A review of feature selection methods in medical applications

Beatriz Remeseiro[a], Veronica Bolon-Canedo[b]

[a]*Department of Computer Science, Universidad de Oviedo*
Campus de Gijón s/n, 33203 Gijón, Spain
[b]*Centro de Investigación CITIC, Department of Computer Science, Universidade da Coruña*
Campus de Elviña s/n, 15071 A Coruña, Spain

Abstract

Feature selection is a preprocessing technique that identifies the key features of a given problem. It has traditionally been applied in a wide range of problems that include biological data processing, finance, and intrusion detection systems. In particular, feature selection has been successfully used in medical applications, where it can not only reduce dimensionality but also help us understand the causes of a disease. We describe some basic concepts related to medical applications and provide some necessary background information on feature selection. We review the most recent feature selection methods developed for and applied in medical problems, covering prolific research fields such as medical imaging, biomedical signal processing, and DNA microarray data analysis. A case study of two medical applications that includes actual patient data is used to demonstrate the suitability of applying feature selection methods in medical problems and to illustrate how these methods work in real-world scenarios.

Keywords: Feature selection, High dimensionality, Pattern recognition, Medical imaging, biomedical data

---

*Principal corresponding author: Beatriz Remeseiro
Email addresses: `bremeseiro@uniovi.es` (Beatriz Remeseiro), `veronica.bolon@udc.es` (Veronica Bolon-Canedo)

1. Introduction

Supervised learning is among the most active fields in machine learning. It involves training a predictive model with a set of samples that includes the target outputs so that once the model is trained, it can deduce the output for samples that have not been observed yet. Depending on the output type (discrete or continuous), this problem is categorized as classification or regression.

The data used by supervised learning algorithms can typically be interpreted as a matrix composed of samples (rows) and features (columns) that define the data. Machine learning algorithms typically require a sufficiently large number of samples because a small number of samples can lead to overfitting and reduced generalization capacity. However, they do not require a large number of features because of reasons associated with *curse of dimensionality* [1] that concerns itself with the problem of analyzing high-dimensional datasets that do not occur in low-dimensionality settings. To overcome the curse of dimensionality, which is related to the difficulty of optimization performed by exhaustive enumeration in product spaces, the dataset needs to be reduced. This is achieved by finding a matrix that has fewer columns and is similar to the original one. Because this matrix is composed of a smaller number of features, it can be used more efficiently than the original matrix. *Dimensionality reduction* can be defined as the process of finding matrices with fewer columns.

Feature selection is one of the techniques used for dimensionality reduction; in this technique, relevant features are selected, and irrelevant and redundant features are discarded [2]. A reduction in input dimensionality can improve performance either by decreasing the learning speed and model complexity or by increasing generalization capacity and classification accuracy. The selection of suitable features can also reduce the measurement cost and improve the understanding of the problem. In some cases, the impact of feature selection is impressive; for example, in microarray data analysis, it is possible to use just 2 of 7,129 features to improve classification performance (see the Leukemia dataset in [3]).

Because of the advantages mentioned earlier, feature selection is being actively applied in real-world problems, mostly classification and regression problems. Feature selection has been successfully used to solve problems in different fields that include microarray analysis, image classification, facial recognition, and text classification [4]. One of the most productive and emerging fields in feature selection and machine learning applications is medicine, where the goal is not only to reduce the dimensionality of the problems but also to reduce the costs involved; such as the extraction of information from images [5], or understanding the reasons behind disagreements regarding disease diagnosis among image-analysis experts [6].

We provide an extensive review of feature selection methods applied in medicine in the last five years that, in some cases, were used to develop ad hoc for solving specific problems. First, we describe the main medical areas where feature selection has been applied—specifically, medical imaging, biomedical signal processing, and DNA microarray data—and follow up with recent developments in each area. Then, we describe an application of feature selection to two real-world medical problems involving image analysis and demonstrate the improvements resulting from the application of feature selection.

The rest of this paper is structured as follows. Section 2 provides necessary background on medical applications for machine learning experts who may not be familiar with certain concepts. Section 3 describes basic feature selection concepts and the most popular techniques. Section 4 describes recent medical applications that have benefited from feature selection. Section 5 describes a case study of two medical applications. Finally, Section 6 concludes the paper.

2. Medical applications

In the digital era, people are able to generate and store data at an unprecedented rate. This explosion in available data for further analysis is as evident in medicine as it is elsewhere. Numerous artificial intelligence methods have been applied in various medical problems with the aim of automating time-

consuming, and often subjective, manual tasks implemented by practitioners in diverse specialties. In this section, we briefly overview the medical terms used in this paper, early initiatives, medical datasets, and relevant studies.

2.1. Terms

Data used in machine learning applications can be classified into two main categories. The first is structured data: traditional datasets that can be stored in a matrix such that each row corresponds to a sample, and one or more columns include the value to be predicted. The second is unstructured data: for example, waves representing sound, text, or images. Generally, machine learning algorithms can be applied directly to structured data, whereas a preliminary feature extraction step is necessary for unstructured data. Now, we describe three main types of medical data: two different types (two-dimensional and one-dimensional) of unstructured data represented by medical imaging and biomedical signal processing, and one type of structured data, namely DNA microarray data.

2.1.1. Medical imaging

Medical imaging is a very active field in image analysis and pattern recognition. It entails the application of various image analysis methods that include image classification or segmentation to medical images [7] such as X-rays, computed tomography (CT) scans, magnetic resonance images (MRI), retinographies, and ultrasound images. Figure 1 shows some representative examples.
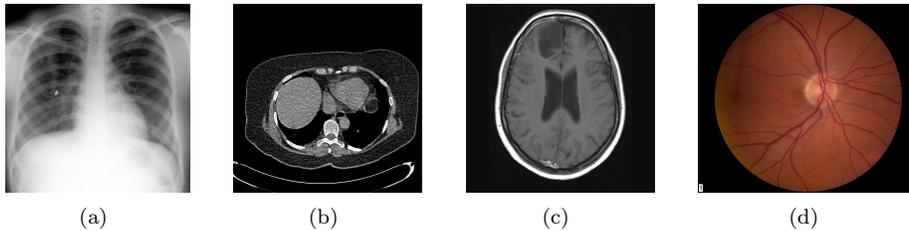

(a)   (b)   (c)   (d)

Figure 1: Examples of different medical images: (a) chest X-ray [8], (b) colon CT section [9], (c) brain MRI section [10], and (d) retinographic image [11].

Many methods have been proposed for performing automatic analysis of digital medical data for screening [12], diagnosis [13], and treatment purposes [14]. These methods often focus on the extraction of features; that is, the computation of image properties such as color, texture, or shape. However, some features may be redundant or irrelevant for a given medical problem, particularly when using general purpose methods or a combination of techniques. This fact, in addition to the large dimensionality of data (high-resolution medical images), make the use of feature selection methods indispensable.

