Computer Modeling in Engineering & Sciences
DOI:10.32604/cmes.2020.011380
Review

Importance of Features Selection, Attributes Selection, Challenges and Future Directions for Medical Imaging Data: A Review

Nazish Naheed[1], Muhammad Shaheen[1], Sajid Ali Khan[1], Mohammed Alawairdhi[2,*] and Muhammad Attique Khan[3,*]

[1]Faculty of Engineering & Information Technology, Foundation University Islamabad, Islamabad, Pakistan
[2]College of Computing and Informatics, Saudi Electronic University, Rayadh, Saudi Arabia
[3]Department of Computer Science, HITEC University, Taxila, Pakistan
[*]Corresponding Authors: Mohammed Alawairdhi. Email: malawairdhi@seu.edu.sa;
Muhammad Attique Khan. Email: attique@ciitwah.edu.pk
Received: 04 July 2020; Accepted: 17 July 2020

Abstract: In the area of pattern recognition and machine learning, features play a key role in prediction. The famous applications of features are medical imaging, image classification, and name a few more. With the exponential growth of information investments in medical data repositories and health service provision, medical institutions are collecting large volumes of data. These data repositories contain details information essential to support medical diagnostic decisions and also improve patient care quality. On the other hand, this growth also made it difficult to comprehend and utilize data for various purposes. The results of imaging data can become biased because of extraneous features present in larger datasets. Feature selection gives a chance to decrease the number of components in such large datasets. Through selection techniques, ousting the unimportant features and selecting a subset of components that produces prevalent characterization precision. The correct decision to find a good attribute produces a precise grouping model, which enhances learning pace and forecast control. This paper presents a review of feature selection techniques and attributes selection measures for medical imaging. This review is meant to describe feature selection techniques in a medical domain with their pros and cons and to signify its application in imaging data and data mining algorithms. The review reveals the shortcomings of the existing feature and attributes selection techniques to multi-sourced data. Moreover, this review provides the importance of feature selection for correct classification of medical infections. In the end, critical analysis and future directions are provided.

Keywords: Medical imaging; imaging data; feature selection; data mining; attribute selection; medical challenges; future directions

1 Introduction

Feature selection or feature reduction is one of the most critical steps in computer vision and image processing. A feature selection algorithm selects the most relevant features from the feature vector and drops the irrelevant attributes. This is also an active research area in machine learning and pattern recognition [1]. In medical imaging, a substantial number of images are being processed for recognition [2]. In some cases, images can be in high resolution, which contains more attributes [3]. Due to this large data, a greater dimensionality feature set is obtained in the feature extraction procedure. All the extracted attributes are not useful, and a higher dimensionality vector affects the performance of the model in terms of time cost and accuracy. Feature selection techniques are utilized to reduce the dimensionality as well as reduce the costs of extraction of information and understanding of the model. The importance of feature selection can be described by an example of using only 2 features from 7129 features to enhance the classification performance [4].

Recently, several feature selection algorithms have been introduced and utilized by the researchers for the recognition of medical images because of the advantages of these methods. These algorithms include genetic algorithm (GA), entropy selection, Particle Swarm Optimization (PSO), and grasshopper and many more. These techniques enhanced both accuracy and time performance. GA is a common feature selection method used nowadays. GA has the ability to give the exact or estimated best solutions. This is based on the theory of evolution and genetics of natural selection [5]. A multi-task hypergraph based feature selection method [6] is introduced for classification of Alzheimer's disease. This method illustrates the high-order dimensionality for each modality. A novel hybrid supervised feature selection method [7] is proposed for the classification of brain tumours from MRI scans. Furthermore, there are different feature selection methods that have developed for the recognition of medical images [8].

New technologies and applications are easily accessible, which results in producing an enormous amount of data [9]. Data can be in the form of reports, design visuals, video and sound arrangements and so forth [10,11]. Analysts in 1991 anticipated that measure of data gets multiplied at regular intervals. The increase in the size of datasets makes it difficult to comprehend and utilize data for various purposes [12]. Today bulk of high dimensional data is stored in online databases, which make its use more difficult and challenging [13]. Data mining (DM) is data science that discovers knowledge from huge databases. In DM, Knowledge is discovered from historical data. Data is analyzed, and Information is learned to extract knowledge [14]. In the context of data management, discovery is made on data to find meaningful patterns. Knowledge discovery in databases is not new and proved its analytical capabilities for data analysis [15]. Application of data mining on different types of data, including graphs, numbers, text, and web, among others are establishing the worth of this domain in different applications. Data mining applies analytical techniques to databases to extract hidden and non-trivial patterns from data. Its applications are broader and not limited to education, telecommunication, superstores, energy sector and bioinformatics.

Knowledge discovery techniques for classification of datasets are divided into two broader categories. (i) Supervised classification. (ii) Unsupervised classification. If the user is provided with class labels in datasets, the applied classification technique will be supervised and the classification done in the absence of class labels is called unsupervised classification. As discussed above that the large datasets have many issues and one of them is the difficulty in accurate classification after applying data mining algorithms, because of extraneous features present in larger datasets. Feature selection and reduction is commonly applied to avoid unnecessary features in the dataset.

Feature selection is a method for ousting the unimportant features and selecting a subset of components which produces prevalent characterization precision. The issue of finding the quality of an attribute is all inclusive in inductive machine learning or data mining algorithms. Effective mining can't be accomplished when inconsequential properties are a piece of forecast process. Such properties are represented in the form of features in a learning problem. Feature Selection is completed by setting a few criteria which assess the value of every competitor attribute [16]. Many words are used for attribute determination like feature selection, feature reduction, dimensionality reduction and attribute selection [17]. All of these have some differences and similarities. Feature selection is the process of selecting some relevant features from a complete feature set which satisfy the qualifying criteria of selection. In feature reduction, new features are created from existing features hence reducing overall number. Dimensionality reduction is all about transforming data into low-dimensional space where features are reduced but their impact on dataset remains. Whereas, attribute selection, is a mathematical representation within which a criterion is specified against which an attribute is evaluated. If it satisfies the criteria, the attribute is selected [18,19].

Feature selection is used in many application domains [20]. It is useful when high dimensional data is supplied with lots of irrelevant and less significant data. Recently, feature selection techniques have been used in many domains such as image recognition [21], image retrieval [22], nobreakbioinformatic data analysis [23], Skin Cancer [24–26], Lungs Cancer [27,28], brain tumor [29,30], Stomach Infections [31,32], Alzheimer's [33,34], dental infections [35] and name a few more [18,36,37]. When correct attributes are selected computations are reduced and time is saved.

Fig. 1a shows the working of the general feature selection method. From original feature set, a subset of features is generated on the basis of some search strategy. Once the subset is generated, all the features of this subset have to undergo through the evolution criteria, if they pass the criteria they are gathered in the final feature set and the process of evaluation goes on. Applications of feature selection techniques in some of the prominent domains are given in Fig. 2.

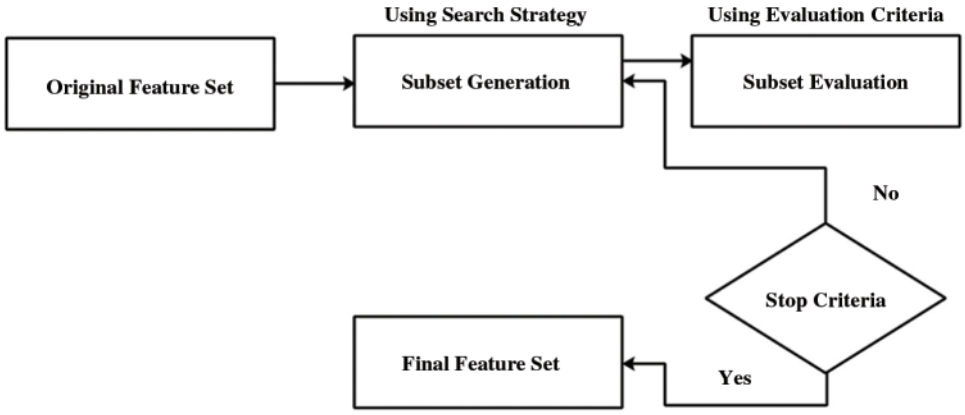

Figure 1: Irrelevant redundant feature detection

Fig. 2 shows that the feature selection techniques are most prominently used in sentiment analysis and image processing. The information of both domains contains a large number of attributes to describe data. Like all other domains, feature selection techniques evolved over time. This paper reviewed feature selection techniques and algorithms especially those which used

data mining techniques for feature selection, with their pros and cons. The paper is organized in a systematic way. Different categories of feature selection methods are given in the next section. Filter algorithms, wrapper feature selection algorithms and attribute selection measures are reviewed in Sections 3–5 respectively. In Section 6, feature selection techniques utilizing data mining are given. The conclusion is given in Section 7.

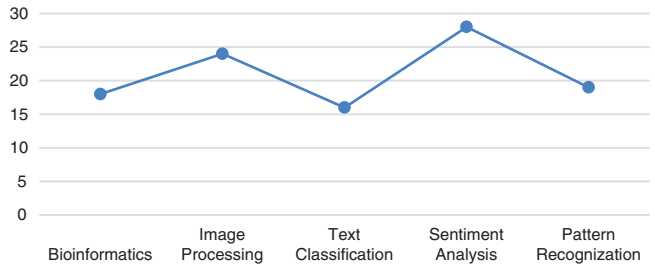

Figure 2: Feature selection techniques usage in different domains

2 Feature Selection Methods

On the premise of the sort, feature selection strategies are gathered in three classifications based on selection criteria [38], Filter, wrapper or embedded approach. In the filter-based approach, no data mining algorithm is used as evaluation criteria. Filter method filter outs irrelevant attributes before induction process starts. In filter-based techniques, quality of an attribute is assessed against some measure. The process of selecting attributes stops when every attribute gets evaluated. Top most attributes are selected which best satisfies the evaluation criteria. Wrapper based techniques require one settled data mining algorithm and utilize its performance as an assessment model. Choice of data mining algorithm ought to satisfy two conditions: (a) It ought to be fit for streamlining characteristics as low as could be allowed; (b) Calculation ought to be exceptionally computationally effective. Wrappers techniques are computationally costly but easy when contrasted with filter strategy. Embedded based techniques exploit both filter and wrapper techniques by changing assessment measures at various stages. The embedded strategy needs to diminish abundant computational time of wrapper approach. The algorithms proposed in the literature for feature selection are placed in one of these three categories. Algorithm wise division is shown in Fig. 3. Fig. 3 shows that in literature filter method algorithms are most commonly used than the wrapper and embedded algorithms, the distribution shows that 47% filter algorithms**,** 29% wrapper algorithm whereas only 24% embedded algorithms are discussed.

Figs. 4–6 shows the advantages and disadvantages of all the three methods of feature selection in a graphical form. For example, it is shown that embedded or filter method have better computational complexity than the wrapper. Filter method doesn't depend on classifier but wrapper and embedded do.

