



A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction

Rizgar R. Zebari^{1,*}, Adnan Mohsin Abdulazeez², Diyar Qader Zeebaree³, Dilovan Asaad Zebari⁴, Jwan Najeeb Saeed⁵

¹IT Department. Technical College of Informatics Akre, Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq, rizgar.ramadhan@dpu.edu.krd

²Presidency of Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq, Adnan.mohsin@dpu.edu.krd

^{3,4}Research Center of Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq, dqszeebaree@dpu.edu.krd, dilovan.zebari@dpu.edu.krd

⁵IT Department. Duhok Technical Institute, Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq, jwan.najeeb@dpu.edu.krd

* Correspondence: rizgar.ramadhan@dpu.edu.krd

Abstract

Due to sharp increases in data dimensions, working on every data mining or machine learning (ML) task requires more efficient techniques to get the desired results. Therefore, in recent years, researchers have proposed and developed many methods and techniques to reduce the high dimensions of data and to attain the required accuracy. To ameliorate the accuracy of learning features as well as to decrease the training time dimensionality reduction is used as a pre-processing step, which can eliminate irrelevant data, noise, and redundant features. Dimensionality reduction (DR) has been performed based on two main methods, which are feature selection (FS) and feature extraction (FE). FS is considered an important method because data is generated continuously at an ever-increasing rate; some serious dimensionality problems can be reduced with this method, such as decreasing redundancy effectively, eliminating irrelevant data, and ameliorating result comprehensibility. Moreover, FE transacts with the problem of finding the most distinctive, informative, and decreased set of features to ameliorate the efficiency of both the processing and storage of data. This paper offers a comprehensive approach to FS and FE in the scope of DR. Moreover, the details of each paper, such as used algorithms/approaches, datasets, classifiers, and achieved results are comprehensively analyzed and summarized. Besides, a systematic discussion of all of the reviewed methods to highlight authors' trends, determining the method(s) has been done, which significantly reduced computational time, and selecting the most accurate classifiers. As a result, the different types of both methods have been discussed and analyzed the findings.

Keywords: dimension reduction, dimension reduction techniques, feature selection, feature extraction.

Received: April 29, 2020 / Accepted: May 13, 2020 / Online: May 15, 2020

I. INTRODUCTION

Nowadays, data mining and knowledge discovery have a great role in several digital applications. Knowledge is detected by processing and analyzing a large amount of the previously collected data [1]. Data generated in a huge volume in different fields, and it is on continuous growth in size, complexity, and dimensionality [2, 3]. A dataset with high dimensionality features its numerous features, but few samples have a direct relation with data mining and machine learning tasks [4, 5]. Therefore, these issues of data become a big challenge for extracting potentially useful, and ultimately understandable

patterns or information in almost every data mining task. Also, working in high dimensional data increases the difficulty of knowledge discovery and pattern classification because there are a lot of redundant and irrelevant features. Reducing high dimensional datasets to a low dimensional dataset by filter or remove redundant and noise information is a method to solve this problem, and this is known as dimensionality reduction [6].

Dimensionality reduction is a process for decreasing features' dimensionality, but the data is still present. In the reduced or low dimension dataset, the crucial features remain even if some particular pattern vanishes [7, 8]. Also, it utilizes

to reduce the size of input data and then preserve much variance of essential features compared to the dataset with the larger size. In real-world data, it will become easy to detect and use for data mining applications and gain high accuracy performance [1, 9]. Moreover, the role of dimensionality reduction is to enhance the accuracy and efficiency of the data mining computation, and it is considered as a vital preprocessing step. Furthermore, it provides several advantages such as eliminating irrelevant, redundant patterns in the dataset; as a result, to reduce the time and amount of memory required for processing such data [1, 10]. By reducing the dataset, the quality of data will improve, the algorithm will work efficiently, achieve better accuracy, and pattern design and examination will be clearer for researchers [11]. Additionally, reducing the cost of computing, improving dimensions visualization, and enhancing the results [12, 13].

This work reviewed more than forty articles of feature selection and feature extraction that have been introduced and published in the last three years.

II. DIMENSIONALITY REDUCTION TECHNIQUES

Dimensionality reduction is the operation of transforming the high dimensional representation of data in low dimensional representations. With the massive growth in high dimensional data, the use of various dimensionality reduction techniques has become popular in many areas of use. Moreover, several modern approaches are continually emerging. Dimensionality reduction techniques transform the original dataset having high dimensionality and turn it into a new dataset representing low dimensionality while maintaining as much as possible the original meanings of the data. The low dimensional representation of the original data contributes to solving the dimensionality curse problem. The low dimensional data can be easily analyzed, processed, and visualized [14]. Several benefits can be obtained due to applying the dimensionality reduction techniques applied to a dataset. (i) As the number of dimensions comes down, data storage space can be reduced. (ii) It takes less computation time only. (iii) Redundant, irrelevant, and noisy data can be removed. (iv) Data quality can be improved. (v) Some algorithms do not perform well on a greater number of dimensions taken. So, reducing these dimensions helps an algorithm to work efficiently and improves accuracy. (vi) It is challenging to visualize data in higher dimensions. So, reducing the dimension may allow us to design and examine patterns more clearly. (vii) It simplifies the process of classification and also improves efficiency [15, 16]. Generally, the dimensionality reduction techniques can be classified into two main groups, or in other words, the dimensionality reduction is achieved through two different techniques: feature selection and feature extraction. In feature selection, information can be lost since some features should be excluded when the process of feature subset choice by doing this information can be reduced. However, in feature extraction, the dimension can be decreased without losing much initial feature dataset [2, 10, 14, 17]. Table I provides a descriptive summary of the methods of dimension reduction.

TABLE I. THE SUMMARY OF DIMENSION REDUCTION TECHNIQUES

Method	Main concept	Pros	Cons
Feature extraction	Summarize the dataset by creating linear combinations of the features	Preserves the original, relative distance between covers latent structure, objects	Not sufficient enough in the existing of a huge number of irrelevant features
Feature selection	A sublist of relevant features can be selected depending on defined criteria	Strong against irrelevant features	Latent structure does not cover

A. Feature Selection

Feature selection is utilized to reduce the dimensionality impact on the dataset through finding the subset of feature which efficiently define the data [18, 19]. It selects the important and relevant features to the mining task from the input data and removes redundant and irrelevant features [20, 21]. It is useful for detecting a good subset of features that is appropriate for the given problem [2, 22]. The main purpose of feature selection is to construct a subset of features as small as possible but represents the whole input data vital features [11, 23]. Feature selection provides numerous advantages: reduce the size of data, decrease needed storage, prediction accuracy improvement, overfitting evading, and reduce executing and training time from easily understanding variables. Feature selection algorithm phase is divided into two-phase such as (i) Subset Generation: (ii) Subset Evaluation: In subset Generation, we need to generate subset from the input dataset and to use Subset Evaluations we have to check whether the generated subset is optimal or not [24, 25]. “Fig.1” shows the overall method of the feature selection process.

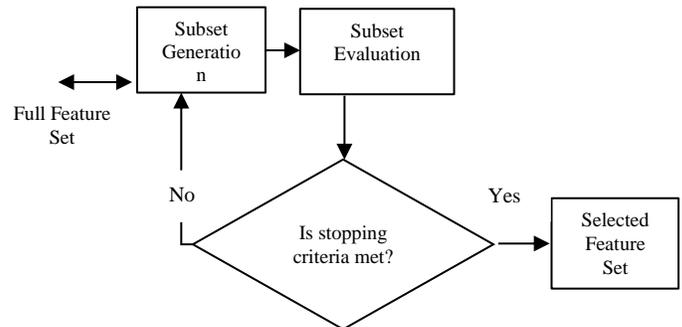


Fig. 1. Process of feature selection.

B. Feature Selection Problems

Various issues can benefit from the feature selection techniques application. High dimension, low sample size data are becoming more popular in different fields. Many of the features of these problems do not facilitate an adequate classification. More so, the imbalance problem happens when one of the two classes has more samples than other classes. Many algorithms neglect the minority sample when concentrating on a major sample classification. However, the minority samples are crucial but seldom occurred. Moreover, in

machine learning, the shift of the dataset is a popular problem that happens when the joint distribution of inputs and outputs varies between training and test stages. A special case of dataset shift, which happens when only the input distribution changes is called Covariate shift. Furthermore, the reduction of the dimensionality and consequently feature selection is one of the most common techniques of noisy data elimination. Eventually, misclassification costs and test costs are the two most significant kinds of cost in cost-sensitive learning [26-28].

C. Feature Selection Methods

Feature selection aims to select a feature subset from the original set of features based on a/the feature’s relevance and redundancy. Originally evaluation methods in feature selection are divided into four kinds: filter, wrapper, embedded [10, 14, 18], and hybrid [20, 29]. Recently, another type of evaluation method is developed, i.e., ensemble feature selection [30, 31]. “Fig. 2” depicts the hierarchy of feature selection techniques.

