorations in the Microstructure Cognition, 1*(282–317), 2.
Hinton, G., et al. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine, 29*(6), 82–97.

Hira, Z. M., & Gillies, D. F. (2015). A review of feature selection and feature extraction methods applied on microarray data. *Advances in Bioinformatics, 2015*.
Ho, T. K., & Basu, M. (2002). Complexity measures of supervised classification problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(3), 289–300.
Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.
Hou, C., Nie, F., Li, X., Yi, D., & Wu, Y. (2013). Joint embedding learning and sparse regression: A framework for unsupervised feature selection. *IEEE Transaction on Cybernetics., 44*(6), 793–804.
Hu, R., et al. (2017). Graph self-representation method for unsupervised feature selection. *Neurocomputing, 220*, 130–137.
Huang, H.-L., & Chang, F.-L. (2007). ESVM: Evolutionary support vector machine for automatic feature selection and classification of microarray data. *Biosystems, 90*(2), 516–528.
Huang, S. H. (2015). Supervised feature selection: A tutorial. *Artificial Intelligence Research, 4*(2), 22–37.
Inza, I., Sierra, B., Blanco, R., & Larrañaga, P. (2002). Gene selection by sequential search wrapper approaches in microarray cancer class prediction. *Journal of Intelligent & Fuzzy Systems, 12*(1), 25–33.
Javed, K., Babri, H. A., & Saeed, M. (2010). Feature selection based on class-dependent densities for high-dimensional binary data. *IEEE Transaction on Knowledge and Data Engineering, 24*(3), 465–477.
Kadota, K., Tominaga, D., Akiyama, Y., & Takahashi, K. (2003). Detecting outlying samples in microarray data: A critical assessment of the effect of outliers on sample classification. *Chem-Bio Informatics Journal, 3*(1), 30–45.
Kar, S., Das Sharma, K., & Maitra, M. (2015). Gene selection from microarray gene expression data for classification of cancer subgroups employing PSO and adaptive K-nearest neighborhood technique. *Expert Systems with Applications, 42*(1), 612–627.
Kashyap, H., Ahmed, H. A., Hoque, N., Roy, S., & Bhattacharyya, D. K. (2014). Big data analytics in bioinformatics: A machine learning perspective. *Journal of LATEX Class File, 13*(9), 1–20.
Khan, M. W., & Alam, M. (2012). A survey of application: Genomics and genetic programming, a new frontier. *Genomics, 100*(2), 65–71.
Khodarev, N. N., et al. (2003). Receiver operating characteristic analysis: A general tool for DNA array data filtration and performance estimation. *Genomics, 81*(2), 202–209.
Kim, J., et al. (2005). Identification of potential biomarkers of genotoxicity and carcinogenicity in L5178Y mouse lymphoma cells by cDNA microarray analysis. *Environmental and Molecular Mutagenesis, 45*(1), 80–89.
Kira, K., & Rendell, L. A. (1992). The feature selection problem: Traditional methods and a new algorithm. *AAAI, 2*, 129–134.
Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In *Machine learning proceedings 1992* (pp. 249–256). Elsevier.
Kong, C. S., Yu, J., Minion, F. C., & Rajan, K. (2011). Identification of biologically significant genes from combinatorial microarray data. *ACS Combination Science, 13*(5), 562–571.
Kononenko, I. (1994). Estimating attributes: Analysis and extensions of RELIEF. In *Proceedings of the European conference on machine learning* (pp. 171–182).
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* (pp. 1097–1105).
Krzanowski, W. J. (1987). Selection of variables to preserve multivariate data structure, using principal components. *Journal of the Royal Statistical Society, Series C (Applied Statistics), 36*(1), 22–33.
Lan, L., & Vucetic, S. (2011). Improving accuracy of microarray classification by a simple multi-task feature selection filter. *International Journal of Data Mining and Bioinformatics, 5*(2), 189–208.
Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997). Face recognition: A convolutional neural-network approach. *IEEE Transaction on Neural Networks, 8*(1), 98–113.
Lazar, C. et al. (2012). A survey on filter techniques for feature selection in gene expression microarray analysis, vol. 9, no. 4, pp. 1106–1119.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.
Lee, C.-P., & Leu, Y. (2011). A novel hybrid feature selection method for microarray data analysis. *Applied Soft Computing, 11*(1), 208–213.
Lena, P. D., Nagata, K., & Baldi, P. F. (2012). Deep spatio-temporal architectures and learning for protein structure prediction. *Advances in neural information processing systems* (pp. 512–520).
Leung, Y., & Hung, Y. (2008). A multiple-filter-multiple-wrapper approach to gene selection and microarray data classification. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 7*(1), 108–117.
Li, F., Piao, M., Piao, Y., Li, M., & Ryu, K. H. (2014). A new direction of cancer classification: Positive effect of low-ranking MicroRNAs. *Osong Public Health and Research Perspectives, 5*(5), 279–285.
Li, H., Jiang, T., & Zhang, K. (2004). Efficient and robust feature extraction by maximum margin criterion. *Advances in neural information processing systems* (pp. 97–104).
Li, J., Chen, Z., Wei, L., Xu, W., & Kou, G. (2007). Feature selection via least squares support feature machine. *International Journal of Information Technology & Decision Making, 6*(04), 671–686.
Li, L., et al. (2005). A robust hybrid between genetic algorithm and support vector machine for extracting an optimal feature gene subset. *Genomics, 85*(1), 16–23.
Liang, S., Ma, A., Yang, S., Wang, Y., & Ma, Q. (2018). A review of matched-pairs feature selection methods for gene expression data analysis. *Computational and Structural Biotechnology Journal, 16*, 88–97. 10.1016/j.csbj.2018.02.005.
Liu, H., Motoda, H., Setiono, R., & Zhao, Z. (2010). Feature selection: An ever evolving frontier in data mining. *Feature selection in data mining* (pp. 4–13).
Liu, R., Rallo, R., & Cohen, Y. (2011). Unsupervised feature selection using incremental least squares. *International Journal of Information Technology & Decision Making, 10*(06), 967–987.
Liu, Y., Nie, F., Wu, J., & Chen, L. (2013). Efficient semi-supervised feature selection with noise insensitive trace ratio criterion. *Neurocomputing, 105*, 12–18.
López, V., Fernández, A., García, S., Palade, V., & Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Science (Ny)., 250*, 113–141.
Lorena, A. C., Costa, I. G., Spolaôr, N., & De Souto, M. C. P. (2012). Analysis of complexity indices for classification problems: Cancer gene expression data. *Neurocomputing, 75*(1), 33–42.
Lovato, P., Bicego, M., Cristani, M., Jojic, N., & Perina, A. (2012). Feature selection using counting grids: Application to microarray data. In *Proceedings of the joint IAPR international workshops on statistical techniques in pattern recognition (SPR) and structural and syntactic pattern recognition (SSPR)* (pp. 629–637).
Luo, T., Hou, C., Nie, F., Tao, H., & Yi, D. (2018). Semi-supervised feature selection via insensitive sparse regression with application to video semantic recognition. *IEEE Transactions on Knowledge and Data Engineering, 30*(10), 1943–1956.
Luo, Y., Tao, D., Xu, C., Li, D., & Xu, C. (2013). Vector-valued multi-view semi-supervsed learning for multi-label image classification. In *Proceedings of the twenty-seventh AAAI conference on artificial intelligence*.
Lusa, L. (2012). Evaluation of smote for high-dimensional class-imbalanced microarray data. In *Proceedings of the 2012 eleventh international conference on machine learning and applications: 2* (pp. 89–94).