2.1.2. Biomedical signal processing

Biomedical signal processing automates the measurement and analysis of biological signals in medicine that most frequently include brain signals (electroencephalograms, EEG), heart signals (electrocardiograms, ECG), and muscle signals (electromyograms, EMG). Figure 2 shows a representative example of biomedical signals.
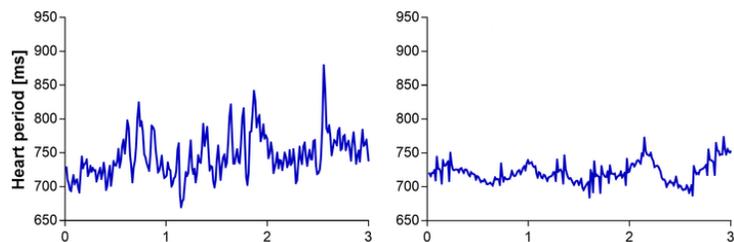

Figure 2: Example traces of heart rate variability: healthy heart (left), and myocardial infarction (right). *Source* Wikimedia Commons.

Biomedical signals have been automatically analyzed for diagnostic [15], monitoring [16], and rehabilitation purposes [17]. Researchers in this field have focused on developing new signal processing techniques that provide practitioners with real-time data for clinical decision-making. Depending on the particular application, these methods include the representation of biosignals by means of Fourier and wavelet basis functions and auto-regressive parameters. This representation can be viewed as a feature vector that can be further analyzed to identify the most representative features and to reduce the dimensionality of

the final dataset.

2.1.3. DNA microarray data

Biomedicine has been a popular field in machine learning in recent years mainly because of the large amount of data that can be extracted from genetic tissues. In particular, the emergence of DNA microarray datasets has led to the emergence of a new and active area of research in bioinformatics and machine learning. For machine learning purposes, microarray data is usually treated as structured data characterized by having very few samples (often under 100) but many features (in the order of thousands). Handling such a large number of features for just a few samples is a challenge for machine learning researchers because of the possibility of "false positives" that can arise by chance (when constructing the prediction model or when selecting the relevant features or genes) [18]. The literature [19] has demonstrated that only a few of these large number of features or genes present in a DNA microarray are relevant for the purpose of resolving a given classification problem. In such a scenario, feature selection is indispensable for two reasons: to eliminate redundant and irrelevant features and to help experts detect underlying relationships between gene expression and a given disease. Figure 3 shows an example of a typical feature selection process applied to microarray data.

2.2. Early initiatives

Artificial intelligence researchers began to tackle medical problems several decades ago, and the result has been the publication of numerous papers based on applications in the three medical areas mentioned above (medical imaging, biomedical signal processing, and DNA microarray data).

Image analysis began to be used for medical purposes in the late 1960s. CT scanning was invented around the same time and continues to be considered one of the most promising areas for medical image analysis [20]. Image analysis has been applied to many different medical conditions, including cancer [21] and retinal diseases [22], and for different purposes, including screening [23]
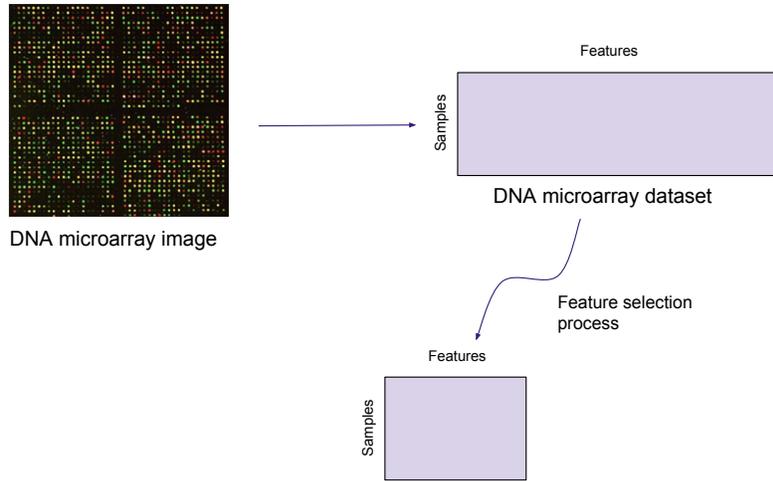
6

Figure 3: Typical feature selection process applied to microarray data. *Source* Wikimedia Commons.

and diagnosis [24]. Biomedical signal processing dates from around the same time. Several methods have been applied to the analysis of EEG [25], ECG [26], and EMG [27] signals. DNA microarrays started to emerge in the 1990s. General approaches extract initial measurements of gene expressions and then apply supervised learning to classify the data. One of several applications is to cancer microarray data classification [28, 29].

2.3. Medical datasets

Public datasets are of utmost importance in testing innovative approaches against state-of-the-art approaches. Several repositories of benchmark datasets, referring to different medical problems, have been made publicly available. Table 1 lists some examples.

The available datasets include unstructured data (such as images and signals) or a set of features extracted from those data, or both. As stated in Section 2.1, machine learning algorithms can be applied directly in the latter case, whereas, in the former case, some prior feature extraction is required. The UCI Repository includes datasets that are particularly useful for general

Table 1: Representative datasets used as benchmarks for various medical problems.

UCI Machine Learning Repository – University of California Irvine [On-line] `http://archive.ics.uci.edu/ml/`
Image Databases at the Image Sciences Institute – Utrecht University [On-line] `https://www.isi.uu.nl/Research/Databases/DRIVE/`
INSPIRE Datasets – University of Iowa Health Care [On-line] `https://medicine.uiowa.edu/eye/inspire-datasets`
VOPTICAL Databases – Universidade da Coruña [On-line] `http://www.varpa.es/research/optics.html#databases`
PhysioBank Databases – National Institute of General Medical Sciences, National Institute of Biomedical Imaging and Bioengineering [On-line] `https://physionet.org/physiobank/database/`
BCI Competitions – Technische Universität Berlin [On-line] `http://www.bbci.de/competition/`
ArrayExpress – European Bioinformatics Institute [On-line] `http://www.ebi.ac.uk/arrayexpress/`
Gene Expression Omnibus – National Institutes of Health [On-line] `http://www.ncbi.nlm.nih.gov/geo/`
Cancer Program Datasets – Broad Institute: [On-line] `http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi`
Lung Image Database Consortium image collection (LIDC-IDRI) – National Cancer Institute, Foundation for the National Institutes of Health, Food and Drug Administration [On-line] `https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI`
Kent Ridge Biomedical Data Set Repository – Institute for Infocomm Research [On-line] `http://research.i2r.a-star.edu.sg/rp/`
Kaggle Datasets and Competitions, – Different institutions share their datasets of a great variety of problems, including medicine and healthcare [On-line] `https://www.kaggle.com/`

machine learning; some of them correspond to various medical problems (for example, cervical cancer and liver disorders).

These datasets have been used in many other research papers, among which we will mention only a few examples. Datasets from the UCI repository, including *hepatitis*, *liver*, and *thyroid*, have been used to evaluate the performance of a new framework for cost-based feature selection [30]. Sasikala et al. [31] analyzed the impact of selecting significant features on accurate diagnosis, making use of medical datasets obtained from the UCI and Kent Ridge repositories. Regarding specific diseases or fields of study, INSPIRE and DRIVE are very

popular datasets for retinal imaging and have been used for problems such as artery and vein classification [32], while VOPTICAL databases have been used for the diagnosis of diseases such as evaporative dry eye or meibomian gland dysfunction [33].

2.4. Related work

Other reviews of medical applications can be found in the recent literature. However, to the best of our knowledge, no study has focused on analyzing the effectiveness of feature selection for medical datasets. Studies that have specialized in medical imaging include reviews of image segmentation applied to breast ultrasound [34] and of biomedical image retrieval systems [35]. Other studies have focused on specific diseases: for example, cancer diagnosis [36] and Alzheimer's disease [37].