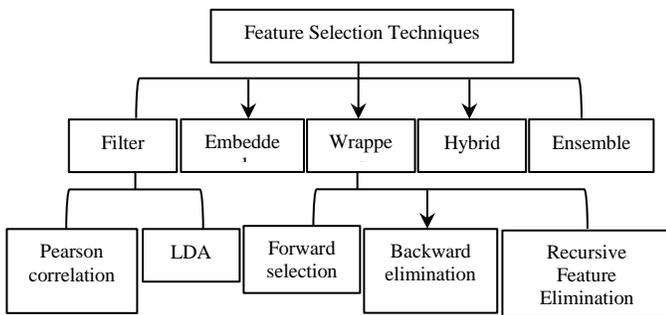


Fig. 2. Hierarchy of feature selection techniques.

Filter is considered the earliest method and also known as an open-loop method. It checks the features relying on the intrinsic characteristics prior to the learning tasks. It mainly measures the feature characteristics depending on four different kinds of measurement criteria, i.e. information, dependency, consistency, and distance [32]. In the filter method, the feature selection process is performed independently of the data mining algorithm. It uses statistical standards for evaluating the ranking of the subset [17, 24]. Moreover, this technique is to perform good performance and high-efficiency computing, easily scalable in high dimensional datasets, and outperformed the wrapper technique. The primary downside of this method is that it neglects the integration between the selected subset and the performance of the induction algorithm [10, 22, 26].

Wrapper, or it also can be called a close-loop method, wraps the feature selection around the learning algorithm and uses the accuracy of the performance or the error rate of the classification process as a criterion of feature evaluation. By decreasing the estimation error of a specific classifier, it chooses the most discriminative subset of features. The wrapper method performs feature selection based on the performance of the learning algorithm; it selects the most optimal feature for the prediction algorithm. Hence it achieves better performance and high accuracy compared to the filter algorithm [22, 27, 33]. The main disadvantage of this approach is computing complexity and more exposure to overfitting in comparison to the filter approach. Most wrapper methods are multivariate;

thus, they need extensive computation times to achieve the convergences and can be intractable for large datasets [33, 34]. “Fig. 3” shows the involved steps in the wrapper method.

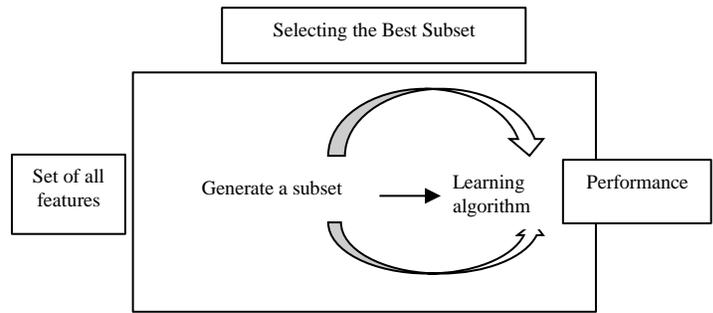


Fig. 3. Wrapper method for feature selection.

Embedded method is a built-in feature selection mechanism that embeds feature selection in the learning algorithm and uses its properties to guide feature evaluation. The embedded method is more effective and more tractable than the wrapper method computationally while retaining similar performance. This is because the embedded method avoids the repeated execution of the classifier and the examination of every feature subset. The embedded method combines the qualities of both filter and wrapper methods. It selects features during the implementation of the mining algorithm, and hence it has less computational expensiveness [32, 35]. Steps involved in the embedded method are shown in “Fig. 4”.

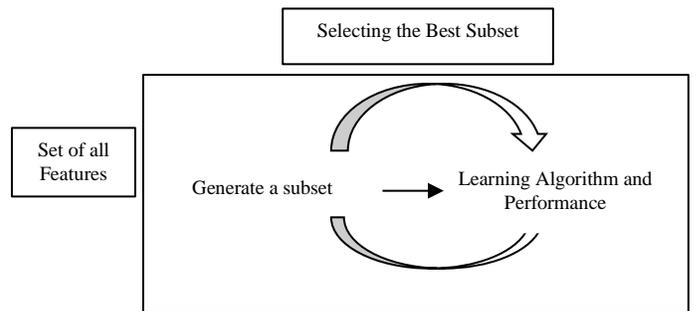


Fig. 4. Embedded method for feature selection.

Hybrid and ensemble methods the recent developments in feature selection can be represented in the hybrid method. Thus, it can be developed either by integrating two various methods (e.g. wrapper and filter), two methods with the same criteria, or two feature selection approaches. In the hybrid method, the advantages of both methods can be inherited by combining their complementary strengths [36]. The combination of filter and wrapper methods is the most common hybrid method [37]. However, **ensemble** method is a method that aims at building a group of feature subsets and then producing an aggregated result out of the group [38]. This method is depending on various subsampling techniques where a particular feature selection method is implemented on a variety of subsamples, and the obtained features are merged to create a more stable subset. Table II describes the advantages and disadvantages of each method

TABLE II. ADVANTAGE AND DISADVANTAGE OF FEATURE SELECTION METHODS.

Method	Advantages	Disadvantages
Filter	<ul style="list-style-type: none"> - It works faster than wrapper - Scalable - Classifier independent - The computational complexity is better than wrapper - Better generalizable property 	<ul style="list-style-type: none"> - The interaction between classifiers is neglected - The dependency among features is ignored
Wrapper	<ul style="list-style-type: none"> - It Interacts with classifier - Consider the dependence among features - Higher performance accuracy than filter 	<ul style="list-style-type: none"> - More prone to over-fitting - Classifier specific - Need expensive computation
Embedded	<ul style="list-style-type: none"> - Interact with classifier - Better computational complexity than wrapper - Higher performance accuracy than filter - Less prone to over-fitting than wrapper - Consider the dependency among features 	<ul style="list-style-type: none"> - Classifier specific
Hybrid	<ul style="list-style-type: none"> - The performance accuracy is higher than filter - Better computational complexity than wrapper - Less prone to over-fitting than wrapper 	<ul style="list-style-type: none"> - Classifier specific
Ensemble	<ul style="list-style-type: none"> - Less exposed to over-fitting - It has more flexibly and robustness upon high dimensional data 	<ul style="list-style-type: none"> - An ensemble of classifiers is not easy to understand

III. FEATURE EXTRACTION

Feature extraction method extracting new features from original dataset, and it is very beneficial when we want to decrease the number of resources required for processing without missing relevant feature dataset. Feature extraction can also decrease the number of additional features for an offered study. Feature extraction produces a remarkable transformation of first features to create more significant features. Feature extraction is a process for creating new features that depend on the original input feature set to decrease the high dimensionality of the feature vector. The transformation method is done by algebraic transformation, and according to some optimization criteria [39, 40]. Also, feature extraction has the ability to handle essential information during dealing with high dimensional issues [41, 42]. These dimensionality reduction techniques aim to not lose a large amount of information during the feature transformation process by conserving the original relative distance between features and cover the original data potential structure [10]. Feature extraction is less exposed to overfitting and perform good accuracy for the classification in comparison to the feature selection methods. However, the data description is lost occasionally after the transformation, and the cost of this process is expensive in several datasets [43, 44].

Feature extraction algorithms are classified into linear and nonlinear algorithms [39, 43, 45]. However, the ideal feature extraction based dimensionality reduction methods is Principal Component Analysis (PCA) [46, 47], Multi-Dimensional Scaling (MDS), Isometric Mapping (ISOMAP), Locally Linear Embedding (LLE), Linear Discriminant Analysis (LDA), Latent Semantic Indexing (LSI) and clustering methods [48, 49]. “Fig. 5” depicts the overall process of feature extraction method.



Fig. 5. The process of feature extraction

In Compare with feature selection, feature extraction diminishes the number of variables by transforming a considerable number of attributes into a set of reduced features. Feature extraction tries to find a significant low-dimensional representation of high-dimensional data. In other words, a lesser information loss and a higher discriminating power can be guaranteed using feature extraction rather than feature selection. However, the application of feature extraction in sentiment analysis has been hindered by several difficulties. First, many feature extraction algorithms that have been classified as nonlinear methods cannot perform the mapping from a high-dimensional space to a low-dimensional space, thereby prohibiting the training of a practically usable classification model and resulting in a loss of data interpretability. An expert system that can automatically determine the sentiment of documents must obtain a low-dimensional representation of new documents during the test phase. Furthermore, unlike most feature selection techniques, feature extraction methods are usually unsupervised; that is, label information cannot be utilized during the process of dimension reduction. When implementing machine learning algorithms for sentiment analysis, those documents with label information must be used to train a predictive model. This same information can also help feature extraction methods create excellent features [45, 50].

In this paper, a comprehensive review is performed for the latest and most efficient methods that have been performed by researchers in the past three years to reduce data dimensions in various areas of machine learning and data mining. Also, the details of each method, such as used algorithms/approaches, datasets, utilized classifiers, and the results obtained are summarized. Moreover, we performed a scientific analysis of the studied feature selection and feature extraction methods. Furthermore, we highlighted the most widely used approaches, the best-reduced dimension, the best decreased computational time, and the highest achieved accuracy methods. This paper is organized as follows: the introduction is done in section I, section II presents the dimensionality reduction techniques then a review about these techniques is done III, the broad discussion is illustrated in section V, while section IV contains the conclusion.

IV. AREVIEW ON DIMENTIONALITY REDUCTION METHODS

A. Feature Selection Methods

Dimensionality reduction utilizes the feature selection methods to select relevant features. In this study, we discuss several recent works for feature selection. The kinds of literature on feature selection methods are summarized in Table III.