2.1 Feature Selection in Medical Data

Feature selection and extraction techniques are widely used in medical datasets. The role of data analysis techniques in the domains of medical and healthcare is multifaceted. The data used for making different types of diagnosis (differential and non-differential), treatment, prognosis and analysis is of diverse nature. The types of data that are mined in medical applications for concluding one of the aforementioned processes include microarray data [39], data related to heart issues [40], medical imaging data [41] and others. The data used for analysis in medical

domain is usually dense and sparse, for example, the data may be in the form of simple and complex images, arrays, X-rays, medical resonance imaging (MRI), radiotherapic data and immunohistochemistry [42]. The usual or conventional data is also used in medical domain in combination with these complex data structures to conclude a diagnosis or treatment. This sparsity of data increases the number of features presents in the data and can cause curse of dimensionality [43]. Numerous feature selection and extraction methods are used on this data for analyzing relevant features only.

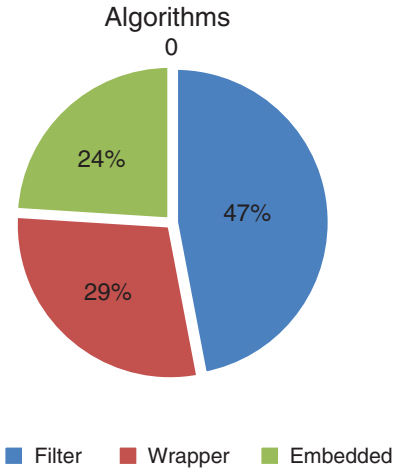

Figure 3: Algorithms of feature selection techniques

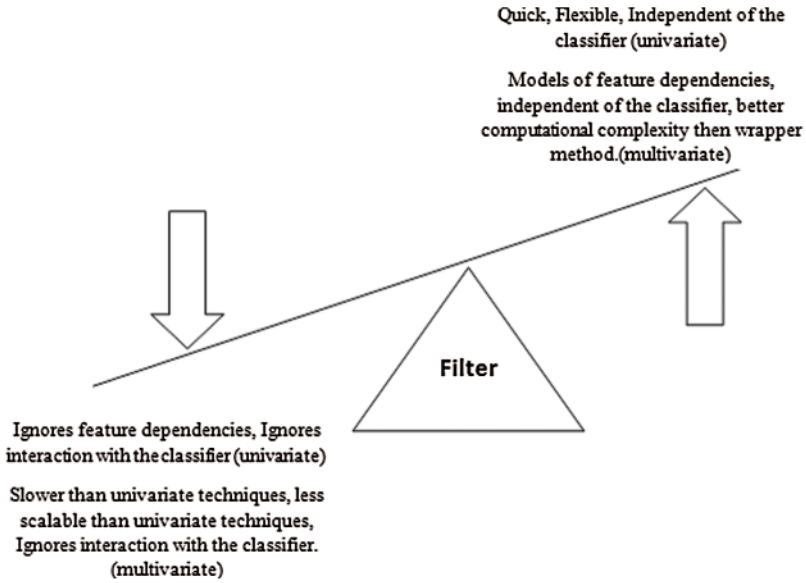

Figure 4: Advantages and disadvantages of the filter method

In the extant literature, these techniques are broadly classified into the techniques for structured and unstructured data. Unstructured data in this domain includes medical imaging data,

genome sequences, medical digital signals and microarray data [44]. Medical images are complex or at least time consuming to analyze not only for the naïve user but also for the specialists related to the domain. These images are available in different formats and their redundant features are eliminated by using different feature selection techniques for X-Rays [45], MRI [46], Ultrasound [47], CT Scans [48], and PET scans [49]. Data mining and machine learning techniques are used with these images either as standalone source of data or in combination with other data, for prognosis, differential diagnosis and treatment.

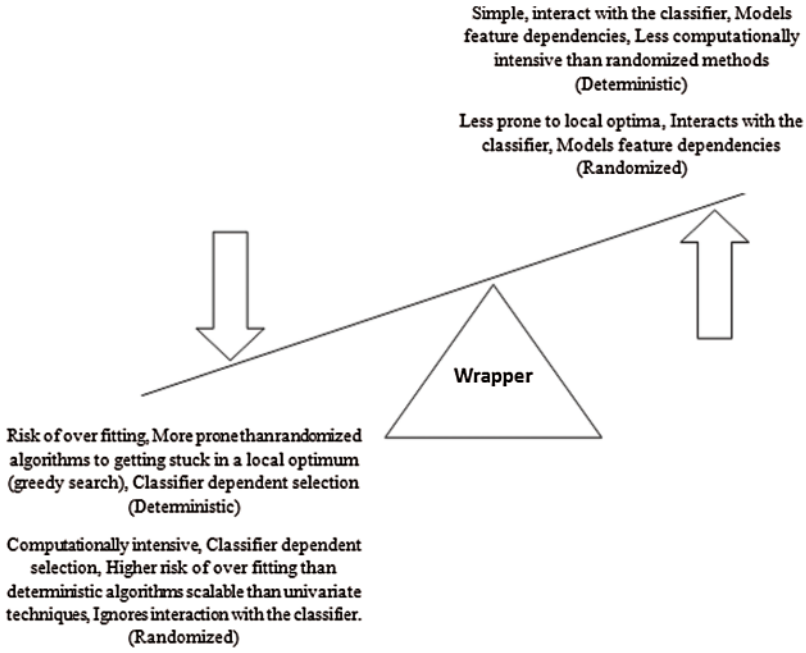

Figure 5: Advantages and disadvantages of the wrapper method

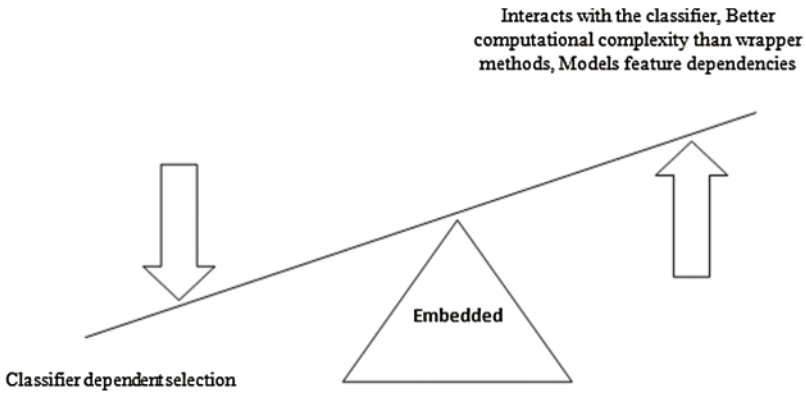

Figure 6: Advantages and disadvantages of the embedded method

The feature selection algorithms citable from literature for the above tasks converts the data to a more compact form capable of extracting implicit patterns by exploiting data mining

and machine learning techniques but are constrained by the interpretability and integrity [50]. Another major limitation on application of feature selection techniques on medical image data is numerousity of the algorithms all of which are closely related to each other [51]. The algorithms which are commonly used for feature selection from medical images are given in Sections 2.2–2.4 and 3.1. As mentioned earlier, feature selection techniques are also used on medical data for genome analysis [52]. Genome analysis is primarily done for detecting genetic patterns, diseases, similarities and dissimilarities. This is done by analyzing genome sequences arranged in different orders. A systematic survey of feature selection techniques used in bioinformatics is presented by [53]. A taxonomy of feature selection techniques with all advantages of filter, wrapper and embedded feature selection techniques used on bioinformatics data is presented in [54]. The filter algorithms in the paper are classified into univariate, multivariate, wrapper methods into deterministic, random and embedded methods are classified into classifier dependent and independent methods. Some ensemble feature selectors are also reviewed in [55]. There are some feature selection algorithms cited in the literature used on medical data for biomedical signal analysis [56]. Biomedical signals are prominently discussed in the literature in the context of electroencephalogram (EEG) signal analysis [57], heart disease prediction [58], brain-computer interface and electromyography (EMG) analysis. Many sequence analysis methods attempt to recognize short, more or less conserved signals to patronized protein sequences.

Another citable attempt made in the literature is related to the application of feature selection techniques on microarray data [59]. Microarray databases are rich repositories of genetic data and are mostly used for cancer differential diagnosis [60]. Genetic algorithms are also very popular for feature selection from microarray data [61]. Manonmani et al. [62] in the first part of their paper presented a review of feature selection techniques used on healthcare datasets. The distribution of these applications is shown in Fig. 7. The feature selection techniques used for the above tasks are presented and reviewed starting from Section 2.2.

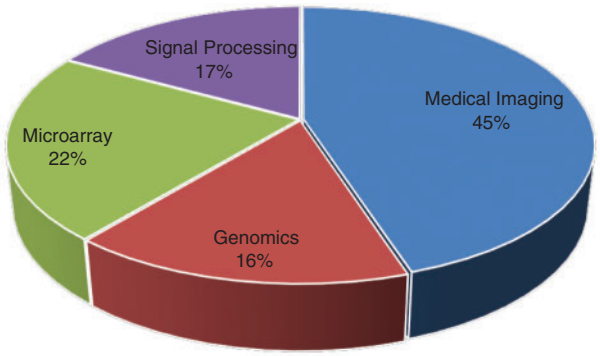

Figure 7: Percent distribution of feature selection techniques in literature

2.2 Filter Algorithms

Filter methods use evaluation criteria against which each attribute is evaluated and filter outs irrelevant attributes before induction process starts. The process of selecting attributes stops when every attribute is evaluated against some measure. Topmost attributes are selected which best satisfies the evaluation criteria. Some of the filter algorithms are described below.

2.2.1 Focus Algorithm

FOCUS algorithm [63] used exponential search to search for consistent attributes in a dataset in forward direction with consistency as an evaluation measure. The algorithm works incrementally by evaluating individual feature set, at first and then each set of two features and so on. It stops when the global consistent solution is found. It uses the concept of conflicts, which is, identifying same samples with different class values. Algorithm searches for a subset with lesser number of conflicts. FOCUS algorithm is computationally expensive because of its exhaustive search. It is not very effective in noise tolerant and applied mostly to discrete values.

2.2.2 FOCUS-2 Algorithm

FOCUS-2 introduces the concept of conflict in positive and negative sets to reduce search space. In A = (001100), positive sets are those who possess 3 or 5 attributes in it, and others are negative [64]. It makes separate queues to place attribute subsets. The time complexity of the FOCUS-2 algorithm is high because, at first, it prunes on the basis of conflicts, and then it performs the search of relevant attributes. However, the tests performed by FOCUS-2 are lesser than FOCUS.

2.2.3 LVF Algorithm

LVF algorithm [65] do a random search in a random direction with any evaluation measure. Consistency is used as an evaluation measure in LVF. Each iteration, a subset is selected and marked as the best subset. In the next iterations, if the selected subset inconsistency rate becomes lesser than the inconsistency rate of the best subset, then it is replaced. LVF has the capability to handle small noisy datasets, but it outperforms in large, noisy datasets. The time complexity of LVF is high because it constantly checks the consistency levels of attributes. The number of attributes selected by this algorithm is usually larger in number for which it is contemplated for achieving a globally consistent solution [66].