Within the field of signal processing, studies include a review of biomedical signal and image processing for biomedical engineering and clinical procedures [38] and works focused on specific types of biomedical signals: for example, EEG [39], ECG [40], and EMG [41].

Several studies have reviewed microarray data classification [42, 43] and some of them have described the application of feature selection to this problem [44, 45, 46].

3. Feature selection

When dealing with a large number of input features, a typical solution is to use specific techniques to reduce the dimensionality of the original problem, which also sometimes improves learning performance. Dimensionality reduction techniques are typically divided into *feature selection* and *feature extraction* methods. The main difference between them is that feature extraction combines the original features and creates a set of *new* features, while feature selection selects a subset of the *original* features, as depicted in Figure 4. Both techniques have advantages and disadvantages [47]. One of the strong points

of feature extraction is that, because the new set of features is usually smaller than that resulting from feature selection, the discriminative power is better. For this reason, feature extraction is more useful for visualization, being typically used in fields such as image analysis, signal processing, and information retrieval. However, a disadvantage is that combinations of features may have no physical meaning, so feature extraction is not a good approach in respect of readability, interpretability, and transparency—characteristics necessary for the development of trustworthy artificial intelligence [48]. Feature selection, which builds a subset of the *original* features, is advantageous when interpretability and knowledge extraction are crucial, as in medicine, although sometimes this comes at the cost of losing some accuracy. Because our work is based on medical problems, we will focus on feature selection.
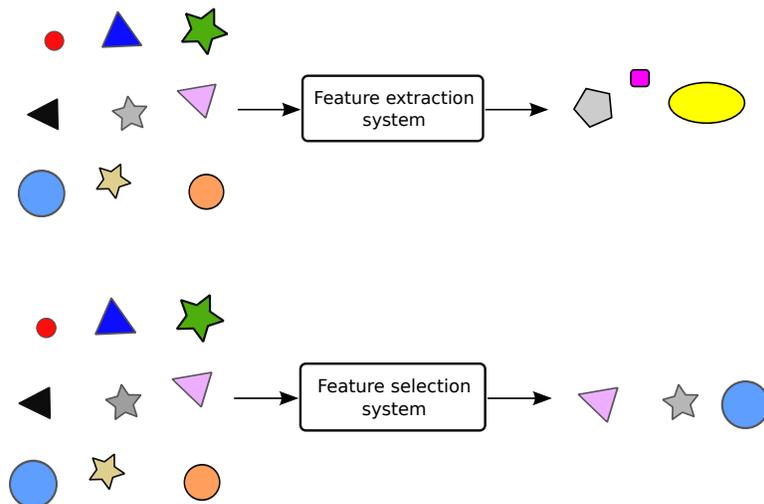

Figure 4: Illustrative example of feature extraction versus feature selection.

Depending on the output, feature selection methods use either *individual evaluation* (also known as feature ranking) or *subset evaluation* of features. In the former, features are assessed individually and assigned a weight that reflects their relevance. In the latter, candidate feature subsets are evaluated using a

10

given measure, so as to select the best feature.

Feature selection methods are also classified into *filters*, *embedded methods*, and *wrappers*, depending on the relationship with the learning method [2]. Filters are independent of any learning method, because the focus is on the general characteristics of the data. Because of their independence from the induction algorithm, they are not computationally costly and have a good generalization capacity. Wrappers and embedded methods both require a learning method to perform feature selection. For wrappers, an induction method evaluates candidate subsets of features. While interaction with the classifier makes wrappers more computationally costly than filters, they tend to perform better. Embedded methods lie between filters and wrappers, because selection is part of the training process for the induction method. The search for the best subset of features is performed during the training of the classifier (for example, while optimizing weights in a neural network) and, for this reason, embedded methods are less computationally costly than wrappers. There is also a tendency to mix algorithms, as in the case of hybrid methods, which usually combine two or more feature selection algorithms of different conceptual origins in a sequential manner. A typical example is to first apply a less computationally costly filter to remove some features and then use a more computationally costly wrapper for fine-tuning.

Most of the new feature selection methods that appear are filters, although we can find representative methods for all three categories. The large number of feature selection methods now available therefore complicates the choice of the best method for a given problem. Below, we describe state-of-the-art methods that have become popular among researchers; some of these methods are used for our case study in Section 5.

- *Correlation-Based Feature Selection* (CFS) [49] is a multivariate filter method that chooses subsets of features that themselves are uncorrelated but show high correlation with the class.

- *Consistency-Based Filter* [50] is a multivariate technique that also selects

11

subsets of features, but selects them according to the degree of consistency with the class, and then uses an inconsistency criterion to determine an acceptable data reduction rate.

- *INTERACT* [51] is an algorithm executed in two steps: first computing the symmetrical uncertainty of all features, and ordering features accordingly, and then individually evaluating the variables by their consistency contribution, so as to select only those whose consistency contribution exceeds a predetermined threshold.

- *Information Gain* (InfoGain) [52] is a simple univariate filter that computes the mutual information for each attribute and class, and then produces an ordered ranking of all of the features.

- *ReliefF* [53] is a popular multivariate filter (an extension of a previous version called Relief [54]) based on nearest neighbors. It works by randomly selecting samples and searching for nearest neighbors from the same class (missing all others). The values of the selected sample are compared with the hits and misses, and then the relevance score for each feature is updated. A useful feature should have values that are similar to examples from the same class and different from examples from the other classes.

- *Recursive Feature Elimination for Support Vector Machines* (SVM-RFE) is an embedded method [55] that, as it trains an SVM classifier in an iterative process, selects features and eliminates the least important ones according to the weights in the SVM solution.

- *Lasso* regularization [56] is an embedded method based on the ℓ_1-norm of the coefficient of a linear classifier $\mathbf{w}$, such that:

$$penalty(\mathbf{w}) = \sum_{i=1}^{m} |\mathbf{w}_i| \qquad (1)$$

and

$$\hat{\mathbf{w}} = \min_{\mathbf{w}} c(\mathbf{w}, X) + \alpha \, penalty(\mathbf{w}) \qquad (2)$$

where $c(\cdot)$ is the classification objective function, $penalty(\mathbf{w})$ is a regularization term, m is the number of features, and α is a regularization parameter to control the trade-off between the $c(\cdot)$ and the penalty [57]. Some of the coefficients of the learned classifier $\mathbf{w}$ may equal zero and, because each coefficient $\mathbf{w}_i$ is associated with a feature f_i, feature selection is achieved by retaining features with non-zero coefficients.

Feature selection has proven to be a successful preprocessing tool for machine learning problems. However, as mentioned, choosing between the growing number of selection methods available is difficult. In a previous work [58], several state-of-the-art feature selection methods were reviewed in terms of their ability to solve common problems such as correlation and redundancy, data non-linearity, noise in the input features, noise in the target class, and (as happens with microarrays) having a number of features much higher than the number of samples. Table 2 summarizes guidance for practitioners [58] depending on the problem to be solved.

Table 2: Recommendations for specific feature selection methods for different scenarios [58]. More stars indicate better behavior.