Churmonge and Jena [51] proposed a method to address the dimensionality issue based on the clustering combined with correlation filter subset selection. The relevant features were found by the K-means clustering algorithm and redundant features found from clusters and removed by the correlation measure. The presented method used on 8 text and 4 microarray datasets and the Naive Bayes (NB) classifier depended on the classification. Furthermore, the authors compared their method performance with the ReliefF and information gain (IG) feature selection methods relating to the accuracy and computational time. The accuracy of the proposed method outperformed both methods in all datasets except two datasets, and in computational, the proposed was faster than other methods in all datasets.

Tan et al. [52] presented a feature selection method based on the evolutionary algorithm (EA) to reduce the dimensionality of motor imagery brain-computer interface from electroencephalogram (EEG) signals. The subset of important features was generated from each iteration of the EA, while the redundant and insignificant features were eliminated. The experiments were performed in two different datasets: EEG dataset and several machine learning datasets. Also, three classifiers depended: support vector machine (SVM), K-Nearest Neighbor (KNN), and discriminant analysis (DA). Also, the performance of the proposed EA – feature selection method was compared with PCA and independent component analysis (ICA), neighborhood component analysis (NCA), and variable-length particle swarm optimization (VLPSO). The results showed that the introduced methods outperformed all the above methods and could achieve high accuracy even with a small subset of the features.

Hafiz et al. [53] investigated the feature selection issues in the power quality events and proposed a two dimensional PSO feature selection method. They depended on the two dimensional in order to efficiently guide the search space of the particle swarm. The noise measurement against the reduced subset was studied by the Gaussian. The used induction algorithms in this study were KNN and Naïve Bayes. Moreover, the proposed method performance was compared with the Genetic Algorithm (GA), Ant Colony Optimization (ACO), Binary PSO (BPSO), Catfish BPSO, and Chaotic BPSO (CHBPSO). The results have shown that the presented method could find an important and robust feature subset and achieve better accuracy than the above-mentioned methods.

Han et al. [54] worked on the limitation of the local linear embedding (LLE) method to propose an unsupervised feature selection mechanism. They depended on the low dimensional space learning and graph matrix learning. The experiments performed in 15 datasets (6 of them from the UGI machine

learning repository and 9 from the feature selection dataset website). Also, the SVM and decision tree (DT) used as classifiers and. In addition, the presented method compared with eight unsupervised feature selection methods. The results demonstrated that the proposed method accomplished better accuracy regarding the SVM and DT classifier except in two datasets and better stability in both classifiers compared to the 8 feature selection methods.

Niu et al. [55] presented a method to deal with multivariate financial time series nonlinearity inherent to improve the accuracy of forecasting and make the financial decision better. The proposed method involved a feature selection part, deep learning framework, and error correlation part. In the feature selection part, the RReliefF algorithm (which is the enhanced version of the ReliefF) cooperated wrapper-based method to remove the redundant feature. Also, the deep learning part has consisted of long-short term memory (LSTM), gated recurrent unit (GRU), and the optimizer based on adaptive moment estimation (Adam). The deep learning part was trained based on the subset generated by the first part. Furthermore, the error correlation used to enhance the accuracy of the method. The method performance validated on 16 benchmarks and three datasets, and the results have shown its superiority.

Jain and Singh [56] proposed a hybrid feature selection method that consisted of ReliefF and PCA algorithms. First, the weight for each feature was calculated in the used datasets, and a set of satisfying features was generated by the first algorithm. The second algorithm was applied in the generated set. In the proposed method, two types of datasets were considered (text and microarray), and the experiments performed in ten datasets. The performance of the method evaluated in terms of a number of the selected features and computation time. The results indicated that the presented method could achieve better performance in low and high dimensional datasets and reduced half of the irrelevant and redundant features.

Hosseini and Moattar [57] presented a feature subset selection method for imbalanced data classification with high dimensional. The authors focused on the feature space, and the method was based on interaction information to improve the search process. Through each iteration of the method, multiple subsets of the features generated, and the best subset involved in the next iteration. In more detail, the candidate features selected by the Symmetric Uncertainty Algorithm (SUA) first. After the multivariate interaction information used to test the candidate feature subset and based on the dominated relationship the best subset of the features selected. Furthermore, KNN, Naïve Bayes and CART were used as classifiers. The efficiency of the proposed method assessed in 13 datasets from different repositories. Additionally, the performance of the method outperformed 10 other feature selection methods in terms of accuracy and in a number of the reduced features.

Manbari et al. [58] proposed a hybrid unsupervised feature selection method based on the clustering and binary ant system. The procedure of the method executed in two stages: In the first,

the clustering was performed in order to cluster the features and in the second stage and from each cluster the best feature calculated through the iteration of the Ant process. Meanwhile, the second stage of the proposed method repeated several times until the dominated features collected. The presented method compared to seven other unsupervised feature selection methods in eight datasets from UCI and Pablo de Olavide research group. Moreover, the comparison of the method with other methods was performed and the assessment done by using four classifiers (SVM, KNN, DT, and RF). The results demonstrated that the performance of the method was better than other methods and significantly minimized computation time.

Qu et al. [59] proposed feature selection for predicting colorectal cancer based on operational taxonomic units. Three feature selection methods were integrated into the proposed method. In first the subset of most significant operational taxonomic units generated by multiple dimension-reduction methods. Then for reducing the dimensionality and increasing the efficiency of the correlation-based feature selection (CFS) and maximum relevance–maximum distance (MRMD) was used as a combined method. Moreover, according to the taxonomy file, the best features selected. The experiment was performed in two datasets, and three classifiers (RF, Naïve Bayes, and DT) were dependent on evaluation. The results illustrated that the correlation-based feature selection method performed better reduction, and the MRMD required more amount of time and memory for computation. Among all used classifiers, the RF achieved better performance.

Umbarkar and Shukla [60] presented a method to improve the performance of the instruction detection system. They used IG, gain ratio (GR), and CFS algorithms for reducing the dimensionality in the used dataset. To obtain the best-reduced set of features, they divided the original set into several parts, and then each algorithm applied to those parts, and the most accurate was selected. The experiment of the proposed system was performed on KDD-Cup 99 dataset, and DT used as the classifier. The results indicated that the correlation-based feature selection method achieved better performance compared to other used algorithms.

Farokhmanesh and Sadeghi [61] proposed a feature selection method based on sparse feature selection and deep neural networks. Initially, Correntropy-induced, Discriminative Least Squares Regression (DLSR), and Sparse Group Lasso (SGL) three methods of sparse were evaluated and compared. Next, the SGL was integrated with a deep neural network, and the performance of this combination also assessed. Meanwhile, the K-means algorithm was used in the SGL method in order to group features. The nearest neighbor (NN) algorithm was used as a classifier in the performance evaluation of the used techniques. The experiment and valuation process of three methods of the sparse and combination method was performed on the MNIST dataset. The results illustrated that the SGL combined with a deep neural network achieved better accuracy.

Duong and Hoang [62] presented a method based on Histogram of Oriented Gradient (HOG) descriptor and feature selection method to classify the rice quality. They extracted HOG features from the rice image, and the score of each feature was calculated by the fisher score feature selection method. In the proposed method, the VNRIICE dataset was used to perform the experiment of the method. Also, the NN classifier was utilized. The results have shown that the fisher score method enhanced the accuracy by 42% of the classification and the computation time reduced.

Alharan et al. [63] proposed a method based on feature extraction and feature selection methods for texture image classification. Firstly, the set of features was extracted from the used datasets by using three approaches (Gray Level Co-occurrence Matrix (GLCM), Local Binary Pattern (LBP), and Gabor filter). After that, the second stage was performed, which was the evaluation of the extracted features, and the assessment was done by five techniques (info gain, Gain ratio, oneR, ReliefF, and symmetric). Then based on the previous assessment, the feature selection was accomplished by utilizing the K-means clustering algorithm. The presented method experiment was done on three datasets, and SVM, NB, and KNN classifiers were used for the classification. The results showed that the NB and KNN achieved better accuracy in the first dataset while the SVM attained superior accuracy in the second and third datasets.

Osman et al. [64] proposed a model involved in feature extraction and feature selection color-based methods for identifying origin automatically. The presented method was performed in three stages. In the first, the skin color information was extracted from human faces by using skin color detection technique. Next, the wrapper subset evaluator and GA method to eliminate the redundant and irrelevant features. Moreover, 1550 face images from different regions were used by the authors, and six classifiers (NB, Bayes Net, KNN, SVM, RF, and Multilayer Perceptron (MLP)) were used for the classification. The results illustrated that the individual color features accuracy was lower than the accuracy of the combined color features. Also, the accuracy of the SVM, NB, and Bayes Net was very low, and hence they could not be used for the proposed method.

Arshak and Eesa [65] proposed a feature selection method for dimensionality reduction based on the cuttlefish algorithm (CFA) for gene classification. The cuttlefish was used to generate a subset of the optimal features. Also, the KNN was used as a classifier for the evaluation and classification of the proposed method. The experiment was performed on eight different datasets from ELVIRA biomedical dataset repository. The performance of the proposed method was compared with SVM and DT and the hidden Markov model in terms of accuracy and computational time. The result demonstrated that the presented achieved better performance in five datasets compared to the other methods.