2.2.4 LVI Algorithm

LVI (Las Vegas Incremental) [67] is a version of LVF, which uses consistency as an evaluation measure. In LVI, it is not necessary to use the whole sample in order to evaluate the measure. The algorithm starts with a small sample of data, and the inconsistency rate is set to 0. The sample is partitioned into two groups. LVI uses a subset of features for group 1 found by LVF, which is further used to check subsets for Group 2. If Group 2 does not exceed the threshold of inconsistency level, LVI stops. Otherwise, this new sample is handed over to LVI and process re-iterates. The process continues until a solution is found otherwise the whole set is returned as a solution. LVI along with LVF algorithms performs an extensive search of attributes, which identifies relevant attributes but at higher complexity rate. LVI initially performs good but, no improvement in the solution is seen in lateral phases concluding in a flat graph [68].

2.2.5 B&B Algorithm

In B&B (Branch and Bound) algorithm, exponential search in backward direction with any monotonic measure is used for evaluation in medical [69]. It starts with complete set. Features are removed using depth-first strategy. It is an optimal search algorithm in which a threshold is defined. Nodes with lower threshold values are not explored because of monotonicity assumption which states that its sub nodes will not generate optimal solution. Algorithm finds relevant attributes by doing branching of dataset. B&B algorithm stops when the possibility of branching stops. This algorithm is storage-intensive, which makes algorithm computationally expensive. ABB

(Automatic branch & bound) is an extension of B&B in which threshold is automatically set. In Fig. 8 from the pie chart it is evident that consistency is commonly used more than one fourth, almost one third monotonicity and mutual information are used whereas relevance and distance are scarcely used [70]. Details of these measures are discussed in Section 5.

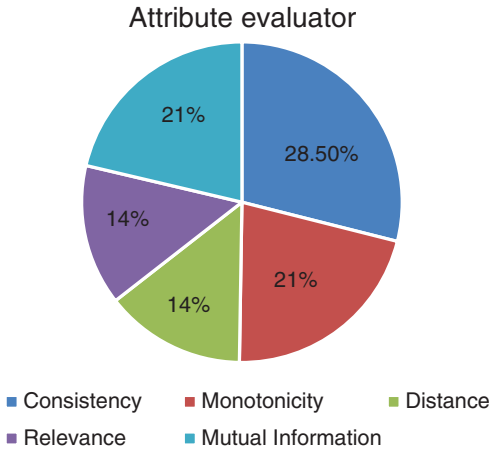

Figure 8: Fractions of different attribute evaluators for feature selection measures

2.2.6 QBB Algorithm

Quick branch and bound (QBB) method uses random and exponential search in the backward direction and uses any monotonic evaluation measure [71]. It is hybrid algorithm made up by combining LVF and ABB. QBB uses LVF to find a good starting point for ABB which latterly explores the search space efficiently. Its execution time is lesser than FOCUS, LVF and ABB algorithms [72].

2.2.7 RELIEF Algorithm

RELIEF algorithm does a random search in the prescribed search space and assigns weights to the attribute which are near to the optimal value [73]. Distance is taken as evaluation measure in this algorithm. From the sample, it chooses an attribute randomly and finds its nearest hit and miss. Nearest hit is considered to be the closest instance of that attribute with same class whereas the nearest miss is the closest instance with a different class. The computation of two nearest neighbors increases time complexity of the algorithm [74].

2.2.8 RELIEF-F Algorithm

RELIEF-F is an improved version of Relief algorithm. The attributes in the RELIEF-F algorithm are selected randomly after which estimated quality values are allocated to the attributes, their nearest hits, and misses. All the hits and misses values are averaged after executing the algorithm for n times for which the time complexity of the algorithm is comparatively lower [75,76].

2.2.9 MRMR Algorithm

Maximum Relevance and Maximum Dependency (MRMR) [77] selects attributes on the basis of two evaluation measures dependency and relevance. If attribute A can be derived from any other attribute B then A is considered to be dependent. Dependency is calculated by

finding correlation among the attributes. Relevance means closely connected. Relevance among attributes is decided on the basis of mutual information shared between attribute and class value. MRMR select those features only whose relevance is maximum and dependency is minimum. The performance of MRMR is claimed to be better than Naïve Bayes, SVM, and LDA algorithms [78].

2.2.10 MRMS Algorithm

It is complex to find dependencies of attributes. Maximum relevance and Maximum significance (MRMS) uses significance of an attribute in place of dependency [79]. It first selects that attribute whose relevance is maximum and then the significance of remaining attributes is compared with the selected attribute [80]. Significance determines mutually exclusive attributes. Zero significance indicates that the removal of the attribute will not affect the accuracy of the results. MRMS algorithm is compared with quick reduct in and gave better results [81].

2.2.11 Red Removing MRLR Algorithm

This is an algorithm based upon maximum mutual information, proposed in Li et al. [82]. Features are selected on the basis of mutual information. The relevance of selected feature is then compared with all candidate attributes. RRMRLR algorithm results in more relevant and less redundant attributes.

2.2.12 Joint Mutual Information Maximization Algorithm

Joint mutual Information Maximization (JMIM) algorithm selects relevant attributes on the basis of mutual information [83,84]. JMIM uses joint mutual information through a greedy approach by taking attributes as a feed-forward network. One of its variants is Normalized joint mutual information maximization (NJMIM) which computes the symmetrical relevance of an attribute. The classification accuracy of JMIM is higher than NJMIM but it seems to be biased towards the attributes with diverse information content [85]. From the above, it can be concluded that consistency attribute evaluator is most prominently used in filter feature selection algorithms. This evaluator keeps only those attributes in the dataset which are the reason for achieving a globally consistent solution. Mutual information and monotonicity are also widely used in filter feature selection algorithms. Algorithms use different search strategies including the exponential, sequential and random search for finding major relevant features. The statistics of use of different search strategies in filter algorithms is given in Fig. 9.

The sequential search is most unmistakably utilized in filter feature selection algorithms. It is preferred due to its optimal results and non-recursive nature. Exponential search is rarely used because it takes more time and has a higher complexity. However random search strategy is hardly used because of its limitations.

2.3 Wrapper Feature Selection Methods

The wrapper model requires one fixed data mining algorithm and uses its performance as an evaluation criterion. Selection of data mining algorithm should fulfill two conditions:

(i) It should be capable of optimizing attributes to as low as possible.
(ii) The algorithm should be highly computationally efficient.

Some wrapper model algorithms are given below.

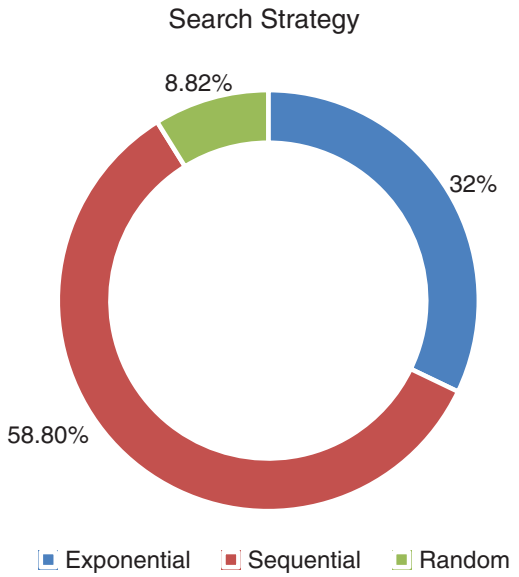

Figure 9: Search strategies used in filter algorithms

2.3.1 SFS Algorithm

Sequential Forward Selection uses heuristic search done in forward direction. It starts with an empty set and adds one feature in the first iteration which is added in the final subset [86]. In next iteration, pairs are made with existing features and new features and best one is selected. The subset is updated with new feature value. In the third iteration, the new feature is evaluated with previously selected features and best feature is selected to form a triplet. It continues until a predefined number of features are selected. SFS finds only those attributes which are needed. SFS gives best results when the size of sample data is small. The main limitation of SFS is its inability to remove features once they are added [87].

2.3.2 SBS Algorithm

A sequential backward selection algorithm (SBS) is exactly the opposite of SFS. All features are initially added to the feature set [88]. An evaluation criterion is used which incrementally removes feature one by one on the basis of some criterion. Removed features are the ones which have no effect on performance. The performance of SBS may be optimal when the large numbers of attributes are present in the dataset [89]. The main limitation of SBS is that it doesn't re-evaluate attributes that are discarded.

2.3.3 SFFS Algorithm

The sequential floating forward search (SFFS) algorithm is based upon backtracking. The addition of this step makes it different from SFS. Like SFS, it initially adds one attribute to the feature set. The attributes are removed on the basis of evaluation measure. SFFS ensures that no relevant attribute is eliminated from feature set. Redundant attributes are not eliminated in this algorithm [90,91].

2.3.4 SFBS Algorithm

The sequential floating backward search (SFBS) algorithm starts with a complete list of features and an objective function. In every iteration, back and forward tracking are done until

the attainment of maximum objective function. Forward and backtracking increase the complexity of the algorithm. The global feature set is conveniently obtained due to its exploration of features which are contributing to achieving objective function [92].

2.3.5 ASFFS Algorithm

Adaptive Sequential Forward Floating Selection (ASFSS) [93] algorithm is a bottom-up approach. It is a more organized search that works in the forward direction. It uses two variables R and O. The former one tells how many features need to be added in inclusion phase and the lateral tells the number of attributes to be removed in exclusion phase. ASFFSS concludes in lesser number of redundant features as compared to SFFS. ASBFS also works exactly in the same way but with the set of all attributes at the starting point. ASFFSS is more complex than ASFBS and it works well on smaller datasets [94].

2.3.6 Plus-L Minus-R Algorithm

Plus-L-Minus-R algorithm [95] is a generalized form of SFS and SBS algorithms. It has two variables, L and R. Variable L and R represents added and removed attributes from a set respectively. If L is greater than R, then it starts with an empty set. It keeps on adding and removing features at the same time. If L becomes lesser than R then it starts from a complete set of attributes. The process continues until defined numbers of attributes are achieved. It tries to compensate limitations of SFS and SBS by adding some backtracking capabilities. The algorithm is more dynamic in behavior as it can be used as SFS and SBS based on the values of L and R. However, no proper procedure for selecting values of L and R is defined for which they are chosen randomly. Fig. 10 shows search strategies used in wrapper algorithms.

Distribution of Search strategies in wrapper algorithms are illustrated in Fig. 10, owing to its peculiar nature in this type of algorithm sequential search is comparatively being used at the maximum on the other hand exponential and random search algorithms are used comparatively at a tangible lesser ratio. Different Data mining techniques are used for determining evaluation criteria in wrapper algorithms. Wrapper methods are computationally expensive and slow as compared to filter algorithms. Fig. 11 shows the division of data mining tasks used in wrapper

Figure 10: Search strategies used in wrapper algorithms

algorithms. Clustering is present as a small minority in Fig. 11 graph and approximately more than three-quarters distribution set out the classification used in wrapper algorithm.