Method	Correlation/ Redundancy	Non- linearity	Input Noise	Target Noise	No. Features >> No. Samples
CFS	★	★	★	★ ★ ★	★ ★ ★ ★
Consistency	★	★	★	★ ★ ★	★ ★
INTERACT	★	★	★	★ ★ ★	★ ★ ★
InfoGain	★	★	★	★ ★ ★	★ ★ ★
ReliefF	★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★
mRMR	★ ★ ★ ★	★ ★ ★	★ ★ ★ ★ ★	★ ★	★
SVM-RFE*	★ ★ ★ ★	★	★	★ ★ ★ ★	★ ★ ★ ★ ★
SVM-RFE**	★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★	★ ★ ★ ★	−
Wrapper SVM	★	★	★ ★ ★	★ ★ ★ ★	★ ★
Wrapper C4.5	★ ★	★ ★ ★	★ ★ ★	★ ★ ★	★ ★ ★

* Linear kernel
** Non-linear kernel

4. Recent work: 2015 to present

In this section, we review some of the feature selection methods applied to medical imaging, biomedical signal processing, and DNA microarray data in the last five years.

4.1. Medical imaging

Medical imaging has transformed health care, with research efforts yielding benefits such as improved patient care and earlier diagnosis. Image analysis techniques can be applied to various real-world scenarios with promising results. However, because medical datasets usually have a large number of features but few samples of a particular disease, preprocessing using feature selection becomes almost mandatory.

Advances in brain imaging enable practitioners to see inside the living brain, and automatic analysis of brain images is particularly useful. Several studies have been conducted of MRI scans, which are frequently used for brain imaging. One example is classification of MRI images of brain tumors [59], based on a novel hybrid supervised feature selection method that selected the most relevant features from an extracted set that included shape and density features. The experimental results demonstrated this hybrid feature selection method to outperform other supervised algorithms.

MRI also allows observation of structural changes in the brain resulting from neurodegenerative diseases, for which a typical biomarker is the hippocampus. Automatic methods for accurate and reproducible delineation of the hippocampus structure are usually voxel-based approaches that extract several local features for each voxel. A study [60] that compared different feature selection techniques (filters, wrappers, and embedded methods), applied to a set of 300 extracted features per voxel, demonstrated no degradation in performance when the number of features per voxel was significantly reduced.

MRI and positron emission tomography (PET) are commonly used for the diagnosis of Alzheimer's disease. To diagnose this disease from automatically

analyzed MRI and PET images, one study [61] applied multitask learning and a feature selection method based on relational information, that uses the $\ell_{2,1}$-norm regularization as part of the objective function, with the experiments yielding good results in comparison with state-of-the-art methods. Another study used features computed from MRI images in an endeavor to discriminate between different cognitive states related to Alzheimer's disease [62]. Feature selection methods and random forest were used to select the most suitable features from different brain areas, and ensemble classification was applied to obtain the final predictions, with the results demonstrating the capacity of the selected features to identify biomarkers for Alzheimer's.

Another common medical imaging application is the analysis of chest CT scans. For lung disease diagnosis, a feature selection method to select the most relevant signs from CT images [63] was based on using the Fisher criterion and genetic optimization, to evaluate subsets of features to find the optimal subset. The computational efficiency of the method was demonstrated in experiments, which also demonstrated better accuracy than using the whole set of features. Another novel feature selection method [64] was applied to classifying CT images of fibrotic interstitial lung disease. The algorithm, used in conjunction with two different classifiers, selected features extracted from texture analysis, based on their robustness to variations in CT technical settings. The results demonstrated the method to be highly effective in varied CT conditions, because classification performance was not compromised after removing irrelevant features.

A growing interest is prostate image segmentation, for which a semi-automatic method has been described [65] that consists of two novel algorithms for transductive feature selection and regression. Experiments on data from prostate CT images demonstrate the suitability of the method, with results improving on the state-of-the-art. MRI images have also been used for prostate image segmentation [66], based on a supervised segmentation method that uses feature selection to detect relevant features from a dictionary consisting of texture properties obtained from MRI images. Preliminary results reported good performance and a more rapid segmentation process. Yet another prostate-focused study [67]

15

described a novel feature selection method that combines SVM and recursive feature elimination with an absolute cosine filter: SVM-RFE(AC). Applied to prostate histopathological images, the method selected non-redundant features for different tissue components.

Eye images have been widely analyzed to detect and diagnose various diseases. Glaucoma, considered to be one of the main factors causing blindness, can be detected early on using a computational decision support system. One approach to glaucoma detection (specifically, to detect primary angle-closure glaucoma), based on optical coherence tomography (OCT) images, used the rank combination (RC) method [68]. The top-ranked features were identified by various feature selection methods, and a rank combination determined the most discriminating features. Experiments demonstrated that the feature subset resulting from application of this method yielded more competitive results than individually applied filter methods.

Another eye disease is retinopathy of prematurity (ROP), an important cause of childhood blindness. Diagnosis from eye fundus images is subjective, so there is great variability between clinicians. A study to analyze the reasons for this inter-expert variability, using machine learning methods [6], used feature selection methods to select the features that seemed to be relevant for each expert and applied similarity measures to the features relevant to each. The study improved diagnostic accuracy and achieved standardization among clinicians.

The quality of eye fundus images, commonly used in ophthalmology, plays a key role in any automatic method. An objective quality assessment approach has been described [69] that applied different feature selection techniques (CFS, consistency-based filter, and INTERACT) to the selection of the most relevant texture features extracted from eye fundus images; this approach resulted in a reduced number of features without compromising accuracy.

Dry eye syndrome (a known public health problem) is diagnosed by analyzing tear film lipid layer patterns. A computer-aided system to support dry eye diagnosis (CASDES) [33] defines tear film maps using color and texture features, which enables evaluation of the prevalence and distribution of inter-

16

ference patterns observed in the tear film lipid layer. To reduce computational requirements (memory and time), an ad-hoc approach based on the CFS filter was used, with experiments demonstrating that reducing the number of features did not degrade performance and allowed real-time analysis. The influence of eye color on tear film classification was analyzed elsewhere [70] using feature selection; by identifying relevant features according to each type of eye, this outperformed previous approaches.

4.2. Biomedical signal processing

Biomedical signals are analyzed and measured in clinical medicine for prevention, diagnosis, and monitoring purposes. Given the amount of data, and the importance of interpretation, feature selection methods have many interesting applications in this area.

The electrical activity of the brain is reflected in EEG signals, which have been demonstrated to be very useful in studies of the brain, with various signal processing techniques applied to their analysis for a range of diseases, including sleep-related disorders. In a study that focused on the classification of K-complexes (a type of waveform found in EEG signals) [71], a feature selection approach was applied to amplitude and duration measurements (which are the features used to classify waveforms), with CFS achieving maximum accuracy using just 36% of the whole feature set.

The brain-computer interface (BCI) technique allows communications between users and systems with no intervention by muscles or external devices. The idea is to acquire brain signals to analyze electrical changes. A novel technique has been described to select optimal combinations of features for a functional near-infrared spectroscopy BCI [72]. This hybrid approach successfully used genetic algorithms and SVM to determine combinations of just two and three features. Another new hybrid method, based on particle swarm optimization (PSO) and a novel neighborhood rough set classifier (NRSC) has been successfully applied to multiclass classification, resulting in a suitable feature set that achieved high classification accuracy [73].

17

ECG signals to detect heart abnormalities and cardiovascular diseases can be obtained in a noninvasive way. To classify ECG patterns on the basis of selecting optimal feature subsets, a technique based on a genetic algorithm has been developed [74]. The genetic algorithm was modified with a variable-range searching strategy, and an SVM was used to classify the patterns. The dimensionality reduction achieved with this method resulted in better performance than other well-known methods, such as sequential forward selection (SFS) and sequential backward selection (SBS).