Zeebaree et al.[66, 67] proposed a feature selection method based on the Convolutional neural network (CNN) for

classifying and identifying the cancer type in the microarray cancer data. In the presented method, the cancer data was transformed into the array after the files of data were opened. Next, the cancer data organized as matrix vectors, then the CNN was applied for the classification. The experiment was accomplished in ten cancer datasets, and the performance of the method was compared with mSVM-RFE-iRF and varSeIRF methods. The results indicated that the proposed CNN method achieved better classification accuracy compared to the other methods. Also, it outperformed the other methods in terms of decreasing cancer’s genes.

Balasaraswathi [68] introduced a feature selection method for intrusion detection systems based on CAF and membrane computing (MC). In the proposed method membrane computing was integrated with the cuttlefish algorithm aiming to enhance the feature selection process. Moreover, two datasets of the intrusion detection system were dependent on performing the method experiment. The performance of the cuttlefish with and without MC was illustrated. Furthermore, the comparison of the proposed method’s performance and a number of various methods of feature selection was accomplished. The results show the accuracy and computation time of the CAF combined with MC was better than all other methods.

Kaur and Singh [69] proposed a method for image steganalysis based on feature selection and PSO. First, the predominant features were selected by mutual information. Moreover, for selecting dominant features, adaptive PSO was

used by the authors. 10000 stego images have been taken from the BOSSbase dataset for the experiment of the presented method. The classification accuracy of the methods was evaluated with SVM, KNN, and DT classifiers and compared with three other PSO based methods. The results revealed that the proposed method outperformed other methods.

Fatima et al. [70] proposed an optimized feature selection method for detecting malware in the android platform. They used evolutionary GA to reduce the feature dimensionality to 50% of the original dataset and then to train the classifier in order to be capable of detecting the malware features. They used two APks sets (Malware/Good ware) and two classifiers (SVM and neural network). The experimental results demonstrated that the SVM accuracy was 96.6%, and neural network accuracy was 95.2%. The authors conducted the performance of the presented method could be enhanced by utilizing larger datasets.

Widiyanti and Endah [71] worked on a study for recognition of the music emotion based on the feature selection algorithms. In the first several features were extracted from the used datasets. After that, three feature selection algorithms, namely Sequential Forward Selection (SFS), Sequential Backward Selection (SBS), and ReliefF were used to identify emotional features. Some emotion classification has been used, such as sad, angry and happy, etc. the experimental performed on songs dataset and SVM classifier was used to compare the performance of the used algorithms. The results explained that the accuracy of the ReliefF algorithm was lower than other algorithms that obtained a similar accuracy.

TABLE III. FEATURE SELECTION METHODS SUMMARIZATION

Ref.	Year	Dataset	Technique(s)	Computation Time	Classifier(s)	Accuracy
[51]	2018	Text and Microarray	Correlation and K-means	0.5 to 10.24 seconds	NB	The best 99.0.2% and the worst 68.02%
[53]	2018	Power quality	PSO	20.9 seconds	KNN and NB	KNN:99.68% and NB:99.44%
[56]	2018	Text and Microarray	ReliefF and PCA	1 to 29 seconds	-	-
[60]	2018	KDD-Cup 99	IG, GR and correlation	-	DT	Correlation: 92.65%, IG: 92.33% and GR: 92.54%
[64]	2018	Human images	wrapper subset and GA	-	NB, KNN, SVM, RF and MLP	NB: 57%, SVM:60%, KNN:64%, RF: 75% and MLP: 71%
[65]	2018	ELVIRA	CFA	0.049 to 2.11 seconds	KNN	100%
[66]	2018	Different cancer datasets	CNN	-	-	100%
[68]	2018	KDDCUP'99	CFA and MC	0.15 by CFA and 0.11 CFAMC	j48	96.66%
[71]	2018	Song dataset	SFS, SBS and ReliefF	-	SVM	43%
[57]	2019	UCI, KEEL and GitHub	SUA and interaction information	126.57 seconds	KNN, NB and CART	100% for KNN, NB and CART
[58]	2019	UCI	K-means and binary Ant	9.939 seconds	SVM, KNN, DT and RF	-
[59]	2019	[72] and[73]	CFS and MRMD	-	RF, NB, and DT	RF achieved better accuracy
[61]	2019	MNIST	DLSR, SGL and Deep Learning	-	NN	96.77%

[62]	2019	VNRICE	HOG and fisher score	-	NN	93.34%
[63]	2019	Kelberg, And Brodatz	IG, ReliefF, symmetric and K-means	-	SVM, NB and KNN	99.95% by KNN and NB, and 99% by SVM
[70]	2019	APKs	GA	3.76 to 8.57 seconds	SVM and neural network	SVM: 96.6% and neural network: 95.2%
[52]	2020	EEG	EA	-	SVM, KNN and DA	SVM:99.25%, KNN:99.13% and DA:100%
[54]	2020	UGI	LLE	-	SVM and DT	SVM:97.71% and DT:97.20%
[55]	2020	SZCI and DJIA	RReliefF and LSTM	-	-	98.62%

B. Feature Extraction Methods

In previous literature, the dimensionality reduction uses the feature selection methods to select the relevant features have been presented. The remainder of the aim of this paper reviews is to review the latest literature related to feature extraction and dimensionality reduction techniques. Table IV illustrates the summary of the recent literature.

Moghaddam et al. [41] proposed a method known as spectral segmentation and integration (SSI) as supervised feature extraction for hyperspectral images. The developed method divided pixels' spectral signature curve to channels. Then a mean weighted operator was used for integration of each channel band in order to extract new features in a very minimal number compared to the original bands. Moreover, the PSO algorithm was used to merge spectral signature curve pixel segments so as to reduce the dimensionality of the image and to increase the class accuracy. In the proposed technique, the SVM was used as a classifier, and two datasets were used. The experimental results confirmed the SSI method outperformed other feature extraction methods such as PCA, SRS, NWF, DAFE, PCA, SELD, BCC, and CBF.

Berber [74] worked on malignant masses in mammograms based on the feature extraction. The researcher presented Gray Level Co-occurrence Matrix (GLCM) texture feature extraction by three hybrid methods that were used in the proposed method. The three hybrid methods called Wavelet CT1, Wavelet CT2, and ST-GLCM. The interesting point of the image was divided into sub-image then contrast stretching stage was used prior to feature extraction. Then the sub-image has been applied for the methods of feature extraction. Next, the GLCM extracted the seven-feature texture and have been merged with seven statistical features. Moreover, two datasets images were used in this research and SVM classifier. The proposed methods outperformed the multi-resolution feature extractions methods in terms of the number of the extracted feature. Also, in Area under the Curve (AUC) measure, the researcher methods were superior to other feature extraction methods.

Rahman et al. [75] worked on the emotion recognition task. They used PCA and t-statistical to reduce the dimensionality of extracted features from emotional signals of EEG. The proposed method was applied to the dataset called SJTU emotion EEG. The emotional state with extracted features has been classified by four classifiers: SVM, ANN, LDA, and KNN.

The result indicated that the prepared method outperformed several other emotion recognition methods. Chu et al. [76] proposed a system for extracting features from the machinery fault-based vibration signals. They used three feature extraction methods: Fourier transform frequency spectrum (FTFS), envelopment analysis, and local mean decomposition (LMD). The vibration signals were analyzed by the Fourier transform analysis to get amplitude and phase spectrums. The results have given that the envelope analysis and mean decomposition methods could extract between cancerous and non-cancerous tumors in the breast. The authors used a new threshold to improve LBP texture features and the LBP descriptor for identifying the abnormal cases. In the proposed method, the features extracted by using CNN and SVM were used for the classification. The experimental results have explained that the developed method could classify the ultrasound images with high accuracy and sensitivity.

Li et al. [77] worked on fault diagnosis and used discriminative graph regularized autoencoder (DGAE) to design a feature extraction method. To map process data to the feature space, to avoid manually designing feature problems and to ensure that the data characteristics is truly reflected by the learned feature reflect they used advanced neural network structure. Furthermore, the neural network structure model is integrated with the graph to learn internal representation and to preserve locality. Also, to improve the performance of the classification, training samples for the label information were embedded to the graph. NN was used as a classifier. In comparison with other fault diagnosis feature extraction methods the proposed method achieved better performance.

Nagarajan et al. [78] used Empirical Mode Decomposition (EMD) to propose two feature extraction methods of mammogram image. The interesting point of the images divided into a group of different frequency components. Then they performed their first method based on Bi-dimensional Empirical Mode Decomposition (BEMD). Through these groups, the GLCM and gray level run matrix features were extracted. However, the extracted features by BEMD first feature extraction method were less orthogonal to each other.

Therefore, the researchers proposed the second feature extraction which was a modified version of the first one and called (MBEMD). The SVM and LDA classifier were used in this research. Furthermore, the proposed method applied in different databases and obtained steady performance.

Rabin et al. [79] used a short-time Fourier transform (STFT) to extract features from the human hand movement-based signal of electromyography (EMG). Because the space of the extracted features was large, they used PCA and diffusion maps (DM) to reduce the dimensionality of the extracted features. Also, to perform the comparison for the performance of both methods with different sizes of the training set. The researchers used KNN as a classifier in this study. The result indicated the DM technique outperformed the PCA method in case of the limited training set. However, in the large training set, both methods achieved high performance in the term of the classification.