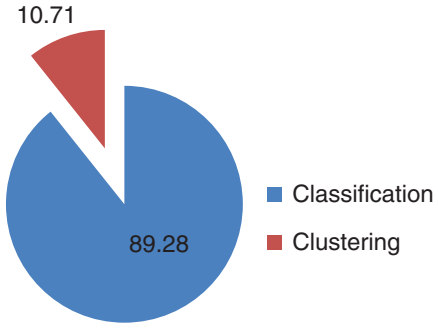

Figure 11: Data mining tasks used in wrapper algorithms

2.4 Embedded Feature Selection Methods

The embedded model tries to take advantage of both filter and wrapper methods by changing evaluation measures at different stages. It searches for features at the time of training. Embedded method wants to reduce the excess computational time of wrapper approach. Some algorithms of the embedded method are discussed below.

2.4.1 BDSFS Algorithm

Boosted Decision Stump Feature Selection (BDSFS) uses information gain criteria to select features. Number of features to be selected is assigned to a variable say k. The algorithm runs for k times. At every step, it ignores all the features that were selected previously. Hence focus is on unselected features with highest information gain [96,97]. It performs the greedy search. Boosting assigns weights. Attribute that correctly predicts class value will have the highest weight. BDSFS is focused upon diversity of an attribute. It picks up attributes which have more information by performing the greedy search.

2.4.2 BDSFS-2 Algorithm

BDSFS specified the number of features to be selected which is considered to be an undesirable property in this algorithm. Stopping criteria are assigned on the basis of learning algorithm used with BDSFS-2. Features with maximum information gain are inserted in the final subset. This algorithm gave comparatively better results. Selection not only depends upon information content of an attribute but also on the learning algorithm used as stopping criteria.

2.4.3 BBHFS Algorithm

Boosting-based hybrid feature selection (BBHFS) technique selects features based on the value of information gain. It uses the learning algorithm to identify features whose weight are high. Stopping criteria is same as of BDSFS-2. It is fast and classification accuracy results are good. Performance of BDSFS-2 and BBHFS is equal and comparable with BDSFS.

2.4.4 SVM-RFE Algorithm

SVM-RFE algorithms change the weight of a feature on the basis of the linear discriminant function of weights [98]. The new objective function is classified by using SVM to perform

recursive feature elimination. SVM-RFE is used for binary class classification. It uses normalization for minimization of SVM problem [99]. Finding optimal objective function is difficult in SVM-RFE.

2.4.5 LFS Algorithm

Lazy feature selection algorithm selects and eliminates features from feature space and also does predictions. It uses K-NN algorithm to predict class label of an attribute mostly used in text categorization problem. Scattering of attributes is taken benefit of and used as a feature selection method. It is classifier dependent. Any change in the classifier changes the working of the algorithm [100,101].

Figs. 12 and 13 shows that all embedded techniques use sequential search strategy at most to find out relevant attributes. A large portion demonstrates the use of classification technique precisely two third and a good proportion of clustering just over a third is used in embedded algorithms.

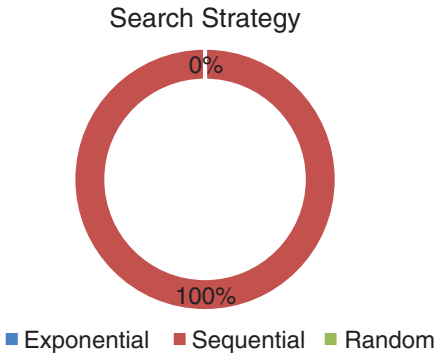

Figure 12: Search strategies used in embedded algorithms

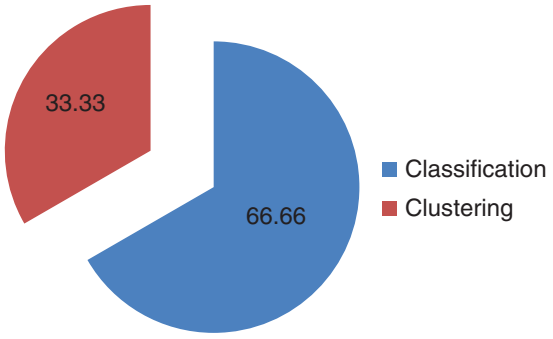

Figure 13: Data mining tasks used in embedded algorithms

In embedded algorithms, classification data mining algorithms are mostly used as evaluation criteria than clustering algorithms. A summary of comparable feature selection algorithms is given in the table below; comparison of algorithms were based on TC (time complexity), SC (space complexity), VT (voice tolerance) and OF (objective function), algorithms with low cost and high cost are almost of equal ratio and, most of them are having VT, binary datasets are only in few algorithms (JMIM, SFBS, ASBFS, ASFFS), I-eval measure is present in almost all

algorithms except MRMS and MRMR, OF is only in PLMR whereas backtracking is present in BDSFS-2 and BBHFS. A summary of FS technique is given in Tab. 1. All the feature selection algorithms are represented in a taxonomic manner in Tab. 2 with its pros and cons and suggested enhancements.

Table 1: Summary of feature selection algorithms

	Algo	Cost*	VT	Search	Binary dataset	1-eval measure	Complex*	T.C*	O.F	S.C	Back tracking
Filter	Focus	High	×	×	×	✓	High	High	×	Low	×
Algorithms	Focus-2	High	✓	×	×	✓	High	Low	×	Low	×
	LVF	Low		×	×	✓	High	Low	×	High	×
	LVI	High	×	✓	×	✓	Low	Low	×	Low	×
	QBB	Low	✓	×	×	✓	High	Low	×	High	×
	B&B	Low	✓	×	×	✓	High	Low	×	High	×
	Relief	Low	✓	×	×	✓	Low	High	×	Low	×
	Relief-F	Low	✓	✓	×	✓	Low	Low	×	Low	×
	RMRLR	–	✓	×	×	✓	Low	Low	×	Low	×
	MRMS	High	✓	×	×	×	Low	High	×	High	×
	MRMR	High	✓	×	×	×	Low	Low	×	High	×
	NJMIM	Low	✓	×	×	✓	High	Low	×	Low	×
Wrapper	JMIM	Low	✓	×	✓	✓	Low	Low	×	Low	×
Algorithms	SBS	High	✓	✓	×	✓	Low	High	×	Low	×
	SFS	High	✓	✓	×	✓	Low	High	×	Low	×
	SFBS	High	✓	✓	✓	✓	Low	High	×	Low	×
	SFFS	High	✓	✓	×	✓	Low	High	×	Low	×
	ASBFS	Low	×	✓	✓	✓	Low	Low	×	High	×
	ASFFS	Low	×	✓	✓	✓	High	Low	×	High	×
Embedded	PLMR	Low	×	✓	×	✓	Low	Low	✓	Low	×
Algorithms	BDSFS	Low	×	✓	×	✓	Low	High	×	Low	×
	BDSFS-2	Low	×	×	×	✓	Low	High	×	Low	✓
	BBHFS	Low	×	✓	×	✓	High	Low	×	Low	✓
	SVM-RFE	Low	×	✓	×	✓	Low	High	✓	Low	×
	LFS	Low	×	✓	✓	✓	Low	Low	×	High	×

*Cost, Complexity, T.C. and S.C are classified into low and high on the basis of certain thresholds.

3 Attribute Selection Measures

In addition to the above stated three major feature selection methods, literature is evident of a few attribute selection measures on the basis of which attributes are either eliminated from the datasets or classified on the basis of their prominence. There are a few evaluation measures which are used to select or reject features [18]. There are almost 29 measures discussed which are used in induction and pruning phase. Most of these attribute selection measures are used in decision tree algorithms where they evaluate attribute importance. Some attribute selection measures are discussed below.

3.1 Information Gain

Information gain is computed from the joint relation between two values. Equality in the values of numerator and denominator indicates independence of both hence generating

Table 2: Taxonomic summary of filter, wrapper and embedded algorithms

Category	Pros	Cons	Discussion
Filter Algorithms	1. The feature selection algorithm in which consistency is considered as an evaluation measure mostly reach globally optimal features from within a feature set. 2. Most of the algorithms are able to handle noisy datasets. 3. The algorithms are based on single evaluation measure for which they are simple to apply. 4. The datasets on which these algorithms are applied are mostly multi class datasets.	1. The time complexity of the algorithms may be high because they mostly work on multiple subsets on original feature set. For each subset, the algorithms completes the whole cycle of execution which makes the algorithms computationally expensive. 2. The algorithms mostly used single evaluation measure for which they can be applied to a specific class of datasets where the features could be reduced with one aspect only. 3. The algorithms are not commonly used in the datasets used for prediction purposes.	Filter algorithms in general are computationally expensive and time-space complex. The algorithms are mostly applied on structured datasets where the features are extracted through single evaluation measure. Datasets like multimedia data are sparse in nature for which the complexity of the algorithms make them less appropriate for its direct application. Similarly the subsets of sparse datasets are also sparse for which the cyclic execution of such algorithms incurs much cost.
Wrapper Algorithms	1. In these algorithms, optimization of the time and space is the part of feature selection algorithm which addresses the problem of time-space complexity. 2. The target of wrapper based algorithms is to achieve globally optimal solution. 3. The algorithms mostly used incremental approach which may reduce the overall execution time for rapid achievement of global optima. 4. Like filter based algorithms, sequential search strategy is used for finding the relevant features which makes the algorithms simple. 5. Sequential search strategy increases the simplicity but decreases sparsity.	1. The algorithms are good for smaller datasets with less number of features. 2. Most of the algorithms do not have a backpropagation mechanism for which the solution once reached may not be undone. 3. A few algorithms selects random features which may add some bias to the final feature set. 4. Despite incorporating optimality mechanism, the algorithms are computationally expensive.	The wrapper algorithms despite using optimization techniques of setting objective function and developing a strategy to address the constraints are computationally expensive. These algorithms are mostly applied on structured datasets and to a smaller class of unstructured datasets like text. The algorithms are found to be used on multimedia data but the complexity is seen to be higher which is because of the incremental approach used in the algorithms. Another reason for its limited application in multimedia systems is the limited use of backpropagation mechanism which is the demand of recursive algorithms applied on multimedia data. Such datasets expect exponential search for selecting appropriate features.

(Continued)

Table 2 (continued).

Category	Pros	Cons	Discussion
Embedded Algorithms	1. These algorithms attempted to add optimal features of both the filter and wrapper techniques. 2. The complexity of the algorithms is comparatively low. 3. These algorithms used optimization techniques for achieving an objective function. 4. The algorithms used sequential search strategy at large.	1. The algorithms developed in this class are smaller in number than the previous categories. 2. Greedy search algorithms are used in a few algorithms which introduced recursiveness in the feature selection process. 3. The iterations or execution cycles are reduced by stopping criteria which when achieved are stopped to proceed.	Embedded algorithms are less in number for which the pros and cons of these algorithms need further deliberations in the literature. These algorithms attempt to combine the pros of wrapper and filter techniques. The algorithms attempt to reach global optimal and consistent solution in a sequential manner for which the time complexity of fewer algorithms is relatively lower. Embedded algorithms are applied on unstructured data for which their complexity in the literature seem higher in the application domain but should not be compared with the complexity of the algorithms applied on unstructured data.