EMG signals are used to evaluate and record the electrical activity produced by active muscles and their neural strategies. To assess the relationship between features extracted from EMG signals and target classes in rehabilitation robotics, the coefficient of determination (R^2) value computed from a one-way analysis of variance was used [75] to select the most relevant EMG features, demonstrating the suitability of the R^2 value for feature selection. One study used an SFS algorithm to determine the most relevant EMG features for the prosthesis field [76], with experiments demonstrating the SFS method to be an efficient feature selection algorithm to classify hand and finger movements.

For the purpose of pain-based emotion classification, many types of signals have been used, including ECG and EMG signals [77]. Key sources of information were determined using an interpretable feature selection protocol based on a topologically informed chart. The results led to the conclusion that ECG features are functionally equivalent to the combination of EMG signals and skin conductance levels.

4.3. DNA microarray data

Researchers working with microarray datasets started to apply feature selection methods to reduce dimensionality almost from the outset. As mentioned in Section 3, the most popular feature selection methods are filters, because they are independent of the learning method and are less computationally costly than other approaches. This is particularly important when dealing with microarray data, because the small number of samples may cause data overfitting, which

makes the use of wrappers inadvisable. Most recent feature methods developed for, or applied to, microarray data are filters. For example, modifications of the minimum redundancy maximum relevance (mRMR) algorithm were developed specifically to cope with microarray data. One such modification is Temporal Minimum Redundancy – Maximum Relevance (TMRMR) [78], which is claimed to be able to handle multivariate temporal data without previous data flattening. This is done by using F-statistics, to compute the relevance of each gene, and dynamic time warping to deal with redundancy.

Genetic algorithms are widely used to deal with microarray data. A recently proposed hybrid feature selection method includes the state-of-the-art CFS in the first phase, followed by improved-binary particle swarm optimization (iBPSO) [79]. The method, evaluated for 11 benchmark microarray datasets, obtained promising results in respect of both classification accuracy and dimensionality reduction. Among evolutionary computation techniques, bacterial-inspired algorithms have recently gained popularity for feature selection. In one study, feature selection driven by a bacterial colony optimization algorithm reduced computational complexity, while improving search capacity, even for discrete optimization problems [80]. In another study, two hybrid feature selection algorithms that combined a wrapper and a filter based on binary differential evolution yielded competitive results in comparison with results for nine state-of-the-art methods [81]. Yet another study proposed a hybrid feature selection algorithm combining mutual information maximization with an adaptive genetic algorithm (AGA) [82]. As explained above, genetic algorithms are very widely used in this domain, particularly in combination with other methods, in the form of hybrids.

Embedded methods are popular because they perform feature selection and classification simultaneously. A modification of Lasso for tumor classification in two-class and multiclass datasets has recently been proposed [83]. Nonetheless, the high dimensionality of microarray data frequently means that wrappers and embedded methods cannot be used. This issue inspired an attempt to reduce SVM-RFE time consumption by proposing an improved version of RFE called

19

RFE with variable step size (VSSRFE) [84]. The idea is to use a large step that decreases in size as the number of selected features decreases. The same study also proposed a more efficient implementation of SVM, called large-scale linear SVM (LLSVM), which, when combined with VSSRFE, results in a competitive and rapid feature selection method. Another approach to reducing computation time is to use distributed feature selection [85], based on a framework that distributes the data by features in disjoint subsets and then merges the results of the base feature selectors. Compared with standard centralized approaches, the framework, tested with five state-of-the-art filters on eight microarray datasets, obtained impressive reductions in computation time while maintaining (or even improving) classification accuracy.

5. Case study

In this section, we describe two applications of feature selection in real-world medical imaging problems that yielded promising results.

5.1. Datasets

The first dataset was composed of 34 images to diagnose ROP, a major cause of childhood blindness that mainly affects infants with a low birth weight. For diagnosis, clinical experts usually consider two issues related to the retina: arterial tortuosity and venous dilation. For the purpose of image analysis, information regarding these two features was extracted based on blood vessel information and using either points or segments. To represent each image, we calculated statistical measures including the minimum, maximum, and mean. From the 34 images diagnosed by 22 experts, 66 features were extracted (see further details in [6]).

The second dataset was composed of 105 images to diagnose evaporative dry eye (EDE), a multi-factorial disorder causing eye redness and discomfort that is experienced by a high percentage of people of different ages. People with dry eyes find it difficult to perform common activities, such as reading or working on

a computer. Because of its multi-factorial nature, a large number of diagnostic tests are employed as part of the clinical routine. One is visual assessment of lipid layer patterns and their categorization into four grades [86]. An automatic method to perform this clinical test [87] is based on color and texture as image properties. The best results were obtained when co-occurrence features were extracted as texture information, and the Lab color space was used. Despite the good results, the time needed (38 s) to compute the necessary features was excessive, thereby preventing clinical use of the proposed methodology. The dataset contains 588 features that were extracted from the 105 images using the Lab color space and the co-occurrence features technique [88].

5.2. Results

For ROP, we developed an automatic diagnostic system based on the application of feature selection prior to classification [6]. For classifiers, we selected four popular algorithms belonging to different families (SVM, C4.5, Naive Bayes, and k-NN), which are available in Weka [89]. For feature selection methods, we used six state-of-the-art algorithms: CFS, INTERACT, InfoGain, CFS, ReliefF, and SVM-RFE. Table 3 shows the classification results after selecting the relevant features for determining whether the image represented ROP. We also present the results for no feature selection (No FS) in the first row.

Table 3: Classification error obtained using leave-one-out cross-validation. Best results are in bold.

Method	C4.5	NB	k-NN	SVM
No FS	**29.41**	**11.76**	38.24	20.59
CFS	35.29	14.71	**26.47**	20.59
Cons	32.35	14.71	32.35	20.59
INTERACT	32.35	14.71	**26.47**	17.65
InfoGain	**29.41**	20.59	29.41	17.65
ReliefF	35.29	**11.76**	35.29	**11.76**
SVM-RFE	**29.41**	14.71	**26.47**	17.65

For EDE, and aiming to reduce computation time, we observed that several statistics needed to be computed to extract the textural features. However,

since the corresponding time varied, this caused substantial time variations depending on the choice of a particular feature. We therefore applied a feature selection method that considered the time cost of the features and only selected those features that have maximum relevance and also reduce the corresponding cost [90]. To allow users to fine-tune the trade-off between relevance and cost, we made it possible to change the value of a parameter called λ (higher values corresponded to larger weights for cost) in the optimization function. The proposed approach was used together with three popular feature selection methods, namely, ReliefF, CFS, and mRMR.

Table 4 shows ReliefF results for the points that form the Pareto front: that is, the points that achieved either the minimum classification error or the minimum cost when classified with a SVM. Because ReliefF is a ranker method, a threshold is required; in this case we opted to retain the top 25, 35, and 50 features, as shown in the first column of the table. The second column indicates the values of the parameter λ, while the third and fourth columns represent the classification error and time for each combination of number of features and value of λ. As the table shows, SVM produced the lowest classification error (6.64) with the 50 top-ranked features, with $\lambda = 0$, meaning that time cost was not considered at all. The best time cost result was achieved for the 25 top-ranked features, for any value of λ, but error increased by at least 3.72%. An optimal solution that compromised between error and time cost resulted from the 35 top-ranked features and $\lambda = 1$, for which the error was 7.55% (not much greater than the minimum error) and the time cost was 7.11 ms (significantly better than the 47.98 ms for the combination that achieved the minimum error).