Kuncan et al. [80] worked on the diagnosis of bearing fault for the rotary machine. They extracted features from the vibration signals of bearing by applying their proposed method called a one-dimensional ternary pattern (1D-TP). The signals were collected from three different datasets in size, speed, and parts. Moreover, Random Forest (RF), KNN, SVM, Bayes Net, and ANN were used as classifiers with the ternary pattern. The results indicated that the proposed method could extract efficient features from vibration signals for classification. Also, all used classifiers attained effective accuracy.

Liu et al. [81] built a method for feature extractions based on the incorporation of the discriminant analysis and the low ranks representation of the original data samples. The supervised proposed feature extraction method is called discriminative low rank preserving projection (DLRPP). The presented method performance was compared with seven other feature extraction methods (LPP, LSDA, DPSR, LRRDP, LSPP, LRPP and FOLP) on images of six different datasets. The DLRPP method achieved better performance in the term of recognition rate.

Ma and Yuan [82] proposed a method for extraction features from images based on the deep CNN and PCA. They used a neural network to extract features. Due to the high dimension of the extracted features, they improved and optimized the PCA algorithm by deep learning through simulation experiments. Then the researchers compared the performance of the PCA before and after improvement. The memory usage before optimizing the algorithms was more than 6000 MB and after optimizing the memory utilization decreased to less than 1000 MB. Also, there was a big difference in the time consumed by the PCA algorithm before and after optimization. Rather than the performance of the improved PCA was effective, also the classifier accuracy was enhanced which was done by the SVM algorithm.

Sellami and Farah [83] presented a combination of feature extraction and band selection methods to reduce the dimensionality of hyperspectral images. They used several feature extraction methods such as PCA, TLPP, KPCA and LE as linear, nonlinear methods. Also, the researchers utilized MI, DM, CBS, and PA supervised and unsupervised band selection methods. The SVM algorithm was used as a classifier in the proposed implementation. They combine both groups of the algorithms as follows: TLPP/CBS, PCA/CBS, KPCA/BS, PCA/MI, LE/CBS, TLPP/MI, LE/MI, and KPCA/MI. The authors used two datasets (the first dataset's images were from Indian Pines and the second dataset's images were from Pavia

University). The results revealed that the classification accuracy of the TLPP/CBS was better compared to all other combination methods for both used datasets.

Chen et al. [84] proposed a study on X-ray images for determining bone age. They used a deep neural network to extract features from the X-ray images. The extracted features include: Glutamate cysteine ligase modifier subunit (GCLM) features and Local Binary Patterns (LBP) features in the image. In more detail, the features of the X-ray hand bone image could be automatically extracted using deep learning and the bone age was assessed automatically by the convolution neural network. Moreover, the PCA algorithm was used to reduce the dimensionality of the extracted features. The extracted features were classified by the SVM classifier. The test data, training data and verify data established by a captured image from several males and females and different ages. The results proved that the presented study achieved better performance compared to other methods in this field.

Jin et al. [85] proposed a feature extraction algorithm for JPEG steganalysis based on the adaptive scale adjustment algorithm. In this algorithm, the scale of feature extraction was adjusted adaptively according to the quality of JPEG images. They mainly depended on the Boss Base 1.01 database, and they applied their algorithm on the MD-CFR feature. The result has shown that the performance of the steganalysis improved by the proposed method. Also, the dimensionality of extracted features was reduced by the presented method and hence it could be used based on residual images in the other steganalysis methods.

Liu et al. [67] presented a method for feature extraction depending on graph-based space to construct an optimal algorithm for semi-supervised learning. Particularly the presented method was a combination of sparse representation, discriminative projection, and manifold learning for dimensionality reduction. They designed and performed their method to obtain semi-supervised feature extraction and spars structure local manifolds at the same time. Moreover, the optimal value was accomplished by modifying the similarity matrix in each iteration. The experiment was executed in five datasets and the performance of the presented method compared with six other methods (PCA, MSEC, DLSR, NLDLSR, SOGFS, and SDR). The offered method outperformed all other methods in all used datasets.

Lin et al. [68] worked on the discriminative graph signal to propose a feature extraction method that could extract good features to perform the desired classification. All the training samples in the contained graph were established. Moreover, they used eigenvector decomposition in order to attain the Fourier base of the graph. Numerous discriminative signals were extracted concurrently for achieving high accuracy especially in a problem that has multiple classes. The proposed method was performed in four different experiments with several datasets. The results indicated that the presented methods could achieve encouraging performance, and it was considered to be more effective with supervised classification.

Kasongo and Sun [69] proposed a method for wireless intrusion detection systems based on the feed-forward Deep

Neural Network (DNN). They used a wrapper feature extraction unit with the DNN framework in order to extract the optimal feature vector. Then the extra trees algorithm was used as a classifier. Moreover, two intrusion detection datasets (UNSW-NB15 and the AWID) were dependent as well to examine their methods' efficiency. Moreover, the experiment of the proposed method performed in two different types of attacks, which were binary and multiclass. Furthermore, the researchers compared their method with RF, KNN, SVM, DT, and NB. The proposed method of detection accuracy outperformed the other methods. Whereas the attack classification by feed-forward deep neural network was better than all the above classifiers.

Liu and Sui [70] worked on different methods to minimize the dimensionality in content-based public cultural video retrieval. The public cultural videos' content features extracted by using the combination of the deep learning framework (Caffe) and Alex's net network model. Due to the high dimension of the extracted features, the researchers used PCA to reduce that dimensionality. The researchers examined their work in several videos. The results indicated that the video contents of the used datasets could effectively be compressed by the PCA algorithm, and only minor contents of the video retrieval were lost while lowering the dimensionality of the extracted feature.

Dehzangi and Sahu [71] worked on human activity recognition. They used spectral and temporal analysis to extract features from the Inertial Measurement Unit. Moreover, several feature extraction methods evaluated and particularly the methods based on the time and frequency domain such as power spectral density and Autocorrelation. Also, a number of the classifiers (DT, KNN, SVM, Neural network, and Ensemble bagged) were utilized for human activity recognition that was used in the proposed system. In addition to reducing the dimensionality of the extracted features, the researchers used PCA and KPCA. Although they used feature selection and

transformation methods to reduce dimensionality. The results showed that using the Random forest feature selection method with Ensemble bagged classifier and using Neighborhood component analysis along with Ensemble bagged classifier achieved better accuracy.

Zhang et al. [86] proposed a system for the hyperspectral images to minimize the dimensionality. They depended on the sparse graph and spatial as the integrated method. They utilized PCA and entropy rate in order to divide hyperspectral images to superpixel patch. Moreover, trained data of the graph was constructed by using superpixel segmentation. Then, they extracted spatial-spectral information when the sparse and low-rank graphs generated on the obtained data. After that, to transform the graph embedding to nonlinear space and map the input data into a high-dimensional space, they used the kernel trick. The proposed method was evaluated by two datasets (Indian Pines data set and the University of Pavia (Pavia-U) data set) and the SVM algorithm was dependent as a classifier. The results show that the accuracy of the presented method is higher than other methods.

Alipourfard et al. [87] worked on the hyperspectral images high dimensionality and proposed a system to reduce it. The proposed system was a combination of CNN and the subspace feature extraction method. The authors reduced the dimensionality of the hyperspectral images by the subspace method in order to generate high-quality training samples for the convolutional neural network and for logistic regression that they used as the classifier. Moreover, the presented method was examined by the researchers in two famous two datasets (Indiana Pines and the Pavia University scenes). The experimental results proved that the proposed method accuracy has been improved and achieved higher marks, even under the limited samples of the training samples.

TABLE IV. FEATURE EXTRACTION METHODS SUMMARY

Ref.	Year	Dataset	Technique(s)	Computation Time	Classifier(s)	Accuracy
[74]	2018	DDSM and MIAS	Wavelet and GLCM	-	SVM	97.89%
[76]	2018	-	FTFS and LMD	-	-	-
[83]	2018	Indian Pines and Pavia University	PCA, TLPP, KPCA, and LE	PCA: 10, TLPP: 12, KPCA: 12 and LE: 28	SVM	96.96%
[88]	2018	public cultural	deep learning, Alex net, and PCA	-	-	-
[89]	2018	UCI	Spectral, temporal analysis and PCA	-	DT, KNN, SVM, Neural network and Ensemble bagged	KNN: 95.4%, DT: 90.9%, SVM: 93.2%, Neural Network: 90.6% and Ensemble bagged: 96.9%
[90]	2018	Indian Pines and Pavia	sparse graph, spatial and PCA	-	SVM	93.01%
[91]	2018	Indiana Pines and the Pavia	CNN and subspace feature	-	logistic regression	98.3%
[92]	2019	US breast datasets	LBP and CNN	-	SVM	96%
[80]	2019	TE process	DGAE and neural network	2.47 seconds	NN	83.95%
[78]	2019	MIAS, DDSM and MGM Hospital	EMD	-	SVM and LDA	SVM: 96.2% and LDA: 92.59%