0 result [102,103]. An attribute with greater information gain tells that it contains maximum information which shows its diversity. Information gain is biased towards more diverse attribute. Information gain of an attribute is computed by the following equation.

$$I(C,A) = \log 2\left(\sum_{i=1}^{nc} ri\right) + \log 2\left(\sum_{j=1}^{nA} r.j\right) - \log 2\left(\sum_{i=1}^{nC}\sum_{j=1}^{nA} rij\right) = \log 2\left(\frac{\left(\sum_{i=1}^{nC} ri\right)\left(\sum_{j=1}^{nA} r.j\right)}{\sum_{i=1}^{nC}\sum_{j=1}^{nA} rij}\right) \quad (1)$$

3.2 Kapur's Entropy

Kapur's entropy is a multilevel unsupervised automatic thresholding technique in which the value of entropy is threshold based on the segmented classes of the image. Kapur's entropy is taken as an optimization objective function with a set of constraints applied on it. The sum of entropies of distinct classes $C_1, C_2, \ldots, C_n$ where $P_1, P_2, \ldots, P_n$ are the probabilities of distinct classes in an image is denoted as Kapur's entropy. Maximize objective function as given below is used to get optimal threshold values [104].

$$Fkapur(T1, T2, \ldots, Tn) = MaxH(T1, T2, \ldots, Tn) \quad (2)$$

where, Ti represent threshold.

3.3 Symmetric Information Gain Ratio 1

Information gain may result in biased symmetry which is avoided by combining the attribute and class as given in the equation below; Attribute is represented by A and class by C.

$$Isgr1\ (C,A) = \frac{Igain\,(C,A)}{H(CA)} = \frac{H\,(C) + H\,(A) - H(CA)}{H(CA)} \tag{3}$$

3.4 Symmetric Information Gain Ratio 2

Another way to do normalization of information gain is by dividing by the sum of individual entropies of an attribute with the class value. Both of these normalization techniques try to compensate the biasness of preferring many-valued attributes. Symmetric information gain is computed by following formula.

$$Isgr2 = \frac{Igain(C,A)}{H\,(A) + H(C)} = \frac{H\,(C) + H\,(A) - H(CA)}{H\,(A) + H(C)} \tag{4}$$

3.5 Quadratic Information Gain

In previous information gain computations, Shannon entropy was used which is not the only type of entropy. Another form of entropy that we can generate is quadratic entropy. In this type of entropy, guesses are made to select alternatives. The correctness of the guesses cannot be predicted but the frequency of incorrectness can be determined. Quadratic entropy is used to derive quadratic information gain. The formula is given below.

$$I^2_{gain}\,(C,A) = H^2\,(C) + H^2\,(A) - H^2(CA) \tag{5}$$

3.6 Shannon Measure of Information

Shannon proposed a measure of information of uncertainty and unlikelihood of occurrence of an event known as Shannon entropy [105]. The basic axiom of Shannon entropy for selection of an attribute is that the attribute with low probability value carries more information than the attribute with higher probability value. The value of Shannon probability is calculated by the equation given below [105].

$$S = -\sum_{i=1}^{n} pi \log_2 pi \tag{6}$$

3.7 Correlation Analysis

Correlation analysis is also used for selection of prominent attributes from larger datasets. It is a bivariate analysis that is used to measure the strength of association between two variable. The outcome of correlation analysis ranges from -1 to $+1$ where -1 indicates maximum negative correlation, $+1$ represents maximum positive correlation and 0 indicates independence of the variables. In order to select attributes based on correlation value, a threshold value is selected [106]. This is a popular attribute selection technique and is available in multiple forms like pearson correlation, spearman ranked correlation,kendall correlation coefficient,

canonical correlation etc. Correlation between two variables is calculated by the following equation [107].

$$Corr_s = \frac{Cov(X, Y)}{\partial_X \partial_Y} \tag{7}$$

3.8 Canonical Correlation Analysis

Canonical correlation analysis is a method of inferring information from cross-variance matrices. CCA finds linear combinations between two variables having maximum correlation. The basic idea of CCA is to find an index that may expose a link between the variables between which the correlation is to be sought out. If X and Y are assumed to be two random variables, canonical correlation analysis searches for vectors a and b so that the relation of two indices $a^T X$ and $b^T Y$ can be quantified [108].

$$\rho\,(a, b) = \rho a^T \mathrm{X} \mathrm{b}^T \mathrm{Y} \tag{8}$$

3.9 Metaheuristic Selection Algorithms

A metaheuristic is a process to generate a heuristic that can serve as solution to an optimization problem. Metaheuristic usually do not guarantee a globally optimal solution. There are multiple classifications of metaheuristic like local and global search metaheuristic or single *vs.* population based metaheuristic or parallel and nature inspired. These techniques include particle swarm optimization (PSO), artificial bee colony optimization, ant colony optimization, genetic algorithms and simulated annealing. The use of these algorithms on medical datasets has shown reduced execution time and earlier convergence [109].

3.10 Gini Index

Quadratic information gain (QIG) is very much similar to conventional info gain. Both measures are considered to be biased. Gini index is another measure which is used in different classification algorithms, removes biasness of information gain by penalizing attribute [110,111]. It favors attributes with larger partitions and mostly used in CART tree type. Gini index classifies data by using the squared proportion of classes. Attribute with the lowest value of Gini index is selected. Gini index is computed by using following formula.

$$Gini_{sym}\,(C, A) = \frac{H^2\,(C) - H^2\,(C|A) + H^2\,(A) - H^2(A|C)}{H^2\,(C) + H^2(A)} \tag{9}$$

3.11 Modified Gini Index

Modified Gini index for the first time undermined the diversity of an attribute as to be taken for measuring the worth of an attribute. It squares the probabilities to increase the effect of heuristics. More probable values having higher influence are selected. The formula is given below.

$$Gini_{mod}\,(C, A) = \sum_{j=1}^{nA} \frac{p.j^2}{\sum_{j=1}^{nA} p.j^2} \sum_{i=1}^{nc} pi|j^2 - \sum_{i=1}^{nc} pi^2 \tag{10}$$

3.12 Relief Measure

Relief measure is closely related to Gini index. It is developed for classification tasks and assesses how well an attribute predicts class values [112]. Good prediction is achieved when

every value of attribute corresponds to a unique value of the class attribute. The formula is given below.

$$Relief\,(C,A) = \frac{\sum_{i=1}^{nc}\sum_{j=1}^{nA} pij^2}{\sum_{i=1}^{nc} pi^2} - \frac{\sum_{j=1}^{nA} pj^2 - \sum_{i=1}^{nc}\sum_{j=1}^{nA} Pij^2}{1-\sum_{i=1}^{nC} Pi^2} \tag{11}$$

3.13 Weight of Evidence

Weight of evidence is also used in classification tasks. In order to calculate the weight of evidence, the quotients that can be considered odd are computed by odds $(n) = n/(1-n)$. Like relief measure, it also evaluates that how important this attribute is in the prediction of class label [113]. Greater the value of an attribute, better it would be. The formula is given below.

$$w_{Evid}\,(c_i,A) = \sum_{j=1}^{nA} p.j \left| log2 \frac{P_{i|j}/(1-P_{i|j})}{pi./(1-pi)} \right|, \quad i=1,2 \tag{12}$$

3.14 Relevance Measure

Like relief measure, in the calculation of relevance, an attribute is given priority if its every value corresponds to a unique value of class. Conditional probabilities of an attribute and class are calculated which shows their relevance. The formula is given below.

$$R\,(C,A) = 1 - \frac{1}{nc-1}\sum_{j=1}^{nA}\sum_{i=1,i\neq i\max(j)}^{nc} \frac{pij}{pi} \tag{13}$$

3.15 $\chi 2$ Measure

χ^2 Measure is well known in statistics, here it also does the same. It computes individual squared difference between two distributions. Whereas information gain measure finds the difference between actual joint and independent distribution of two attributes. Equation is given below.

$$X^2\,(C,A) = \sum_{i=1}^{nc}\sum_{j=1}^{nA} \frac{(E_{ij}-N_{ij})^2}{E_{ij}}, \quad where\ E_{ij} = \frac{Ni.Nj}{N} \tag{14}$$

3.16 Specificity Gain

The measure is well known in statistics, here it also does the same. It computes the individual squared difference between two distributions. Whereas information gain measure finds the difference between actual joint and independent distribution of two attributes. The equation is given below.

$$S_{gain}\,(C,A) = nonspec\,(\pi_c) + nonspec\,(\pi_A) - nonspec\,(\pi_{CA}) \tag{15}$$

3.17 Symmetric Specificity Gain

In similarity with information gain, there are many ways to reduce the biasness of multi-valued attributes. Symmetric specificity gain also tries to eliminate biasness by using specificity gain. The formula is given below.

$$S_{sgr}^{(1)}\,(C,A) = \frac{S_{gain}(C,A)}{nonspec(\pi CA)} \quad \text{and} \quad S_{sgr}^{(2)}\,(C,A) = \frac{S_{gain}(C,A)}{nonspec\,(\pi A) + nonspec(\pi c)} \tag{16}$$

3.18 Possibilistic Mutual Information

Information gain was first known as mutual information because it finds the actual joint distribution and independent distributions. Mutual information is the combined probability of class and an attribute. The formula is given below.

$$d_{mi}(C,A) = -\sum_{i=1}^{nc}\sum_{j=1}^{nA}\pi ijlog2\frac{\pi ij}{\min(\pi i,\pi j)} \tag{17}$$

3.19 Feature Selection Evaluation Criteria's

In data mining, classification, and feature selection, different algorithms used different criteria to evaluate the attributes. From literature, it has been observed that most of the algorithms use consistency as an evaluation measure. A detailed breakdown of literature citations in this regard are given in Fig. 14.

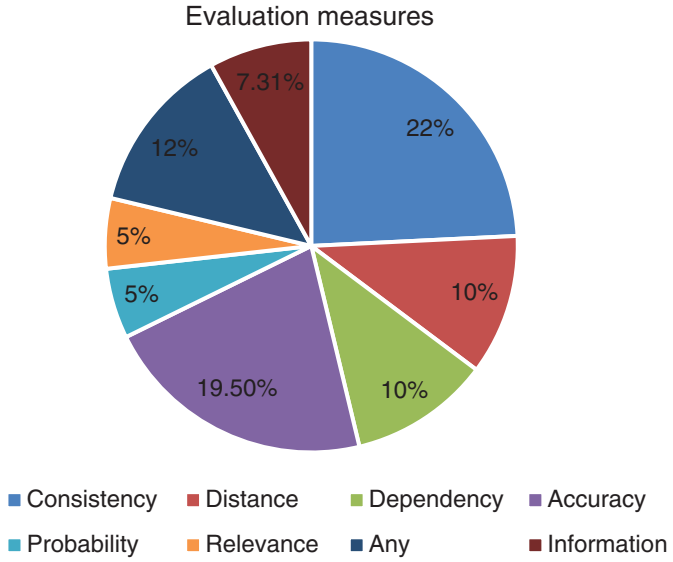

Figure 14: Feature selection evaluation criteria's

Consistency evaluation measure is a recursive process which continuously checks the consistency level achieved by selecting or rejecting an attribute. It is computational overhead to compute distances of attributes in the same class, or across the classes. Relevance evaluation obviously selects relevant attributes at the cost of redundancy in a dataset. Whereas, information evaluation measure picks up attribute that is most informative but is found to be biased towards multi-valued attributes.