5.3. Discussion

Both cases demonstrate the suitability of feature selection as a preprocessing step. For ROP diagnosis, and for all classifiers, there was always at least one feature selection method that equaled, or improved on, the test error when no feature selection was applied [6]. As for EDE diagnosis, while previous

Table 4: Classification error (%) and time (ms) using 10-fold cross-validation for the Pareto front points. The best results are in bold.

Features	λ	Error	Time
	0.75	10.36	**3.50**
25	2	10.55	**3.50**
	5	14.36	**3.50**
	30	15.18	**3.50**
	1	7.55	7.11
35	2	11.36	7.04
	5	13.18	7.00
	25	16.09	7.00
	0	**6.64**	47.98
	1	9.36	10.62
50	2	9.36	10.62
	25	14.27	10.62
	30	14.36	10.62

approaches that did not apply feature selection achieved promising accuracy, the time required to train the system was unacceptable for real-time clinical routine use. The analysis thus demonstrates that feature selection techniques can lead to a significant reduction in computation time, making the methodology suitable for clinical use to diagnose EDE.

6. Conclusions

In this study, we reviewed recent approaches to medical applications of feature selection; we showed that feature selection is a useful preprocessing tool that not only reduces the number of input features, thereby saving on future data collection but also helps practitioners in understanding the underlying causes of certain diseases.

We considered three main types of medical applications, namely medical imaging, biomedical signal processing, and DNA microarray data, wherein feature selection methods are commonly used to solve problems associated with these types of medical applications. For each category, we provided brief background information, described early initiatives, listed widely used repositories

23

of medical data, and examined recent studies on applications. We also demonstrated the suitability of applying feature selection in two real-world ophthalmology problems. In one case, feature selection outperformed previous classification results; in the second case, feature selection reduced the computation time required to extract the image features that had previously prevented the real-time use of a computer-aided system.

With regard to future scope, it is important not to overlook the emerging Big Data scenario wherein millions of samples and features need to be simultaneously dealt with. The state-of-the-art feature selection methods were not developed under that assumption; thus, most of them are unable to handle vast datasets. Therefore, it is imperative to develop more sophisticated feature selection methods (for example, parallel programming) that can tackle Big Data. Another issue to be considered is that real-time feedback is of paramount importance; this suggests that online feature selection methods are required, which is still a challenge for researchers.

Acknowledgments

References

[1] R. Bellman, Dynamic Programming, Princeton University Press, 1957.

[2] I. Guyon, S. Gunn, M. Nikravesh, L. A. Zadeh, Feature Extraction: Foundations and Applications, Springer Verlag, 2006.

[3] V. Bolón-Canedo, N. Sánchez-Maroño, A. Alonso-Betanzos, On the effectiveness of discretization on gene selection of microarray data, in: International Joint Conference on Neural Networks, 2010, pp. 18–23.

[4] V. Bolón-Canedo, N. Sánchez-Maroño, A. Alonso-Betanzos, Recent advances and emerging challenges of feature selection in the context of big data, Knowledge-Based Systems 86 (2015) 33–45.

[5] B. Remeseiro, V. Bolon-Canedo, D. Peteiro-Barral, A. Alonso-Betanzos, B. Guijarro-Berdinas, A. Mosquera, M. G. Penedo, N. Sanchez-Marono, A methodology for improving tear film lipid layer classification, IEEE Journal of Biomedical and Health Informatics 18 (4) (2014) 1485–1493.

[6] V. Bolón-Canedo, E. Ataer-Cansizoglu, D. Erdogmus, J. Kalpathy-Cramer, O. Fontenla-Romero, A. Alonso-Betanzos, M. Chiang, Dealing with inter-expert variability in retinopathy of prematurity: A machine learning approach, Computer Methods and Programs in Biomedicine 122 (1) (2015) 1–15.

[7] R. Eisenberg, A. Margulis, A Patient's Guide to Medical Imaging, Oxford University Press, 2011.

[8] A. J. Chandrasekhar, Chest X-ray Atlas (2002) [cited June 2019].
URL http://www.meddean.luc.edu/lumen/meded/medicine/pulmonar/cxr/atlas/cxratlas_f.htm

[9] C. D. Johnson, The National CT Colonography Trial (2014) [cited June 2019].
URL https://wiki.cancerimagingarchive.net/display/Public/CT+COLONOGRAPHY

[10] F. N. L. for Cancer Research, The Cancer Genome Atlas-Low Grade Glioma (2014) [cited June 2019].
URL https://wiki.cancerimagingarchive.net/display/Public/TCGA-LGG

[11] M. Niemeijer, X. Xu, A. Dumitrescu, P. Gupta, B. van Ginneken, J. Folk, M. Abramoff, INSPIRE-AVR: Iowa Normative Set for Processing Images of the Retina Artery Vein Ratio (2011) [cited June 2019].
URL https://medicine.uiowa.edu/eye/inspire-datasets

[12] N. Linder, R. Turkki, M. Walliander, A. Mårtensson, V. Diwan, E. Rahtu, M. Pietikäinen, M. Lundin, J. Lundin, A malaria diagnostic tool based on computer vision screening and visualization of Plasmodium falciparum candidate areas in digitized blood smears, PLoS One 9 (8) (2014) e104855.

[13] M. J. Budoff, J. S. Shinbane, Cardiac CT imaging: diagnosis of cardiovascular disease, Springer, 2016.

[14] K. D. Fritscher, M. Peroni, P. Zaffino, M. F. Spadea, R. Schubert, G. Sharp, Automatic segmentation of head and neck CT images for radiotherapy treatment planning using multiple atlases, statistical appearance models, and geodesic active contours, Medical physics 41 (5) (2014) 1–11.

[15] A. K. Tiwari, R. B. Pachori, V. Kanhangad, B. K. Panigrahi, Automated diagnosis of epilepsy using key-point-based local binary pattern of EEG signals, IEEE Journal of Biomedical and Health Informatics 21 (4) (2017) 888–896.

[16] I. Mporas, V. Tsirka, E. I. Zacharaki, M. Koutroumanidis, M. Richardson, V. Megalooikonomou, Seizure detection using EEG and ECG signals for computer-based monitoring, analysis and management of epileptic patients, Expert Systems with Applications 42 (6) (2015) 3227–3233.

[17] K. K. Ang, K. S. G. Chua, K. S. Phua, C. Wang, Z. Y. Chin, C. W. K. Kuah, W. Low, C. Guan, A randomized controlled trial of EEG-based motor imagery brain-computer interface robotic rehabilitation for stroke, Clinical EEG and Neuroscience 46 (4) (2015) 310–320.

[18] G. Piatetsky-Shapiro, P. Tamayo, Microarray data mining: facing the challenges, ACM SIGKDD Explorations Newsletter 5 (2) (2003) 1–5.

[19] T. Golub, D. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. Mesirov, H. Coller, M. Loh, J. Downing, M. Caligiuri, et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (5439) (1999) 531–537.

[20] R. C. Gonzalez, R. E. Woods, Digital image processing (3rd edition) (2008).