Maldonado, S., & López, J. (2018). Dealing with high-dimensional class-imbalanced datasets: Embedded feature selection for SVM classification. *Applied Soft Computing, 67*, 94–105.
Maldonado, S., Weber, R., & Basak, J. (2011). Simultaneous feature selection and classification using kernel-penalized support vector machines. *Inf. Sci. (Ny)., 181*(1), 115–128.
Mao, Z., Cai, W., & Shao, X. (2013). Selecting significant genes by randomization test for cancer classification using gene expression data. *Journal of Biomedical Informatics, 46*(4), 594–601.
Maqlin, P., Thamburaj, R., Mammen, J. J., & Manipadam, M. T. (2015). Automated nuclear pleomorphism scoring in breast cancer histopathology images using deep neural networks. In *Proceedings of international conference on mining intelligence and knowledge exploration* (pp. 269–276).
Martín-Merino, M., & De Las Rivas, J. (2009). Improving k-nn for human cancer classification using the gene expression profiles. In *Proceedings of the international symposium on intelligent data analysis* (pp. 107–118).
Masci, J., Meier, U., Cireşan, D., & Schmidhuber, J. (2011). Stacked convolutional auto-encoders for hierarchical feature extraction. In *Proceedings of the international conference on artificial neural networks* (pp. 52–59).
Meyer, P. E., Schretter, C., & Bontempi, G. (2008). Information-theoretic feature selection in microarray data using variable complementarity. *IEEE Journal of Selected Topics in Signal Processing, 2*(3), 261–274.
Michiels, S., Koscielny, S., & Hill, C. (2005). Prediction of cancer outcome with microarrays: A multiple random validation strategy. *Lancet, 365*(9458), 488–492.
Min, S., Lee, B., & Yoon, S. (2017). Deep learning in bioinformatics. *Briefings in Bioinformatics, 18*(5), 851–869. 10.1093/bib/bbw068.
Minsky, M., & Papert, S. (1969). *Perceptron: An introduction to computational geometry*: 19. Cambridge: The MIT Press.
Mitchell, T. M. (1997). Machine learning.
Mohammadi, M., Noghabi, H. S., Hodtani, G. A., & Mashhadi, H. R. (2016). Robust and stable gene selection via maximum–minimum correntropy criterion. *Genomics, 107*(2–3), 83–87.
Moreno-Torres, J. G., Sáez, J. A., & Herrera, F. (2012). Study on the impact of partition-induced dataset shift on $ k $-fold cross-validation. *IEEE Transaction on Neural Networks Learning System, 23*(8), 1304–1312.
Mundra, P. A., & Rajapakse, J. C. (2009). SVM-RFE with MRMR filter for gene selection. *IEEE Transaction on Nanobioscience, 9*(1), 31–37.
Mutch, D. M., Berger, A., Mansourian, R., Rytz, A., & Roberts, M.-A. (2001). Microarray data analysis: A practical approach for selecting differentially expressed genes. *Genome Biology, 2*(12) pp. preprint0009-1.
Narayanan, A., Keedwell, E. C., & Olsson, B. (2002). Artificial intelligence techniques for bioinformatics. *Applied Bioinformatics, 1*, 191–222.
Navarro, F. F. G., & Muñoz, L. A. B. (2009). Gene subset selection in microarray data using entropic filtering for cancer classification. *Expert Systems, 26*(1), 113–124.
Nguyen, D. V, & Rocke, D. M. (2002). Multi-class cancer classification via partial least squares with gene expression profiles. *Bioinformatics, 18*(9), 1216–1226.
Nguyen, D. V, & Rocke, D. M. (2002). Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics, 18*(1), 39–50.
Nie, F., Huang, H., Cai, X., & Ding, C. H. (2010). Efficient and robust feature selection via joint ℓ2, 1-norms minimization. In *Advances in neural information processing systems* (pp. 1813–1821).