Feature selection techniques use classification algorithms which counter checks the classification accuracy results. Many techniques, like, Entropy, GA, PSO, Grasshopper, etc. [114–121] are used for this purpose in medical area. Moreover, in data mining, these techniques are also employing as figured Fig. 15. SVM and GA are top most used techniques of data mining in feature selection algorithms with a difference of one point whereas clustering is two points less than GA same as the ARM is two points less than clustering and RF is six points less than ARM.

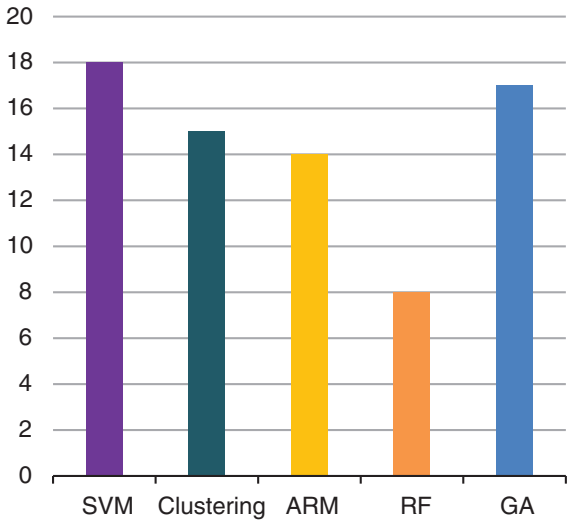

Figure 15: Feature selection algorithms with classification algorithms

4 Future Directions

The feature selection plays a crucial role in the medical imaging domain for classification tasks. The major medical areas in which feature selection techniques are mostly used include skin cancer, brain tumor, lung cancer, stomach diseases, blood diseases, and name a few more. The importance of these areas bases on feature selection techniques is discussed in Section 4. Based on the following points, the feature selection techniques will be considered in future studies.

(a) The selection of robust features can improve the prediction accuracy of attributes based skin lesion classification.
(b) The selection of best features also useful in recognition of multi-type skin lesion classification.
(c) A selection of most robust features reduces the number of predictors, so it is helpful to improve the computation time of skin lesion classification.
(d) Brain datasets like BRATS series are too large in size, and the number of images is in millions. Therefore, the classification process of brain modalities like T1, T1W, T2, and Flair is difficult to classify in relevant categories. Several deep learning techniques are also implemented in literature, but still, they did not achieve the desired accuracy. For this purpose, the feature selection techniques are more useful for brain modalities classification with improved accuracy and low computational time.
(e) Sometimes, the selection of best features is also employing in the segmentation task like a brain tumor and skin lesion. For this purpose, the deep learning models are trained for the detection of lesion area. However, due to the presence of noise and a few other factors, the extracted features do not map the exact lesion area. For this purpose, the feature selection techniques can be used to select only relevant features for an accurate lesion or tumor detection.
(f) The use of feature optimization technique such as genetic algorithm (GA), PSO, Swarm Intelligence, Bee Colony, Whale Optimization, Grasshopper, and few others, are useful in medical imaging and gain a huge success based on few recent articles [122].

(g) The selection of deep learning features also supports better recognition accuracy in the medical domain, especially for a massive amount of imaging data. For deep learning, the meta-heuristic techniques are not useful; therefore, it is essential to implement computational methods like Newton Raphson [123].
(h) The physics-based feature selection techniques like entropy-controlled give vast attention for computer vision researcher whose are working on medical imaging [124,125]. Through entropy-based techniques, the computational time of the implemented CAD system is minimized and improves recognition accuracy.
(i) The performance of each selection method is based on the fitness function. Mostly, researchers utilized FKNN and MSER as a fitness function [2,126,127]. However, the fitness function can be optimized using better techniques like ELM, Neural Network, and Naïve Bayes.

5 Discussion & Conclusion

Due to the exponential growth of information investments in medical data repositories and health service provision, medical institutions are collecting large volumes of data [128]. These data repositories contain details information essential to support medical diagnostic decisions and also improve patient care quality. However, the data is of high-dimensions in which the features numbers are more important than samples. Fast and accurate machine learning systems are desired in the medical domain. Large numbers of samples are required for training by most of the machine learning algorithms because small samples can reduce generalization capacity and also lead to overfitting. Using traditional methods to deal with such data is not a good idea because of the reasons associated with the curse of dimensionality [129]. In order to obtain better accuracy, the selection of important features is required [130–133]. Feature selection is one of the effective methods used to remove the redundant and un-important features before pattern classification [134]. In this study, different approaches to medical applications of feature selection are reviewed. It is demonstrated that the feature selection process is not only useful to reduce the number of features but also can enhance the accuracy rate and thus helps in understanding the underlying cause of diseases. Our discussion spanned feature selection, feature extraction, dimensionality reduction, and attribute selection. Three general approaches of feature selection methods, namely filter, wrapper and embedded methods, are described in detail, and their algorithms are also presented. The attribute evaluation measures are reviewed, and the interesting facts regarding the advantages and disadvantages of feature selection methods and attribute evaluation measures are also discussed. It is concluded that the algorithms of feature selection and reduction are useful for both execution time and accuracy for medical imaging. Moreover, it is also concluded that the selection of relevant feature decreases the overall complexity of a CAD system.

Funding Statement: The author(s) received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Kharrat, A., Mahmoud, N. (2019). Feature selection based on hybrid optimization for magnetic resonance imaging brain tumor classification and segmentation. *Applied Medical Informatics, 41,* 9–23.

2. Majid, A., Khan, M. A., Yasmin, M., Rehman, A., Yousafzai, A. et al. (2020). Classification of stomach infections: a paradigm of convolutional neural network along with classical features fusion and selection. *Microscopy Research and Technique, 83,* 562–576.
3. Wang, S., Tang, C., Sun, J., Zhang, Y. (2019). Cerebral micro-bleeding detection based on densely connected neural network. *Frontiers in Neuroscience, 13,* 422.
4. Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A. (2010). On the effectiveness of discretization on gene selection of microarray data. *2010 International Joint Conference on Neural Networks*, Barcelona, Spain: IEEE, 1–8.
5. Goldberg, D. E. (1989). Genetic algorithms in search. *Optimization, and machine learning.*
6. Shao, W., Peng, Y., Zu, C., Wang, M., Zhang, D. et al. (2020). Hypergraph based multi-task feature selection for multimodal classification of Alzheimer's disease. *Computerized Medical Imaging and Graphics, 80,* 101663.
7. Jothi, G. (2016). Hybrid tolerance rough set-firefly based supervised feature selection for MRI brain tumor image classification. *Applied Soft Computing, 46,* 639–651.
8. Yu, X., Kang, C., Guttery, D. S., Kadry, S., Chen, Y. et al. (2020). ResNet-SCDA-50 for breast abnormality classification. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.
9. Zhang, Y. D., Govindaraj, V. V., Tang, C., Zhu, W., Sun, J. (2019). High performance multiple sclerosis classification by data augmentation and AlexNet transfer learning model. *Journal of Medical Imaging and Health Informatics, 9,* 2012–2021.
10. Solorio-Fernández, S., Carrasco-Ochoa, J. A., Martínez-Trinidad, J. F. (2020). A review of unsupervised feature selection methods. *Artificial Intelligence Review, 53,* 907–948.
11. Wang, S. H., Xie, S., Chen, X., Guttery, D. S., Tang, C. et al. (2019). Alcoholism identification based on an AlexNet transfer learning model. *Frontiers in Psychiatry, 10,* 205.
12. Bommert, A., Sun, X., Bischl, B., Rahnenführer, J., Lang, M. (2020). Benchmark for filter methods for feature selection in high-dimensional classification data. *Computational Statistics & Data Analysis, 143,* 106839.
13. Habib, M., Aljarah, I., Faris, H., Mirjalili, S. (2020). Multi-objective particle swarm optimization: theory, literature review, and application in feature selection for medical diagnosis. *Evolutionary machine learning techniques*. Singapore: Springer, 175–201.
14. Shaheen, M., Shahbaz, M., Guergachi, A. (2013). Context based positive and negative spatio-temporal association rule mining. *Knowledge-Based Systems, 37,* 261–273.
15. Shaheen, M., Khan, M. Z. (2016). A method of data mining for selection of site for wind turbines. *Renewable and Sustainable Energy Reviews, 55,* 1225–1233.
16. Batool, F. E., Attique, M., Sharif, M., Javed, K., Nazir, M. et al. (2020). Offline signature verification system: a novel technique of fusion of GLCM and geometric features using SVM. *Multimedia Tools and Applications,* 1–20.
17. Hussain, N., Khan, M. A., Sharif, M., Khan, S. A., Albesher, A. A. et al. (2020). A deep neural network and classical features based scheme for objects recognition: an application for machine inspection. *Multimedia Tools and Applications*. DOI 101007/s11042-020–08852-3.
18. Alaoui, A., Elberrichi, Z. (2020). Enhanced ant colony algorithm for best features selection for a decision tree classification of medical data. In *Critical Approaches to Information Retrieval Research: IGI Global,* 278–293.
19. Alirezanejad, M., Enayatifar, R., Motameni, H., Nematzadeh, H. (2020). Heuristic filter feature selection methods for medical datasets. *Genomics, 112,* 1173–1181.
20. Wang, S., Yang, M., Li, J., Wu, X., Wang, H. et al. (2017). Texture analysis method based on fractional Fourier entropy and fitness-scaling adaptive genetic algorithm for detecting left-sided and right-sided sensorineural hearing loss. *Fundamenta Informaticae, 151,* 505–521.
21. Rashid, M., Khan, M. A., Sharif, M., Raza, M., Sarfraz, M. M. et al. (2019). Object detection and classification: a joint selection and fusion strategy of deep convolutional neural network and SIFT point features. *Multimedia Tools and Applications, 78,* 15751–15777.