[21] I. Magnin, F. Cluzeau, C. Odet, A. Bremond, Mammographic texture analysis: an evaluation of risk for developing breast cancer, Optical Engineering 25 (6) (1986) 256780.

[22] P. Nagin, B. Schwartz, G. Reynolds, Measurement of fluorescein angiograms of the optic disc and retina using computerized image analysis, Ophthalmology 92 (4) (1985) 547–552.

[23] F. Meyer, Automatic screening of cytological specimens, Computer Vision, Graphics, and Image Processing 35 (3) (1986) 356–369.

[24] K. R. Davis, J. M. Taveras, P. F. F. New, J. A. Schnur, G. H. Roberson, Cerebral infarction diagnosis by computerized tomography: analysis and evaluation of findings, American Journal of Roentgenology 124 (4) (1975) 643–660.

[25] P. Y. Ktonas, N. Papp, Instantaneous envelope and phase extraction from real signals: theory, implementation, and an application to EEG analysis, Signal Processing 2 (4) (1980) 373–385.

[26] J. Cox, F. Nolle, H. Fozzard, G. Oliver, AZTEC, a preprocessing program for real-time ECG rhythm analysis, IEEE Transactions on Biomedical Engineering 15 (2) (1968) 128–129.

[27] C. Hershler, M. Milner, An optimality criterion for processing electromyographic (EMG) signals relating to human locomotion, IEEE Transactions on Biomedical Engineering 25 (5) (1978) 413–420.

[28] J. Tödling, R. Spang, Assessment of five microarray experiments on gene expression profiling of breast cancer, in: 7th IEEE Annual International Conference on Computational Biology, 2003, pp. 10–13.

[29] G. Bloom, I. V. Yang, D. Boulware, K. Y. Kwong, D. Coppola, S. Eschrich, J. Quackenbush, T. J. Yeatman, Multi-platform, multi-site, microarray-based human tumor classification, The American Journal of Pathology 164 (1) (2004) 9–16.

[30] V. Bolón-Canedo, I. Porto-Díaz, N. Sánchez-Maroño, A. Alonso-Betanzos, A framework for cost-based feature selection, Pattern Recognition 47 (7) (2014) 2481–2489.

[31] S. Sasikala, S. A. alias Balamurugan, S. Geetha, Multi Filtration Feature Selection (MFFS) to improve discriminatory ability in clinical data set, Applied Computing and Informatics 12 (2) (2016) 117–127.

[32] F. Huang, B. Dashtbozorg, T. Tan, B. M. ter Haar Romeny, Retinal artery/vein classification using genetic-search feature selection, Computer methods and programs in biomedicine 161 (2018) 197–207.

[33] B. Remeseiro, A. Mosquera, M. G. Penedo, CASDES: a computer-aided system to support dry eye diagnosis based on tear film maps, IEEE Journal of Biomedical and Health Informatics 20 (3) (2016) 936–943.

[34] Q. Huang, Y. Luo, Q. Zhang, Breast ultrasound image segmentation: a survey, International journal of computer assisted radiology and surgery 12 (3) (2017) 493–507.

[35] J. Kalpathy-Cramer, A. G. S. de Herrera, D. Demner-Fushman, S. Antani, S. Bedrick, H. Müller, Evaluating performance of biomedical image retrieval systemsan overview of the medical image retrieval task at imageclef 2004–2013, Computerized Medical Imaging and Graphics 39 (2015) 55–61.

[36] V. K. Sudarshan, M. R. K. Mookiah, U. R. Acharya, V. Chandran, F. Molinari, H. Fujita, K. H. Ng, Application of wavelet techniques for cancer

diagnosis using ultrasound images: A review, Computers in biology and medicine 69 (2016) 97–111.

[37] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, C. Davatzikos, A review on neuroimaging-based classification studies and associated feature extraction methods for alzheimer's disease and its prodromal stages, NeuroImage 155 (2017) 530–548.

[38] J. Rajeswari, M. Jagannath, Advances in biomedical signal and image processing – A systematic review, Informatics in Medicine Unlocked 8 (2017) 13–19.

[39] U. R. Acharya, H. Fujita, V. K. Sudarshan, S. Bhat, J. E. Koh, Application of entropies for automated diagnosis of epilepsy using EEG signals: A review, Knowledge-Based Systems 88 (2015) 85–96.

[40] R. Jaros, R. Martinek, R. Kahankova, Non-adaptive methods for fetal ecg signal processing: A review and appraisal, Sensors 18 (11) (2018) 3648.

[41] N. Nazmi, M. Abdul Rahman, S.-I. Yamamoto, S. Ahmad, H. Zamzuri, S. Mazlan, A review of classification techniques of emg signals during isotonic and isometric contractions, Sensors 16 (8) (2016) 1304.

[42] T. Boongoen, N. Iam-On, A comparative review of graph-based ensemble clustering as transformation methods for microarray data classification, Computational Methods With Applications In Bioinformatics Analysis 20 (2017) 53.

[43] J. W. Lee, J. B. Lee, M. Park, S. H. Song, An extensive comparison of recent classification tools applied to microarray data, Computational Statistics & Data Analysis 48 (4) (2005) 869–885.

[44] Y. Saeys, I. Inza, P. Larrañaga, A review of feature selection techniques in bioinformatics, Bioinformatics 23 (19) (2007) 2507–2517.

[45] Z. M. Hira, D. F. Gillies, A review of feature selection and feature extraction methods applied on microarray data, Advances in bioinformatics 2015.

[46] V. Bolón-Canedo, N. Sánchez-Marono, A. Alonso-Betanzos, J. M. Benítez, F. Herrera, A review of microarray datasets and applied feature selection methods, Information Sciences 282 (2014) 111–135.

[47] Z. A. Zhao, H. Liu, Spectral feature selection for data mining, Chapman & Hall/CRC, 2011.

[48] H. L. E. G. on Artificial Intelligence, Ethic Guidelines for Trustworthy AI, European Commission, 2019.

[49] M. A. Hall, Correlation-based feature selection for machine learning, Ph.D. thesis, The University of Waikato (1999).

[50] M. Dash, H. Liu, Consistency-based search in feature selection, Artificial Intelligence 151 (1) (2003) 155–176.

[51] Z. Zhao, H. Liu, Searching for Interacting Features, in: International Joint Conference on Artificial Intelligence, Vol. 7, 2007, pp. 1156–1161.

[52] M. A. Hall, L. A. Smith, Practical feature subset selection for machine learning, Computer Science 98 (1998) 181–191.

[53] I. Kononenko, Estimating attributes: analysis and extensions of RELIEF, in: European Conference on Machine Learning, 1994, pp. 171–182.

[54] K. Kira, L. A. Rendell, The feature selection problem: Traditional methods and a new algorithm, in: Proceedings of the National Conference on Artificial Intelligence, 1992, pp. 129–129.

[55] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Machine Learning 46 (1-3) (2002) 389–422.

[56] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society: Series B (Methodological) 58 (1) (1996) 267–288.

[57] J. Tang, S. Alelyani, H. Liu, Feature selection for classification: A review, Data classification: algorithms and applications (2014) 37.

[58] V. Bolón-Canedo, N. Sánchez-Maroño, A. Alonso-Betanzos, A review of feature selection methods on synthetic data, Knowledge and information systems 34 (3) (2013) 483–519.