Panigrahi, P. P., & Singh, T. R. (2013). Computational studies on Alzheimer's disease associated pathways and regulatory patterns using microarray gene expression and network data: Revealed association with aging and other diseases. *Journal of Theoretical Biology, 334*, 109–121.
Passi, K., Nour, A., & Jain, C. K. (2017). Markov blanket: Efficient strategy for feature subset selection method for high dimensional microarray cancer datasets. In *Proceedings of the 2017 IEEE international conference on bioinformatics and biomedicine (BIBM)* (pp. 1864–1871).
Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(8), 1226–1238.

Pihur, V., Datta, S., & Datta, S. (2008). Finding common genes in multiple cancer types through meta–analysis of microarray experiments: A rank aggregation approach. *Genomics, 92*(6), 400–403.
Qi, Y., Sun, H., Sun, Q., & Pan, L. (2011). Ranking analysis for identifying differentially expressed genes. *Genomics, 97*(5), 326–329.
Quackenbush, J. (2001). Computational analysis of microarray data. *Nature Reviews Genetics, 2*(6), 418–427.
Rafii, F., Hassani, B. D. R., & Kbir, M. A. (2017). New approach for microarray data decision making with respect to multiple sources. In *Proceedings of the second international conference on big data, cloud and applications* (pp. 1–5).
Raileanu, L. E., & Stoffel, K. (2004). Theoretical comparison between the gini index and information gain criteria. *Annals of Mathematics and Artificial Intelligence, 41*(1), 77–93.
Rajeswari, P., & Reena, G. S. (2011). Human liver cancer classification using microarray gene expression data. *International Journal of Computer Applications, 34*(6), 25–37.
Rangasamy, M. (2009). An efficient statistical model based classification algorithm for classifying cancer gene expression data with minimal gene subsets. *International Journal of Cyber Society and Education, 2*(2), 51–66.
Remeseiro López, B., & Bolon Canedo, V. (2019). A review of feature selection methods in medical applications. *Computers in Biology and Medicine, 112*.
Rendón, E., Alejo, R., Castorena, C., Isidro-ortega, F. J., & Granda-gutiérrez, E. E. (2020). Applied sciences data sampling methods to deal with the big data multi-class imbalance problem. *Applied Science, 10*(1276), 1–15. 10.3390/app10041276.
Revathy, N., & Amalraj, R. (2011). Accurate cancer classification using expressions of very few genes. *International Journal on Computer Applications, 14*(4), 19–22.
Reyes-Nava, A., Cruz-Reyes, H., Alejo, R., Rendón-Lara, E., Flores-Fuentes, A. A., & Granda-Gutiérrez, E. E. (2019). Using deep learning to classify class imbalanced gene-expression. In *Proceedings of Iberoamerican congress on pattern recognition (CIARP): 1* (pp. 46–54). 10.1007/978-3-030-13469-3.
Reyes-Nava, A., Sánchez, J. S., Alejo, R., Flores-Fuentes, A. A., & Rendón-Lara, E. (2018). Performance analysis of deep neural networks for classification of gene-expression microarrays. In *Proceedings of the Mexican conference on pattern recognition* (pp. 105–115).
Ruiz, R., Riquelme, J. C., & Aguilar-Ruiz, J. S. (2006). Incremental wrapper-based gene selection from microarray data for cancer classification. *Pattern Recognition, 39*(12), 2383–2392.
Saeid, M. M., Nossair, Z. B., & Saleh, M. A. (2020). A microarray cancer classification technique based on discrete wavelet transform for data reduction and genetic algorithm for feature selection. In *Proceedings of the 2020 forth international conference on trends in electronics and informatics (ICOEI)(48184)* (pp. 857–861).
Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics, 23*(19), 2507–2517.
SáEz, J. A., Luengo, J., & Herrera, F. (2013). Predicting noise filtering efficacy with data complexity measures for nearest neighbor classification. *Pattern Recognition, 46*(1), 355–364.