[80]	2019	Artificial Fault Type	1D-TP	0.352 second	RF, KNN, SVM, Bayes Net and ANN	100% for all classifiers
[81]	2019	public image databases	DLRPP		NN	-
[82]	2019	-	CNN and PCA	100 to 1300 seconds	SVM	91.3%
[79]	2020	UCI	STFT, PCA and DM	-	KNN	94.8%
[85]	2020	Boss Base 1.01	adaptive scale adjustment	-	-	-
[93]	2020	[94], [95], [96], and [97]	sparse representation and discriminative projection	-	-	-
[86]	2020	UCI	eigenvector decomposition	-	NN, LDA, NB and SVM	97.5%
[87]	2020	UNSW-NB15 and AVID	wrapper feature extraction and DNN	-	Extra trees	99.77%
[41]	2020	ROSIS and AVIRIS	SSI and PSO	-	SVM	84.83 %
[75]	2020	SJTU	PCA and t-statistical	-	SVM, ANN, LDA and KNN	ANN: 86.57%, SVM: 85.85%, LDA: 82.50% KNN:73.42%
[41]	2020	ROSIS and AVIRIS	SSI and PSO	-	SVM	84.83 %

V. DISCUSSION

In the proposed feature selection methods, different techniques/algorithms used to get the dimensionality reduction of the dataset, minimize computation time, and improve classification accuracy. Through the literature and table III there are three methods ([51, 58, 63]) dependent on the clustering technique using K-means. The authors in [51] used K-means for removing non-relevant features, while [58] in the similarity value was used to separate the features in multiple clusters, and in [44] the algorithm was used to divide the features into the most relevant and noisy clusters. The mentioned three methods were performed in the highly dimension datasets such as text, microarray and texture image classification. In [51] the computational time was reduced to more than 50% as the researchers compared with the ReliefF and IG methods. Also, in the [58], the authors compared the fulfillment of their methods with other three ant colony-based feature selection methods and illustrated that the computation time was better 5-8 times than those of the other methods.

Also, about 50% of the reviewed feature selection methods were based on the optimization approaches. Where these methods used several optimized algorithms such as PSO in [53] and [69], EA in [52], Ant in [58], GA in [64] and [70], CAF in [65] and [68] and deep learning in [61] and [65]. In [53] and [69] PSO algorithm was used to incorporate the features information into search space and hence selecting the most desired features and removing not required ones. In [52] EA was used to reduce the dimensionality of the search space by eliminating the unnecessary features from each iteration process, then influential features were selected at the same time. In [58] after the features clustered, the binary Ant used to rank them from each cluster. Then from each iteration of the Ant process, the wanted set of the important features constructed. The GA was used due to its capability for reducing the dataset dimensionality. According to the research [64] the average of reduction was 93%, and in [70] used feature dataset was reduced by 50%. Also, the CFA has a great role in reducing the dimensionality and computational time in [65] the dataset was reduced up to 93% and in [68] the time was less than compared to other optimization algorithms the authors did a comparison with. The deep neural network was used in order to improve classification accuracy. In the literature two methods ([61] and

[66]) based on the neural network where the accuracy of both of them enhanced about 19% and 8.75%.

The traditional feature selection methods were used in different methods such as ReliefF was utilized in [55], [56], [63], and Information Gain used in [60] and [63]. The important task of the ReliefF approached in the used methods was to select relevant features and in the most researches was utilized with other techniques for example in [55] it used with LSTM, in [56] the PCA algorithm was utilized with it, and in [63] IG, symmetric and K-means were used with it. Similarly, the IG algorithm was used to reduce the dimensionality of the datasets and this technique was used with other approaches such as GR, correlation-based feature and symmetric techniques.

In general, the accuracy of the reviewed methods is varied from approach to another one. However, the optimization-based feature selection methods achieved better performance than the traditional methods. The most used classifiers in the methods are SVM and KNN, both were used in 8 methods and NB was used in 6 studies. In some articles, more than one classifier was used. Nevertheless, the SVM obtained better accuracy compared to others in those manuscripts who used multiple classifiers. Also, there is a variety in the computational time and the role of clustering and optimization algorithms was significant for reducing the computational complexity. Whereas, in research [51] the execution time was better, about 50% to 70% when compared with IG and ReliefF techniques. Moreover, in the researches [65] and [70], execution time was reduced by 40% and 50% by the CFA and GA. However, the best computational time was achieved when the Ant system integrated with the clustering algorithm in the research [58] where it was 5-8 times better compared to the other three ACO algorithms.

On the other hand, 21 feature extraction methods have been reviewed in this study and summarized in Table IV. Among them 7 methods ([92], [77], [82], [84], [87], [88], and [91]) depended on deep learning (CNN and DNN). CNN and LBP were used in research [92] and achieved high-efficiency accuracy. In [77] the neural network was integrated with graph autoencoder. Particularly, the neural network was used to map the process data to feature space. In [82] the PCA algorithm was optimized by CNN to efficiently reduce dimensionality in big

data. As a result, about 70% of the features reduced, executed time minimized and memory utilized decreased and hence the accuracy increased by 3%. Also, in research [84] the accuracy was improved around 7% by using CNN and PCA together. In [87] DNN framework was used with a wrapper feature extraction algorithm. In the used datasets the dimensionality was reduced efficiently, and accuracy increased by about more than 6%. In [88] the dimensionality of the public culture videos was reduced proficiently by the deep learning framework and PCA algorithm. The classification accuracy enhanced about 4% by the integrated CNN and feature subspace reduction in the research [91].

Moreover, the PCA algorithms were used in 8 feature extraction methods [70], [79], [82], [83], [84], [88], [89] and [90] which exist in the literature. In [55] the PCA role was to reduce the redundant information rather than extracting the features. Also, in [79], [83], [84], [88] and [89], the dimensionality of the extracted features were reduced by the PCA. While in research [90] the PCA algorithm was used for extracting the initial component in the process of converting the HSI images into a superpixel patch. In research [41] the PSO was used to minimize the dimensionality and enhance the classification accuracy. The wavelet transform was utilized to decrease the number of extracted features in [74]. Also, the study [76] proved that extracting the features in the strong noise was only accomplished by the FTFS when compared with envelopment analysis and LMD. Further, the SVM algorithm was the most used classifier in the summarized feature extraction methods in this paper, it was used in 12 methods.

The best-achieved accuracy of the reviewed methods and from those who depended on the PCA algorithms is research [82] for a reason; they optimized the algorithm by deep learning. Also, the enhanced PCA by the CNN in the research [82] reduced a great computational time, it was 1300 without CNN and became 100 when the PCA algorithm improved by the CNN. But in the research [41] the computational complexity of the SSI that has been combined with PSO was more compared to the spectral region splitting (SRS).

VI. CONCLUSION

The high dimensionality of data has a direct impact on the learning algorithm, computational time, computer resources (memory), and model accuracy. Therefore, reducing dimensionality and tackling its curse became an exciting topic in search and development areas to provide the most reliable, flexible, and high accurate computerized tools and applications. Hence, several methods and techniques accomplished in the last two decades based on the feature selection and feature extraction.

This paper reviews the most recent studies in several fields such as medical disease analysis, ethnicity identification, emotion recognition, genes classification, text classification, image Steganalysis, data visualization, Hyperspectral images classification, network malware detection and several engineering tasks, etc. moreover, the details used techniques/algorithms, datasets, classifiers approaches were used by the authors and attained results relating to the accuracy and computational time are summarized for each of the feature

selection and feature extraction methods. We observed that the trend of the researchers for reducing the dimensionality based on the feature selection methods is to use the optimization algorithms, and about half of the reviewed researches were relying on the different techniques of optimization. Also, the most used classifiers are the SVM and KNN, and the best-achieved accuracy was the SVM algorithm. On the other hand, for feature extraction methods, CNN and DNN techniques take a great role and have been used in 7 methods of the studied research. While the PCA is still a widely used algorithm in the feature extraction works, it has been used in 8 methods. Additionally, the optimized PCA could achieve better performance in terms of accuracy, computational time, and the number of reduced features.

REFERENCES

- [1] N. Sharma and K. Saroha, "Study of dimension reduction methodologies in data mining," in International Conference on Computing, Communication & Automation, 2015, pp. 133-137: IEEE.
- [2] S. Ayesha, M. K. Hanif, and R. Talib, "Overview and comparative study of dimensionality reduction techniques for high dimensional data," Information Fusion, vol. 59, pp. 44-58, 2020.
- [3] D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, and S. R. Zeebaree, "Combination of K-means clustering with Genetic Algorithm: A review," International Journal of Applied Engineering Research, vol. 12, no. 24, pp. 14238-14245, 2017.
- [4] Z. Cheng and Z. Lu, "A novel efficient feature dimensionality reduction method and its application in engineering," Complexity, vol. 2018, 2018.
- [5] D. A. Zebari, H. Haron, D. Q. Zeebaree, and A. M. Zain, "A Simultaneous Approach for Compression and Encryption Techniques Using Deoxyribonucleic Acid," in 2019 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), 2019, pp. 1-6: IEEE.
- [6] M. Li, H. Wang, L. Yang, Y. Liang, Z. Shang, and H. Wan, "Fast hybrid dimensionality reduction method for classification based on feature selection and grouped feature extraction," Expert Systems with Applications, vol. 150, p. 113277, 2020.
- [7] A. P. Pandian, R. Palanisamy, and K. Ntalianis, Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI-2019). Springer Nature, 2020.
- [8] M. A. Mohammed, B. Al-Khateeb, A. N. Rashid, D. A. Ibrahim, M. K. A. Ghani, and S. A. Mostafa, "Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images," Computers & Electrical Engineering, vol. 70, pp. 871-882, 2018.
- [9] O. Saini and S. Sharma, "A review on dimension reduction techniques in data mining," Computer engineering and intelligent systems, vol. 9, pp. 7-14, 2018.
- [10] N. Abd-alsabour, "On the Role of Dimensionality Reduction," JCP, vol. 13, no. 5, pp. 571-579, 2018.
- [11] S. Velliangiri and S. Alagumuthukrishnan, "A Review of Dimensionality Reduction Techniques for Efficient Computation," Procedia Computer Science, vol. 165, pp. 104-111, 2019.
- [12] W. Wang, W.-g. Shen, Y.-x. Sun, B. Chen, and R. Zhu, "Dimensionality reduction via adjusting data distribution density," in 2018 5th International Conference on Systems and Informatics (ICSAI), 2018, pp. 1052-1055: IEEE.
- [13] J. Stuckman, J. Walden, and R. Scandariato, "The effect of dimensionality reduction on software vulnerability prediction models," IEEE Transactions on Reliability, vol. 66, no. 1, pp. 17-37, 2016.