22. Urbanowicz, R. J., Olson, R. S., Schmitt, P., Meeker, M., Moore, J. H. (2018). Benchmarking relief-based feature selection methods for bioinformatics data mining. *Journal of Biomedical Informatics, 85,* 168–188.
23. Naftchali, R. E., Abadeh, M. S. (2017). A multi-layered incremental feature selection algorithm for adjuvant chemotherapy effectiveness/futileness assessment in non-small cell lung cancer. *Biocybernetics and Biomedical Engineering, 37,* 477–488.
24. Rehman, A., Khan, M. A., Mehmood, Z., Saba, T., Sardaraz, M. et al. (2020). Microscopic melanoma detection and classification: a framework of pixel-based fusion and multilevel features reduction. *Microscopy Research and Technique, 83,* 410–423.
25. Saba, T., Khan, M. A., Rehman, A., Marie-Sainte, S. L. (2019). Region extraction and classification of skin cancer: a heterogeneous framework of deep CNN features fusion and reduction. *Journal of Medical Systems, 43(9),* 289.
26. Khan, M. A., Sharif, M. I., Raza, M., Anjum, A., Saba, T. et al. (2019). Skin lesion segmentation and classification: a unified framework of deep neural network features fusion and selection. *Expert Systems,* e12497.
27. Khan, M. A., Rubab, S., Kashif, A., Sharif, M. I., Muhammad, N. et al. (2020). Lungs cancer classification from CT images: an integrated design of contrast based classical features fusion and selection. *Pattern Recognition Letters, 129,* 77–85.
28. Khan, S. A., Nazir, M., Khan, M. A., Saba, T., Javed, K. et al. (2019). Lungs nodule detection framework from computed tomography images using support vector machine. *Microscopy Research and Technique, 82,* 1256–1266.
29. Sharif, M. I., Li, J. P., Khan, M. A., Saleem, M. A. (2020). Active deep neural network features selection for segmentation and recognition of brain tumors using MRI images. *Pattern Recognition Letters, 129,* 181–189.
30. Khan, M. A., Lali, I. U., Rehman, A., Ishaq, M., Sharif, M. et al. (2019). Brain tumor detection and classification: a framework of marker-based watershed algorithm and multilevel priority features selection. *Microscopy Research and Technique, 82,* 909–922.
31. Khan, M. A., Sharif, M., Akram, T., Yasmin, M., Nayak, R. S. (2019). Stomach deformities recognition using rank-based deep features selection. *Journal of Medical Systems, 43(12),* 329.
32. Khan, M. A., Khan, M., Ahmed, A., Mittal, F., Goyal, M. et al. (2020). Gastrointestinal diseases segmentation and classification based on duo-deep architectures. *Pattern Recognition Letters, 131,* 193–204.
33. Zhang, Y., Wang, S., Sui, Y., Yang, M., Liu, B. et al. (2018). Multivariate approach for Alzheimer's disease detection using stationary wavelet entropy and predator-prey particle swarm optimization. *Journal of Alzheimer's Disease, 65,* 855–869.
34. Wang, S. H., Zhang, Y., Li, Y. J., Jia, W. J., Liu, F. Y. et al. (2018). Single slice based detection for Alzheimer's disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization. *Multimedia Tools and Applications, 77,* 10393–10417.
35. Li, Z., Wang, S. H., Fan, R. R., Cao, G., Zhang, Y. D. et al. (2019). Teeth category classification via seven-layer deep convolutional neural network with max pooling and global average pooling. *International Journal of Imaging Systems and Technology, 29,* 577–583.
36. Inbarani, H. H., Azar, A. T., Jothi, G. (2014). Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis. *Computer Methods and Programs in Biomedicine, 113,* 175–185.
37. Shilaskar, S., Ghatol, A. (2013). Feature selection for medical diagnosis: evaluation for cardiovascular diseases. *Expert Systems with Applications, 40,* 4146–4153.
38. Agnihotri, D., Verma, K., Tripathi, P. (2017). Variable global feature selection scheme for automatic classification of text documents. *Expert Systems with Applications, 81,* 268–281.
39. Selvaraj, S., Natarajan, J. (2011). Microarray data analysis and mining tools. *Bioinformation, 6,* 95.
40. Pavithra, V., Jayalakshmi, V. (2019). *A review on predicting cardiovascular diseases using data mining techniques*. pp. 374–380. Springer Cham.

41. Parmar, C., Barry, J. D., Hosny, A., Quackenbush, J., Aerts, H. J. (2018). Data analysis strategies in medical imaging. *Clinical Cancer Research, 24,* 3492–3499.
42. López-Pineda, A., Rodríguez-Moran, M. F., Álvarez-Aguilar, C., Valle, S. M. F., Acosta-Rosales, R. et al. (2018). Data mining of digitized health records in a resource-constrained setting reveals that timely immunophenotyping is associated with improved breast cancer outcomes. *BMC Cancer, 18,* 1–10.
43. Bellman, R. (1957). *Dynamic programming*. p. 95. Princeton, NJ: Princeton University Press.
44. Remeseiro, B., Bolon-Canedo, V. (2019). A review of feature selection methods in medical applications. *Computers in Biology and Medicine, 112,* 103375.
45. Lalka, V. S., Kundeti, S. R., Kumar, V., Vijayananda, J. (2018). A comparative study of feature selection methods for classification of chest X-ray image as normal or abnormal inside AWS ECS cluster. *IEEE International Conference on Cloud Computing in Emerging Markets,* 7–13.
46. Tohka, J., Moradi, E., Huttunen, H., AsDN Initiative (2016). Comparison of feature selection techniques in machine learning for anatomical brain MRI in dementia. *Neuroinformatics, 14,* 279–296.
47. Rosati, S., Molinari, F., Balestra, G. (2011). Feature selection applied to ultrasound carotid images segmentation. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* 5161–5164.
48. Delzell, D. A. P., Peter, T., Smith, M., Magnuson, S., Smith, B. J. (2019). Machine learning and feature selection methods for disease classification with application to lung cancer screening image data. *Frontiers in Oncology, 9,* 1393.
49. Li, H., Xu, C., Xin, B., Zheng, C., Zhao, Y. et al. (2019). 18F-FDG PET/CT radiomic analysis with machine learning for identifying bone marrow involvement in the patients with suspected relapsed acute leukemia. *Theranostics, 9(16),* 4730.
50. AI HLEG (2019). High-level expert group on artificial intelligence: ethics guidelines for trustworthy AI. *European Commission,* 0904.
51. Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge and Information Systems, 34,* 483–519.
52. Tadist, K., Najah, S., Nikolov, N. S., Mrabti, F., Zahi, A. (2019). Feature selection methods and genomic big data: a systematic review. *Journal of Big Data, 6,* 79.
53. Wang, L., Wang, Y., Chang, Q. (2016). Feature selection methods for big data bioinformatics: a survey from the search perspective. *Methods, 111,* 21–31.
54. Saeys, Y., Inza, I., Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics, 23,* 2507–2517.
55. Khaire, U. M., Dhanalakshmi, R. (2019). Stability of feature selection algorithm: a review. *Journal of King Saud University-Computer and Information Sciences*.
56. Wang, H., Jing, X., Niu, B. (2017). A discrete bacterial algorithm for feature selection in classification of microarray gene expression cancer data. *Knowledge-Based Systems, 126,* 8–19.
57. Schroder, M., Bogdan, M., Hinterberger, T., Birbaumer, N. (2003). Automated EEG feature selection for brain computer interfaces, *In First International IEEE EMBS Conference on Neural Engineering,* 626–629.
58. Too, J., Abdullah, A. R., Mohd Saad, N., Tee, W. (2019). EMG feature selection and classification using a pbest-guide binary particle swarm optimization. *Computation, 7(1),* 12.
59. Hira, Z. M., Gillies, D. F. (2015). A review of feature selection and feature extraction methods applied on microarray data. *Advances in Bioinformatics,* 13.
60. 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,* 14863–14868.
61. Fakoor, R., Ladhak, F., Nazi, A., Huber, M. (2013). *Using deep learning to enhance cancer diagnosis and classification*. New York, USA: ACM.
62. Manonmani, M., Balakrishnan, S. (2020). Review of optimization-based feature selection algorithms on healthcare dataset. *Emerging Research in Data Engineering Systems and Computer Communications*. pp. 239–245. Singapore: Springer.

63. Almuallim, H., Dietterich, T. G. (1991). Learning with many irrelevant features. *Citeseer,* 547–552.
64. Almuallim, H., Dietterich, T. G. (1992). Efficient algorithms for identifying relevant features. *Proceedings of the Ninth Canadian Conference on Artificial Intelligence.* Citeseer.
65. Nandi, G. (2011). An enhanced approach to Las Vegas Filter (LVF) feature selection algorithm. *2nd National Conference on Emerging Trends and Applications in Computer Science,* 1–3.
66. Chen, X., Xia, X., Liu, B. (2014). An analysis of hybrid feature selection for improved medical data classification. *Journal of Computational Intelligence and Electronic Systems, 3,* 44–49.
67. Elfangary, L. M. S., Atteya, W. A. (2010). Mining medical databases using incremental enhanced association rules algorithm.
68. Gallego, A. J., Calvo-Zaragoza, J., Fisher, R. B. (2020). Incremental unsupervised domain-adversarial training of neural networks. arXiv preprint arXiv: 200104129.
69. Li, X., Malikić, S., Schäffer, A., Gertz, E., Day, C. et al. (2020). PhISCS-BnB: a fast branch and bound algorithm for the perfect tumor phylogeny reconstruction problem.
70. Gmys, J., Mezmaz, M., Melab, N., Tuyttens, D. (2020). A computationally efficient branch-and-bound algorithm for the permutation flow-shop scheduling problem. *European Journal of Operational Research*.
71. McMahon, G., Burton, P. (1967). Flow-shop scheduling with the branch-and-bound method. *Operations Research, 15,* 473–481.
72. Consul, S., Vikalo, H. (2019). Reconstructing intra-tumor heterogeneity via convex optimization and branch-and-bound search. *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics,* 524–529.
73. He, Y., Zhou, J., Lin, Y., Zhu, T. (2019). A class imbalance-aware relief algorithm for the classification of tumors using microarray gene expression data. *Computational Biology and Chemistry, 80,* 121–127.
74. Chen, Y. X., Tadikamalla, P. R., Shang, J., Song, Y. (2020). Supply allocation: bi-level programming and differential evolution algorithm for natural disaster relief. *Cluster Computing, 23,* 203–217.
75. Peker, M., Ballı, S., Sağbaş, E. A. (2020). Predicting human actions using a hybrid of ReliefF feature selection and kernel-based extreme learning machine. *Cognitive Analytics, Tools, and Applications: Concepts, Methodologies: IGI Global,* 307–325.
76. Kilicarslan, S., Adem, K., Celik, M. (2020). Diagnosis and classification of cancer using hybrid model based on ReliefF and convolutional neural network. *Medical Hypotheses, 137,* 109577.
77. Bugata, P., Drotar, P. (2020). On some aspects of minimum redundancy maximum relevance feature selection. *Science China Information Sciences, 63,* 1–15.
78. Rana, M. M., Ahmed, K. (2020). *Feature selection and biomedical signal classification using minimum redundancy maximum relevance and artificial neural network*. pp. 207–214. Berlin: Springer.
79. Che, J., Yang, Y., Li, L., Bai, X., Zhang, S. et al. (2017). Maximum relevance minimum common redundancy feature selection for nonlinear data. *Information Sciences, 409,* 68–86.
80. Maji, P., Paul, S. (2011). Rough set based maximum relevance-maximum significance criterion and gene selection from microarray data.
81. Liu, W., Zhu, W., Liao, B., Chen, X. (2016). Gene regulatory network inferences using a maximum-relevance and maximum-significance strategy. *PLoS One, 11,* e0166115.
82. Li, Z., Gu, W. (2015). A redundancy-removing feature selection algorithm for nominal data. *PeerJ Computer Science, 1,* e24.
83. Bennasar, M., Hicks, Y., Setchi, R. (2015). Feature selection using joint mutual information maximisation. *Expert Systems with Applications, 42,* 8520–8532.
84. Lan, T., Erdogmus, D., Adami, A., Pavel, M. (2005). Feature selection by independent component analysis and mutual information maximization in EEG signal classification. *IEEE International Joint Conference on Neural Networks,* 3011–3016.
85. Hacilar, H., Nalbantoglu, O. U., Aran, O., Bakir-Gungor, B. (2020). Inflammatory bowel disease biomarkers of human gut microbiota selected via ensemble feature selection methods. arXiv preprint arXiv: 200103019.