Sahu, B., & Mishra, D. (2012). A novel feature selection algorithm using particle swarm optimization for cancer microarray data. *Procedia Engineering, 38*, 27–31.
Sánchez-Maroño, N., Alonso-Betanzos, A., García-González, P., & Bolón–Canedo, V. (2010). Multiclass classifiers vs multiple binary classifiers using filters for feature selection. In *Proceedings of the 2010 international joint conference on neural networks (IJCNN)* (pp. 1–8).
Santos, V., Datia, N., & Pato, M. P. M. (2014). Ensemble feature ranking applied to medical data. *Procedia Technology, 17*, 223–230.
Saqib, P., Qamar, U., Khan, R. A., & Aslam, A. (2020). MF-GARF: Hybridizing multiple filters and GA wrapper for feature selection of microarray cancer datasets. In *Proceedings of the 2020 twenty-second international conference on advanced communication technology (ICACT)* (pp. 517–524).
Selvaraj, S., & Natarajan, J. (2011). Microarray data analysis and mining tools. *Bioinformation, 6*(3), 95.
Shah, M., Marchand, M., & Corbeil, J. (2011). Feature selection with conjunctions of decision stumps and learning from microarray data. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(1), 174–186.
Sharbaf, F. V., Mosafer, S., & Moattar, M. H. (2016). A hybrid gene selection approach for microarray data classification using cellular learning automata and ant colony optimization. *Genomics, 107*(6), 231–238.
Sharma, A., Imoto, S., & Miyano, S. (2011). A top-r feature selection algorithm for microarray gene expression data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 9*(3), 754–764.
Shedden, K., et al. (2008). Gene expression–based survival prediction in lung adenocarcinoma: A multi-site, blinded validation study. *Natural Medicine, 14*(8), 822.
Shekar, B. H., & Dagnew, G. (2020). L1-regulated feature selection and classification of microarray cancer data using deep learning. In *Proceedings of third international conference on computer vision and image processing* (pp. 227–242).
Shreem, S. S., Abdullah, S., & Nazri, M. Z. A. (2014). Hybridising harmony search with a Markov blanket for gene selection problems. *Information Science (Ny)., 258*, 108–121.
Shreem, S. S., Abdullah, S., Nazri, M. Z. A., & Alzaqebah, M. (2012). Hybridizing ReliefF, MRMR filters and GA wrapper approaches for gene selection. *Journal of Theoretical and Applied Information Technology, 46*(2), 1034–1039.
Singh, D., et al. (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell, 1*(2), 203–209.
Song, L., Smola, A., Gretton, A., Bedo, J., & Borgwardt, K. (2012). Feature selection via dependence maximization. *Journal of Machine Learning Research, 13*(May), 1393–1434.
Song, Q., Ni, J., & Wang, G. (2011). A fast clustering-based feature subset selection algorithm for high-dimensional data. *IEEE Transactions on Knowledge and Data Engineering, 25*(1), 1–14.
Stoppiglia, H., Dreyfus, G., Dubois, R., & Oussar, Y. (2003). Ranking a random feature for variable and feature selection. *Journal of Machine Learning Research, 3*(Mar), 1399–1414.
Student, S., & Fujarewicz, K. (2012). Stable feature selection and classification algorithms for multiclass microarray data. *Biology Direct, 7*(1), 33.
Sun, Y., Wong, A. K. C., & Kamel, M. S. (2009). Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence, 23*(04), 687–719.
Svozil, D., Kvasnicka, V., & Pospichal, J. (1997). Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems, 39*(1), 43–62.
Svrakic, N. M., Nesic, O., Dasu, M. R. K., Herndon, D., & Perez-Polo, J. R. (2003). Statistical approach to DNA chip analysis. *Recent Progress in Hormone Research, 58*, 75–94.