- [14] M. Verleysen and D. François, "The curse of dimensionality in data mining and time series prediction," in *International Work-Conference on Artificial Neural Networks*, 2005, pp. 758-770: Springer.
- [15] L. Liu and M. T. Özsu, *Encyclopedia of database systems*. Springer New York, NY, USA., 2009.
- [16] A. Juvonen, T. Sipola, and T. Hämäläinen, "Online anomaly detection using dimensionality reduction techniques for HTTP log analysis," *Computer Networks*, vol. 91, pp. 46-56, 2015.
- [17] X. Huang, L. Wu, and Y. Ye, "A Review on Dimensionality Reduction Techniques," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, no. 10, p. 1950017, 2019.
- [18] D. L. Padmaja and B. Vishnuvardhan, "Comparative study of feature subset selection methods for dimensionality reduction on scientific data," in *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, 2016, pp. 31-34: IEEE.
- [19] M. B. Abdulrazzaq and J. N. Saeed, "A Comparison of Three Classification Algorithms for Handwritten Digit Recognition," in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 2019, pp. 58-63: IEEE.
- [20] A. S. Eesa, Z. Orman, and A. M. A. Brifcani, "A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems," *Expert Systems with Applications*, vol. 42, no. 5, pp. 2670-2679, 2015.
- [21] A. S. Eesa, A. M. A. Brifcani, and Z. Orman, "Cuttlefish algorithm-a novel bio-inspired optimization algorithm," *International Journal of Scientific & Engineering Research*, vol. 4, no. 9, pp. 1978-1986, 2013.
- [22] P. Jindal and D. Kumar, "A review on dimensionality reduction techniques," *International journal of computer applications*, vol. 173, no. 2, pp. 42-46, 2017.
- [23] A. S. Eesa, Z. Orman, and A. M. A. Brifcani, "A new feature selection model based on ID3 and bees algorithm for intrusion detection system," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 23, no. 2, pp. 615-622, 2015.
- [24] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *Journal of King Saud University-Computer and Information Sciences*, 2019.
- [25] S. Visalakshi and V. Radha, "A literature review of feature selection techniques and applications: Review of feature selection in data mining," in *2014 IEEE International Conference on Computational Intelligence and Computing Research*, 2014, pp. 1-6: IEEE.
- [26] C. M. Teng, "Combining noise correction with feature selection," in *International Conference on Data Warehousing and Knowledge Discovery*, 2003, pp. 340-349: Springer.
- [27] H. Zhao, F. Min, and W. Zhu, "Cost-sensitive feature selection of numeric data with measurement errors," *Journal of Applied Mathematics*, vol. 2013, 2013.
- [28] J. N. Saeed, "A SURVEY OF ULTRASONOGRAPHY BREAST CANCER IMAGE SEGMENTATION TECHNIQUES," *Academic Journal of Nawroz University*, vol. 9, no. 1, pp. 1-14, 2020.
- [29] Y. Leung and Y. Hung, "A multiple-filter-multiple-wrapper approach to gene selection and microarray data classification," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 7, no. 1, pp. 108-117, 2008.
- [30] C. Lazar et al., "A survey on filter techniques for feature selection in gene expression microarray analysis," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 9, no. 4, pp. 1106-1119, 2012.
- [31] M. R. Mahmood and A. M. Abdulazeez, "A Comparative Study of a New Hand Recognition Model Based on Line of Features and Other Techniques," in *International Conference of Reliable Information and Communication Technology*, 2017, pp. 420-432: Springer.
- [32] M. Dash and H. Liu, "Feature selection for classification," *Intelligent data analysis*, vol. 1, no. 3, pp. 131-156, 1997.
- [33] D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction: A review," *Egyptian Informatics Journal*, vol. 19, no. 3, pp. 179-189, 2018.
- [34] D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, and D. A. Zebari, "Machine learning and Region Growing for Breast Cancer Segmentation," in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 2019, pp. 88-93: IEEE.
- [35] D. Zebari, H. Haron, and S. Zeebaree, "Security Issues in DNA Based on Data Hiding: A Review," *International Journal of Applied Engineering Research*, vol. 12, no. 24, ISSN, pp. 0973-4562, 2017.
- [36] M. M. Kabir, M. M. Islam, and K. Murase, "A new wrapper feature selection approach using neural network," *Neurocomputing*, vol. 73, no. 16-18, pp. 3273-3283, 2010.
- [37] Y. Peng, Z. Wu, and J. Jiang, "A novel feature selection approach for biomedical data classification," *Journal of Biomedical Informatics*, vol. 43, no. 1, pp. 15-23, 2010.
- [38] Q. Shen, R. Diao, and P. Su, "Feature Selection Ensemble," *Turing-100*, vol. 10, pp. 289-306, 2012.
- [39] M. K. Elhadad, K. M. Badran, and G. I. Salama, "A novel approach for ontology-based dimensionality reduction for web text document classification," *International Journal of Software Innovation (IJSI)*, vol. 5, no. 4, pp. 44-58, 2017.
- [40] D. A. Zebari, H. Haron, S. R. Zeebaree, and D. Q. Zeebaree, "Enhance the Mammogram Images for Both Segmentation and Feature Extraction Using Wavelet Transform," in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 2019, pp. 100-105: IEEE.
- [41] S. H. A. Moghaddam, M. Mokhtarzade, and B. A. Beirami, "A feature extraction method based on spectral segmentation and integration of hyperspectral images," *International Journal of Applied Earth Observation and Geoinformation*, vol. 89, p. 102097, 2020.
- [42] D. M. Sulaiman, A. M. Abdulazeez, H. Haron, and S. S. Sadiq, "Unsupervised Learning Approach-Based New Optimization K-Means Clustering for Finger Vein Image Localization," in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 2019, pp. 82-87: IEEE.
- [43] R. Aziz, C. Verma, and N. Srivastava, "Dimension reduction methods for microarray data: a review," *AIMS. Bioengineering*, vol. 4, no. 1, pp. 179-197, 2017.
- [44] A. S. Eesa, A. M. Abdulazeez, and Z. Orman, "A DIDS Based on The Combination of Cuttlefish Algorithm and Decision Tree," *Science Journal of University of Zakho*, vol. 5, no. 4, pp. 313-318, 2017.
- [45] Z. M. Hira and D. F. Gillies, "A review of feature selection and feature extraction methods applied on microarray data," *Advances in bioinformatics*, vol.170. 2015, 2015.
- [46] A. M. Abdulazeez and A. S. Issa, "Intrusion detection system based on neural networks using bipolar input with bipolar sigmoid activation function," *AL-Rafidain Journal of Computer Sciences and Mathematics*, vol. 8, no. 2, pp. 79-86, 2011.
- [47] D. A. Zebari, H. Haron, S. R. Zeebaree, and D. Q. Zeebaree, "Multi-Level of DNA Encryption Technique Based on DNA Arithmetic and Biological Operations," in *2018 International Conference on Advanced Science and Engineering (ICOASE)*, 2018, pp. 312-317: IEEE.
- [48] O. M. S. Hassan, A. M. Abdulazeez, and V. M. TİRYAKİ, "Gait-based human gender classification using lifting 5/3 wavelet and principal component analysis," in *2018 International Conference on Advanced Science and Engineering (ICOASE)*, 2018, pp. 173-178: IEEE.
- [49] F. P. Shah and V. Patel, "A review on feature selection and feature extraction for text classification," in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2016, pp. 2264-2268: IEEE.