86. Marcano-Cedeno, A., Quintanilla-Domínguez, J., Cortina-Januchs, M., Andina, D. (2010). Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network. *IEEE,* 2845–2850.
87. Bouatmane, S., Roula, M. A., Bouridane, A., Al-Maadeed, S. (2011). Round-Robin sequential forward selection algorithm for prostate cancer classification and diagnosis using multispectral imagery. *Machine Vision and Applications, 22,* 865–878.
88. Liu, H., Jiang, H., Zheng, R. (2016). The hybrid feature selection algorithm based on maximum minimum backward selection search strategy for liver tissue pathological image classification. *Computational and Mathematical Methods in Medicine,* 9.
89. Tsymbal, A., Cunningham, P., Pechenizkiy, M., Puuronen, S. (2003). Search strategies for ensemble feature selection in medical diagnostics. *16th IEEE Symposium Computer-Based Medical System,* 124–129.
90. Ververidis, D., Kotropoulos, C. (2008). Fast and accurate sequential floating forward feature selection with the Bayes classifier applied to speech emotion recognition. *Signal Processing, 88,* 2956–2970.
91. Choi, K. S., Zeng, Y., Qin, J. (2012). Using sequential floating forward selection algorithm to detect epileptic seizure in EEG signals. *IEEE 11th International Conference on Signal Processing, 3,* 1637–1640.
92. Sun, Y., Babbs, C., Delp, E. (2006). A comparison of feature selection methods for the detection of breast cancers in mammograms: adaptive sequential floating search *vs*. genetic algorithm. *IEEE,* 6532–6535.
93. Somol, P., Pudil, P., Novovičová, J., Paclık, P. (1999). Adaptive floating search methods in feature selection. *Pattern Recognition Letters, 20,* 1157–1163.
94. Tarle, B., Chintakindi, S., Jena, S. (2020). Integrating multiple methods to enhance medical data classification. *Evolving Systems, 11,* 133–142.
95. Nakariyakul, S., Casasent, D. P. (2009). An improvement on floating search algorithms for feature subset selection. *Pattern Recognition, 42,* 1932–1940.
96. Levner, I. (2005). Feature selection and nearest centroid classification for protein mass spectrometry. *BMC Bioinformatics, 6(1),* 68.
97. Wosiak, A., Zakrzewska, D. (2014). Feature selection for classification incorporating less meaningful attributes in medical diagnostics. *Federated Conference on Computer Science and Information Systems,* 235–240.
98. Zhou, X., Tuck, D. P. (2007). MSVM-RFE: extensions of SVM-RFE for multiclass gene selection on DNA microarray data. *Bioinformatics, 23,* 1106–1114.
99. Khan, M. A., Sharif, M., Javed, M. Y., Akram, T., Yasmin, M. et al. (2017). License number plate recognition system using entropy-based features selection approach with SVM. *IET Image Processing, 12,* 200–209.
100. Peng, Y., Xuefeng, Z., Jianyong, Z., Yumhong, X. (2009). Lazy learner text categorization algorithm based on embedded feature selection. *Journal of Systems Engineering and Electronics, 20,* 651–659.
101. Wang, S. H., Lv, Y. D., Sui, Y., Liu, S., Wang, S. J. et al. (2018). Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling. *Journal of Medical Systems, 42(1),* 2.
102. Chen, L., Ying, Z., Xing, L. (2020). Information gain regression analysis of diagnostic value of CT in severe acute pancreatitis. *Journal of Medical Imaging and Health Informatics, 10,* 60–64.
103. Martín-Valdivia, M. T., Díaz-Galiano, M. C., Montejo-Raez, A., Ureña-López, L. (2008). Using information gain to improve multi-modal information retrieval systems. *Information Processing & Management, 44,* 1146–1158.
104. Lang, C., Jia, H. (2019). Kapur's entropy for color image segmentation based on a hybrid whale optimization algorithm. *Entropy, 21(3),* 318.
105. Ben-Naim, A. (2017). Entropy, Shannon's measure of information and Boltzmann's H-theorem. *Entropy, 19(2),* 48.

106. Shaheen, M., Shahbaz, M., Guergachi, A., Rehman, Z. (2011). Mining sustainability indicators to classify hydrocarbon development. *Knowledge-Based Systems, 24,* 1159–1168.
107. Kirch, W. (2008). *Encyclopedia of public health: Volume 1: A-H Volume 2: I-Z*. New York: Springer Science & Business Media.
108. Härdle, W., Simar, L. (2007). *Applied multivariate statistical analysis*. New York: Springer.
109. Mahendru, S., Agarwal, S. (2019). Feature selection using metaheuristic algorithms on medical datasets. *Harmony Search and Nature Inspired Optimization Algorithms*. pp. 923–937. Singapore: Springer.
110. Rodriguez-Lujan, I., Huerta, R., Elkan, C., Santa Cruz, C. (2010). Quadratic programming feature selection. *Journal of Machine Learning Research, 11,* 1491–1516.
111. Javidi, M. M., Mansoury, S. (2017). Diagnosis of the disease using an ant colony gene selection method based on information gain ratio using fuzzy rough sets. *Journal of Particle Science & Technology, 3,* 175–186.
112. Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., Moore, J. H. (2018). Relief-based feature selection: introduction and review. *Journal of Biomedical Informatics, 85,* 189–203.
113. Burton, G. A., Chapman, P. M., Smith, E. P. (2002). Weight-of-evidence approaches for assessing ecosystem impairment. *Human and Ecological Risk Assessment, 8,* 1657–1673.
114. Nasir, M., Attique Khan, M., Sharif, M., Lali, I. U., Saba, T. et al. (2018). An improved strategy for skin lesion detection and classification using uniform segmentation and feature selection based approach. *Microscopy Research and Technique, 81,* 528–543.
115. Liaqat, A., Khan, M. A., Shah, J. H., Sharif, M., Yasmin, M. et al. (2018). Automated ulcer and bleeding classification from WCE images using multiple features fusion and selection. *Journal of Mechanics in Medicine and Biology, 18,* 1850038.
116. Khan, M. A., Akram, T., Sharif, M., Shahzad, A., Aurangzeb, K. et al. (2018). An implementation of normal distribution based segmentation and entropy controlled features selection for skin lesion detection and classification. *BMC Cancer, 18(1),* 638.
117. Akram, T., Khan, M. A., Sharif, M., Yasmin, M. (2018). Skin lesion segmentation and recognition using multichannel saliency estimation and M-SVM on selected serially fused features. *Journal of Ambient Intelligence and Humanized Computing,* 1–20.
118. Sharif, M., Tanvir, U., Munir, E. U., Khan, M. A., Yasmin, M. (2018). Brain tumor segmentation and classification by improved binomial thresholding and multi-features selection. *Journal of Ambient Intelligence and Humanized Computing,* 1–20.
119. Afza, F., Khan, M. A., Sharif, M., Rehman, A. (2019). Microscopic skin laceration segmentation and classification: a framework of statistical normal distribution and optimal feature selection. *Microscopy Research and Technique, 82,* 1471–1488.
120. Khan, M. A., Javed, M. Y., Sharif, M., Saba, T., Rehman, A. (2019). Multi-model deep neural network based features extraction and optimal selection approach for skin lesion classification. *International Conference on Computer and Information Sciences,* 1–7.
121. Sharif, M., Khan, M. A., Rashid, M., Yasmin, M., Afza, F. et al. (2019). Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images. *Journal of Experimental & Theoretical Artificial Intelligence,* 1–23.
122. Khan, M. A., Rashid, M., Sharif, M., Javed, K., Akram, T. (2019). Classification of gastrointestinal diseases of stomach from WCE using improved saliency-based method and discriminant features selection. *Multimedia Tools and Applications, 78,* 27743–27770.
123. Khan, M. A., Sharif, M., Akram, T., Bukhari, S. A. C., Nayak, R. S. (2020). Developed Newton–Raphson based deep features selection framework for skin lesion recognition. *Pattern Recognition Letters, 129,* 293–303.
124. Wang, S. H., Cheng, H., Phillips, P., Zhang, Y. D. (2018). Multiple sclerosis identification based on fractional Fourier entropy and a modified Jaya algorithm. *Entropy, 20(4),* 254.
125. Wang, S. H., Muhammad, K., Lv, Y., Sui, Y., Han, L. et al. (2018). Identification of alcoholism based on wavelet Renyi entropy and three-segment encoded Jaya algorithm. *Complexity, 2018,* 3198184.

126. Sharif, M., Khan, M. A., Faisal, M., Yasmin, M., Fernandes, S. L. (2018). A framework for offline signature verification system: best features selection approach. *Pattern Recognition Letters*.
127. Khan, M. A., Akram, T., Sharif, M., Saba, T., Javed, K. et al. (2019). Construction of saliency map and hybrid set of features for efficient segmentation and classification of skin lesion. *Microscopy Research and Technique, 82,* 741–763.
128. Raghupathi, W., Tan, J. (2002). Strategic IT applications in health care. *Communications of the ACM, 45,* 56–61.
129. Bellman, R. E. (1957). *Dynamic programming, ser. Rand Corporation research study*. Princeton: Princeton University Press.
130. Wang, S., Sun, J., Mehmood, I., Pan, C., Chen, Y. et al. (2020). Cerebral micro-bleeding identification based on a nine-layer convolutional neural network with stochastic pooling. *Concurrency and Computation: Practice and Experience, 32,* e5130.
131. Kang, C., Yu, X., Wang, S. H., Guttery, D., Pandey, H. et al. (2020). A heuristic neural network structure relying on fuzzy logic for images scoring. *IEEE Transactions on Fuzzy Systems*.
132. Chong, D. Y., Kim, H. J., Lo, P., Young, S., McNitt-Gray, M. F. et al. (2015). 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),* 144–157.
133. Shi, Y., Gao, Y., Liao, S., Zhang, D., Gao, Y. et al. (2016). A learning-based CT prostate segmentation method via joint transductive feature selection and regression. *Neurocomputing, 173,* 317–331.
134. Sahran, S., Albashish, D., Abdullah, A., Abd Shukor, N., Pauzi, S. H. M. (2018). Absolute cosine-based SVM-RFE feature selection method for prostate histopathological grading. *Artificial Intelligence in Medicine, 87,* 78–90.