Swathi, S., Babu, G. A., Sendhilkumar, R., & Bhukya, S. N. (2012). Performance of ART1 network in the detection of breast cancer. In *Proceedings of international conference on computer design and engineering (ICCDE 2012): 49* (pp. 100–105).
Tamayo, P., et al. (1999). Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proceedings of the National Academy of Sciences of the United States of America, 96*(6), 2907–2912.
Tan, F., Fu, X., Zhang, Y., & Bourgeois, A. G. (2008). A genetic algorithm-based method for feature subset selection. *Soft Computers, 12*(2), 111–120.
Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J., & Church, G. M. (1999). Systematic determination of genetic network architecture. *Nature Genetics, 22*(3), 281–285.
Tong, D. L., & Schierz, A. C. (2011). Hybrid genetic algorithm-neural network: Feature extraction for unpreprocessed microarray data. *Artificial Intelligence in Medicine, 53*(1), 47–56.
Tu, K., Yu, H., Guo, Z., & Li, X. (2004). Learnability-based further prediction of gene functions in gene ontology. *Genomics, 84*(6), 922–928.
Veerabhadrappa, & Rangarajan, L. (2010). Bi-level dimensionality reduction methods using feature selection and feature extraction. *International Journal of Computer Applications, 4*(2), 33–38.
Ventimiglia, G., & Petralia, S. (2013). Recent advances in DNA microarray technology: An overview on production strategies and detection methods. *Bionanoscience, 3*(4), 428–450.
Viloria, A., Bonerge, O., Lezama, P., & Mercado-caruzo, N. (2020). Unbalanced data processing using oversampling: Machine learning unbalanced data processing using oversampling: Machine learning. *Procedia Computer Science, 175*, 108–113. 10.1016/j.procs.2020.07.018.
Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. In *Proceedings of the twenty-fifth international conference on machine learning* (pp. 1096–1103).
Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A., & Bottou, L. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research, 11*(12).
Vukusic, I., Grellscheid, S. N., & Wiehe, T. (2007). Applying genetic programming to the prediction of alternative mRNA splice variants. *Genomics, 89*(4), 471–479.
Wanderley, M. F. B., Gardeux, V., Natowicz, R., & de Pádua Braga, A. (2013). GA-KDE-Bayes: An evolutionary wrapper method based on non-parametric density estimation applied to bioinformatics problems. *ESANN*.
Wang, G., Song, Q., Xu, B., & Zhou, Y. (2013). Selecting feature subset for high dimensional data via the propositional FOIL rules. *Pattern Recognition, 46*(1), 199–214.
Wang, J., Bø, T. H., Jonassen, I., Myklebost, O., & Hovig, E. (2003). Tumor classification and marker gene prediction by feature selection and fuzzy c-means clustering using microarray data. *BMC Bioinformatics, 4*(1), 60.
Wang, L., Chu, F., & Xie, W. (2007). Accurate cancer classification using expressions of very few genes. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 4*(1), 40–53.
Wang, P., et al. (2017). Proportional hybrid mechanism for population based feature selection algorithm. *International Journal of Information Technology & Decision Making, 16*(05), 1309–1338.
Wang, S., Tang, J., & Liu, H. (2015). Embedded unsupervised feature selection. In *Proceedings of the twenty-ninth AAAI conference on artificial intelligence*.
Wang, Y., et al. (2011). Predicting human microRNA precursors based on an optimized feature subset generated by GA–SVM. *Genomics, 98*(2), 73–78.
WHO. (2018). *Cancer facts sheet.* WHO http://www.who.int/mediacentre/factsheets/fs297/en/.
Williams, R. J., & Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural Computation, 1*(2), 270–280.
Xing, E. P., Jordan, M. I., & Karp, R. M. (2001). Feature selection for high-dimensional genomic microarray data. in. *Icml, 1*, 601–608.
Xiong, W., Cai, Z., & Ma, J. (2008). A DSRPCL-SVM approach to informative gene analysis. *Genomics Proteomics Bioinformatics, 6*(2), 83–90.