[59] G. Jothi, H. Hannah-Inbarani, Hybrid Tolerance Rough Set–Firefly based supervised feature selection for MRI brain tumor image classification, Applied Soft Computing 46 (2016) 639–651.

[60] S. Tangaro, N. Amoroso, M. Brescia, S. Cavuoti, A. Chincarini, R. Errico, P. Inglese, G. Longo, R. Maglietta, A. Tateo, et al., Feature Selection based on Machine Learning in MRIs for Hippocampal Segmentation, Computational and Mathematical Methods in Medicine (Article ID 814104) (2015) 1–10.

[61] X. Zhu, H.-I. Suk, L. Wang, S.-W. Lee, D. Shen, A. D. N. Initiative, et al., A novel relational regularization feature selection method for joint regression and classification in ad diagnosis, Medical image analysis 38 (2017) 205–214.

[62] S. I. Dimitriadis, D. Liparas, M. N. Tsolaki, A. D. N. Initiative, et al., Random forest feature selection, fusion and ensemble strategy: Combining multiple morphological MRI measures to discriminate among healhy elderly, MCI, cMCI and alzheimers disease patients: From the alzheimers disease neuroimaging initiative (ADNI) database, Journal of neuroscience methods 302 (2018) 14–23.

[63] X. Liu, L. Ma, L. Song, Y. Zhao, X. Zhao, C. Zhou, Recognizing common CT imaging signs of lung diseases through a new feature selection

method based on Fisher criterion and genetic optimization, IEEE Journal of Biomedical and Health Informatics 19 (2) (2015) 635–647.

[64] D. Y. Chong, H. J. Kim, P. Lo, S. Young, M. F. McNitt-Gray, F. Abtin, J. G. Goldin, M. S. Brown, Robustness-Driven Feature Selection in Classification of Fibrotic Interstitial Lung Disease Patterns in Computed Tomography Using 3D Texture Features, IEEE Transactions on Medical Imaging 35 (1) (2016) 144–157.

[65] Y. Shi, Y. Gao, S. Liao, D. Zhang, Y. Gao, D. Shen, A learning-based CT prostate segmentation method via joint transductive feature selection and regression, Neurocomputing 173 (2016) 317–331.

[66] M. Guinin, S. Ruan, B. Dubray, L. Massoptier, I. Gardin, Feature selection and patch-based segmentation in MRI for prostate radiotherapy, in: IEEE International Conference on Image Processing, 2016, pp. 2663–2667.

[67] S. Sahran, D. Albashish, A. Abdullah, N. A. Shukor, S. H. M. Pauzi, Absolute cosine-based svm-rfe feature selection method for prostate histopathological grading, Artificial intelligence in medicine 87 (2018) 78–90.

[68] S. I. Niwas, W. Lin, X. Bai, C. K. Kwoh, C. C. Sng, M. C. Aquino, P. T. Chew, Reliable feature selection for automated angle closure glaucoma mechanism detection, Journal of Medical Systems 39 (3) (2015) 21.

[69] B. Remeseiro, A. M. Mendonça, A. Campilho, Objective quality assessment of retinal images based on texture features, in: International Joint Conference on Neural Networks, 2017, pp. 4520–4527.

[70] B. Remeseiro, V. Bolón-Canedo, A. Alonso-Betanzos, M. G. Penedo, Learning features on tear film lipid layer classification, in: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 2015, pp. 195–200.

[71] E. Hernández-Pereira, V. Bolón-Canedo, N. Sánchez-Maroño, D. Álvarez-Estévez, V. Moret-Bonillo, A. Alonso-Betanzos, A comparison of perfor-

mance of K-complex classification methods using feature selection, Information Sciences 328 (2016) 1–14.

[72] F. M. Noori, N. Naseer, N. K. Qureshi, H. Nazeer, R. A. Khan, Optimal feature selection from fnirs signals using genetic algorithms for bci, Neuroscience letters 647 (2017) 61–66.

[73] S. U. Kumar, H. H. Inbarani, Pso-based feature selection and neighborhood rough set-based classification for bci multiclass motor imagery task, Neural Computing and Applications 28 (11) (2017) 3239–3258.

[74] L. Lu, J. Yan, C. W. de Silva, Feature selection for ecg signal processing using improved genetic algorithm and empirical mode decomposition, Measurement 94 (2016) 372–381.

[75] N. M. Kakoty, S. M. Hazarika, J. Q. Gan, Emg feature set selection through linear relationship for grasp recognition, Journal of Medical and Biological Engineering 36 (6) (2016) 883–890.

[76] A. Phinyomark, R. N Khushaba, E. Scheme, Feature extraction and selection for myoelectric control based on wearable emg sensors, Sensors 18 (5) (2018) 1615.

[77] E. D. Campbell, A. Phinyomark, E. J. Scheme, Feature extraction and selection for pain recognition using peripheral physiological signals, Frontiers in Neuroscience 13 (2019) 437.

[78] M. Radovic, M. Ghalwash, N. Filipovic, Z. Obradovic, Minimum redundancy maximum relevance feature selection approach for temporal gene expression data, BMC bioinformatics 18 (1) (2017) 9.

[79] I. Jain, V. K. Jain, R. Jain, Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification, Applied Soft Computing 62 (2018) 203–215.

[80] H. Wang, X. Jing, B. Niu, A discrete bacterial algorithm for feature selection in classification of microarray gene expression cancer data, Knowledge-Based Systems 126 (2017) 8–19.

[81] J. Apolloni, G. Leguizamón, E. Alba, Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments, Applied Soft Computing 38 (2016) 922–932.

[82] H. Lu, J. Chen, K. Yan, Q. Jin, Y. Xue, Z. Gao, A hybrid feature selection algorithm for gene expression data classification, Neurocomputing 256 (2017) 56–62.

[83] C. Kang, Y. Huo, L. Xin, B. Tian, B. Yu, Feature selection and tumor classification for microarray data using relaxed lasso and generalized multiclass support vector machine, Journal of theoretical biology 463 (2019) 77–91.

[84] Z. Li, W. Xie, T. Liu, Efficient feature selection and classification for microarray data, PloS one 13 (8) (2018) e0202167.

[85] V. Bolón-Canedo, N. Sánchez-Maroño, A. Alonso-Betanzos, Distributed feature selection: An application to microarray data classification, Applied soft computing 30 (2015) 136–150.

[86] J.-P. Guillon, Non-invasive Tearscope Plus routine for contact lens fitting, Contact Lens and Anterior Eye 21 (1998) S31–S40.

[87] B. Remeseiro, L. Ramos, M. Penas, E. Martinez, M. G. Penedo, A. Mosquera, Colour texture analysis for classifying the tear film lipid layer: a comparative study, in: International Conference on Digital Image Computing Techniques and Applications, 2011, pp. 268–273.

[88] VOPTICAL_I1, VARPA optical dataset acquired and annotated by optometrists from the Optometry Service of the University of Santiago de Compostela (Spain) (2012) [cited June 2019].
URL http://www.varpa.es/research/optics.html#databases

[89] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, The WEKA data mining software: an update, ACM SIGKDD Explorations Newsletter 11 (1) (2009) 10–18.

[90] V. Bolón-Canedo, B. Remeseiro, N. Sánchez-Maroño, A. Alonso-Betanzos, Real-Time Tear Film Classification Through Cost-Based Feature Selection, in: Transactions on Computational Collective Intelligence XX, Springer, 2015, pp. 78–98.