- [50] H. Sadeeq, A. Abdulazeez, N. Kako, and A. Abraham, "A Novel Hybrid Bird Mating Optimizer with Differential Evolution for Engineering Design Optimization Problems," in International Conference of Reliable Information and Communication Technology, 2017, pp. 522-534: Springer.
- [51] S. Chormunge and S. Jena, "Correlation based feature selection with clustering for high dimensional data," Journal of Electrical Systems and Information Technology, vol. 5, no. 3, pp. 542-549, 2018.
- [52] P. Tan, X. Wang, and Y. Wang, "Dimensionality reduction in evolutionary algorithms-based feature selection for motor imagery brain-computer interface," Swarm and Evolutionary Computation, vol. 52, p. 100597, 2020.
- [53] F. Hafiz, A. Swain, C. Naik, and N. Patel, "Efficient feature selection of power quality events using two dimensional (2D) particle swarms," Applied Soft Computing, vol. 81, p. 105498, 2019.
- [54] X. Han, P. Liu, L. Wang, and D. Li, "Unsupervised feature selection via graph matrix learning and the low-dimensional space learning for classification," Engineering Applications of Artificial Intelligence, vol. 87, p. 103283, 2020.
- [55] T. Niu, J. Wang, H. Lu, W. Yang, and P. Du, "Developing a deep learning framework with two-stage feature selection for multivariate financial time series forecasting," Expert Systems with Applications, vol. 148, p. 113237, 2020.
- [56] D. Jain and V. Singh, "An efficient hybrid feature selection model for dimensionality reduction," Procedia Computer Science, vol. 132, pp. 333-341, 2018.
- [57] E. S. Hosseini and M. H. Moattar, "Evolutionary feature subsets selection based on interaction information for high dimensional imbalanced data classification," Applied Soft Computing, vol. 82, p. 105581, 2019.
- [58] Z. Manbari, F. AkhlaghianTab, and C. Salavati, "Hybrid fast unsupervised feature selection for high-dimensional data," Expert Systems with Applications, vol. 124, pp. 97-118, 2019.
- [59] K. Qu, F. Gao, F. Guo, and Q. Zou, "Taxonomy dimension reduction for colorectal cancer prediction," Computational biology and chemistry, vol. 83, p. 107160, 2019.
- [60] S. Umbarkar and S. Shukla, "Analysis of heuristic based feature reduction method in intrusion detection system," in 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), 2018, pp. 717-720: IEEE.
- [61] F. Farokhmanesh and M. T. Sadeghi, "Deep Feature Selection using an Enhanced Sparse Group Lasso Algorithm," in 2019 27th Iranian Conference on Electrical Engineering (ICEE), 2019, pp. 1549-1552: IEEE.
- [62] H.-T. Duong and V. T. Hoang, "Dimensionality Reduction Based on Feature Selection for Rice Varieties Recognition," in 2019 4th International Conference on Information Technology (InCIT), 2019, pp. 199-202: IEEE.
- [63] A. F. Alharan, H. K. Fatlawi, and N. S. Ali, "A cluster-based feature selection method for image texture classification," Indonesian Journal of Electrical Engineering and Computer Science, vol. 14, no. 3, pp. 1433-1442, 2019.
- [64] M. Z. Osman, M. A. Maarof, M. F. Rohani, K. Moorthy, and S. Awang, "Multi-Scale Skin Sample Approach for Dynamic Skin Color Detection: An Analysis," Advanced Science Letters, vol. 24, no. 10, pp. 7662-7667, 2018.
- [65] Y. Arshak and A. Eesa, "A New Dimensional Reduction Based on Cuttlefish Algorithm for Human Cancer Gene Expression," in 2018 International Conference on Advanced Science and Engineering (ICOASE), 2018, pp. 48-53: IEEE.
- [66] D. Q. Zebaree, H. Haron, and A. M. Abdulazeez, "Gene selection and classification of microarray data using convolutional neural network," in 2018 International Conference on Advanced Science and Engineering (ICOASE), 2018, pp. 145-150: IEEE.
- [67] O. Ahmed and A. Brifceni, "Gene Expression Classification Based on Deep Learning," in 2019 4th Scientific International Conference Najaf (SICN), 2019, pp. 145-149: IEEE.
- [68] V. Balasaraswathi, "Enhanced Cuttle Fish Algorithm Using Membrane Computing for feature selection of intrusion detection.",vol.10, special issue,2018.
- [69] J. Kaur and S. Singh, "Feature selection using mutual information and adaptive particle swarm optimization for image steganalysis," in 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO), 2018, pp. 538-544: IEEE.
- [70] A. Fatima, R. Maurya, M. K. Dutta, R. Burget, and J. Masek, "Android Malware Detection Using Genetic Algorithm based Optimized Feature Selection and Machine Learning," in 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), 2019, pp. 220-223: IEEE.
- [71] E. Widiyanti and S. N. Endah, "Feature Selection for Music Emotion Recognition," in 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS), 2018, pp. 1-5: IEEE.
- [72] G. Zeller et al., "Potential of fecal microbiota for early - stage detection of colorectal cancer," Molecular systems biology, vol. 10, no. 11, 2014.
- [73] J. Zackular, M. Rogers, and M. Ruffin, "4th, Schloss PD," The human gut microbiome as a screening tool for colorectal cancer. Cancer Prev Res (Phila), vol. 7, no. 11, pp. 1112-21, 2014.
- [74] M. A. Berbar, "Hybrid methods for feature extraction for breast masses classification," Egyptian informatics journal, vol. 19, no. 1, pp. 63-73, 2018.
- [75] M. A. Rahman, M. F. Hossain, M. Hossain, and R. Ahmmed, "Employing PCA and t-statistical approach for feature extraction and classification of emotion from multichannel EEG signal," Egyptian Informatics Journal, vol. 21, no. 1, pp. 23-35, 2020.
- [76] C. Chu, Z. Zuo-xi, K. Xin-rong, and G. Yun-zhi, "The Research of Machinery Fault Feature Extraction Methods Based On Vibration Signal," IFAC-PapersOnLine, vol. 51, no. 17, pp. 346-352, 2018.
- [77] Y. Li, Y. Chai, H. Zhou, and H. Yin, "A novel feature extraction method based on discriminative graph regularized autoencoder for fault diagnosis," IFAC-PapersOnLine, vol. 52, no. 24, pp. 272-277, 2019.
- [78] V. Nagarajan, E. C. Britto, and S. M. Veeraputhiran, "Feature extraction based on empirical mode decomposition for automatic mass classification of mammogram images," Medicine in Novel Technology and Devices, vol. 1, p. 100004, 2019.
- [79] N. Rabin, M. Kahlon, S. Malayev, and A. Ratnovsky, "Classification of human hand movements based on EMG signals using nonlinear dimensionality reduction and data fusion techniques," Expert Systems with Applications, vol. 149, p. 113281, 2020.
- [80] M. Kuncan, K. Kaplan, M. R. Minaz, Y. Kaya, and H. M. Ertunç, "A novel feature extraction method for bearing fault classification with one dimensional ternary patterns",vol.100,p.346-357.ISA transactions, 2020.
- [81] Z. Liu, J. Wang, G. Liu, and L. Zhang, "Discriminative low-rank preserving projection for dimensionality reduction," Applied Soft Computing, vol. 85, p. 105768, 2019.
- [82] J. Ma and Y. Yuan, "Dimension reduction of image deep feature using PCA," Journal of Visual Communication and Image Representation, vol. 63, p. 102578, 2019.
- [83] A. Sellami and M. Farah, "Comparative study of dimensionality reduction methods for remote sensing images interpretation," in 2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 2018, pp. 1-6: IEEE.

- [84] X. Chen, J. Li, Y. Zhang, Y. Lu, and S. Liu, "Automatic feature extraction in X-ray image based on deep learning approach for determination of bone age," *Future Generation Computer Systems*, 2019 Oct 31.
- [85] Z. Jin, G. Feng, Y. Ren, and X. Zhang, "Feature Extraction Optimization of JPEG Steganalysis Based on Residual Images," *Signal Processing*, Vol.170, p. 107455, 2020.
- [86] W. Lin, J. Huang, C. Y. Suen, and L. Yang, "A feature extraction model based on discriminative graph signals," *Expert Systems with Applications*, vol. 139, p. 112861, 2020.
- [87] S. M. Kasongo and Y. Sun, "A deep learning method with wrapper based feature extraction for wireless intrusion detection system," *Computers & Security*, vol. 92, p. 101752, 2020.
- [88] Y. Liu and A. Sui, "Research on Feature Dimensionality Reduction in Content Based Public Cultural Video Retrieval," in *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, 2018, pp. 718-722: IEEE.
- [89] O. Dehzangi and V. Sahu, "IMU-Based Robust Human Activity Recognition using Feature Analysis, Extraction, and Reduction," in *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018, pp. 1402-1407: IEEE.
- [90] X. Zhang et al., "Spatial-Spectral Graph-Based Nonlinear Embedding Dimensionality Reduction for Hyperspectral Image Classification," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 8472-8475: IEEE.
- [91] T. Alipourfard, H. Arefi, and S. Mahmoudi, "A novel deep learning framework by combination of subspace-based feature extraction and convolutional neural networks for hyperspectral images classification," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 4780-4783: IEEE.
- [92] D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, and D. A. Zebari, "Trainable Model Based on New Uniform LBP Feature to Identify the Risk of the Breast Cancer," in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 2019, pp. 106-111: IEEE.
- [93] Z. Liu, Z. Lai, W. Ou, K. Zhang, and R. Zheng, "Structured optimal graph based sparse feature extraction for semi-supervised learning," *Signal Processing*, vol.170, p. 107456, 2020.
- [94] A. M. Martinez, "The AR face database," *CVC Technical Report*24, 1998.
- [95] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image and vision computing*, vol. 16, no. 5, pp. 295-306, 1998.
- [96] Y. Xu, X. Li, J. Yang, Z. Lai, and D. Zhang, "Integrating conventional and inverse representation for face recognition," *IEEE transactions on cybernetics*, vol. 44, no. 10, pp. 1738-1746, 2013.
- [97] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of 1994 IEEE workshop on applications of computer vision*, 1994, pp. 138-142: IEEE.