Xu, Z., King, I., Lyu, M. R.-T., & Jin, R. (2010). Discriminative semi-supervised feature selection via manifold regularization. *IEEE Transactions on Neural networks, 21*(7), 1033–1047.
Yang, F., & Mao, K. Z. (2010). Robust feature selection for microarray data based on multicriterion fusion. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 8*(4), 1080–1092.
Yang, Y., Shen, H. T., Ma, Z., Huang, Z., & Zhou, X. (2011). L2, 1-norm regularized discriminative feature selection for unsupervised. In *Proceedings of the twenty-second international joint conference on artificial intelligence*.
Ye, Y., Wu, Q., Huang, J. Z., Ng, M. K., & Li, X. (2013). Stratified sampling for feature subspace selection in random forests for high dimensional data. *Pattern Recognition, 46*(3), 769–787.
Yu, H., Gu, G., Liu, H., Shen, J., & Zhao, J. (2009). A modified ant colony optimization algorithm for tumor marker gene selection. *Genomics Proteomics Bioinformatics, 7*(4), 200–208.
Yu, L., & Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the twentieth international conference on machine learning (ICML-03)* (pp. 856–863).

Yu, L., & Liu, H. (2004). Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research, 5*(Oct), 1205–1224.
Zaharia, M., et al. (2016). Apache spark: A unified engine for big data processing. *Communication ACM, 59*(11), 56–65.
Zahiri, J., Yaghoubi, O., Mohammad-Noori, M., Ebrahimpour, R., & Masoudi-Nejad, A. (2013). PPIevo: Protein–protein interaction prediction from PSSM based evolutionary information. *Genomics, 102*(4), 237–242.
Zhang, R., Nie, F., & Li, X. (2018). Feature selection under regularized orthogonal least square regression with optimal scaling. *Neurocomputing, 273*, 547–553.
Zhang, R., Nie, F., Li, X., & Wei, X. (2019). Feature selection with multi-view data: A survey. *Information Fusion, 50*, 158–167.
Zhao, Z., & Liu, H. (2007). Searching for Interacting Features. In *Proceedings of the twentieth international joint conference on artificial intelligence* (pp. 1156–1161).
Zhao, Z., & Liu, H. (2007). Spectral feature selection for supervised and unsupervised learning. In *Proceedings of the twenty-forth international conference on Machine learning* (pp. 1151–1157).
Zhao, Z., Wang, L., & Liu, H. (2010). Efficient spectral feature selection with minimum redundancy. In *Proceedings of the twenty-fourth AAAI conference on artificial intelligence*.
Zhao, Z., Wang, L., Liu, H., & Ye, J. (2011). On similarity preserving feature selection. *IEEE Transactions on Knowledge and Data Engineering, 25*(3), 619–632.
Zheng, C.-H., Huang, D.-S., Kong, X.-Z., & Zhao, X.-M. (2008). Gene expression data classification using consensus independent component analysis. *Genomics Proteomics Bioinformatics, 6*(2), 74–82.
Zhou, N., & Wang, L. (2007). A modified T-test feature selection method and its application on the HapMap genotype data. *Genomics Proteomics Bioinformatics, 5*(3–4), 242–249.
Zhou, P., Hu, X., Li, P., & Wu, X. (2017). Online feature selection for high-dimensional class-imbalanced data. *Knowledge-Based Systems, 136*, 187–199.
Zhu, Z., Ong, Y.-S., & Dash, M. (2007). Markov blanket-embedded genetic algorithm for gene selection. *Pattern Recognition, 40*(11), 3236–3248.
Zibakhsh, A., & Abadeh, M. S. (2013). Gene selection for cancer tumor detection using a novel memetic algorithm with a multi-view fitness function. *Engineering Applications of Artificial Intelligence, 26*(4), 1274–1281.