Research Article

Rupali Tajanpure* and Akkalakshmi Muddana

Circular convolution-based feature extraction algorithm for classification of high-dimensional datasets

Abstract: High-dimensional data analysis has become the most challenging task nowadays. Dimensionality reduction plays an important role here. It focuses on data features, which have proved their impact on accuracy, execution time, and space requirement. In this study, a dimensionality reduction method is proposed based on the convolution of input features. The experiments are carried out on minimal pre-processed nine benchmark datasets. Results show that the proposed method gives an average 38% feature reduction in the original dimensions. The algorithm accuracy is tested using the decision tree (DT), support vector machine (SVM), and K-nearest neighbor (KNN) classifiers and evaluated with the existing principal component analysis algorithm. The average increase in accuracy (Δ) is 8.06 for DT, 5.80 for SVM, and 18.80 for the KNN algorithm. The most significant characteristic feature of the proposed model is that it reduces attributes, leading to less computation time without loss in classifier accuracy.

Keywords: high-dimensional data, convolution, dimensionality reduction, feature extraction

1 Introduction

1.1 Overview

The data generated every day due to Internet-based applications are multidimensional. Data originate from various sources in different forms [46]. It is a very cumbersome task to analyze such extensive data. The accuracy and speed of data analysis depend on the features of the data involved. Features describe instances of data. Features can be constructed, discretized, transformed, or selected from massive data to classify data more accurately. There are three main kinds of features: categorical, ordinal, and quantitative. There is a need for feature transformation to improve the usefulness of features by changing, removing, or adding information. The feature transformation may turn the set of features into another set of features having fewer features. Selecting a subset of a given set of features speeds up learning and helps protect from overfitting [1].

1.2 Dimensionality reduction

Different features of data contain information about the target. If more features characterize data, then information carried by features is more, and hence data analytics should lead to better results. But this is

---

* Corresponding author: Rupali Tajanpure, GITAM University, Hyderabad, Telangana, India, e-mail: rupalidixit18@gmail.com
Akkalakshmi Muddana: GITAM University, Hyderabad, Telangana, India, e-mail: amuddana@gitam.edu

not always true. The presence of irrelevant or redundant features does not contribute to data classification, known as the curse of dimensionality [33,48]. Here the need for data reduction comes. Data reduction is nothing but removing irrelevant or redundant data before going for analysis. Data-reduction algorithms should reduce the size of data without loss of properties of data [7]. Different data-reduction techniques such as dimensionality reduction, data transformation, and numerosity reduction exist. Among these techniques, dimensionality-reduction methods such as principal component analysis (PCA) and discrete wavelet transforms are widespread [2,7,47].

The feature selection technique selects the crucial features of the original data based on some criteria. Feature selection tries to find the best subset of features among competing 2^n candidate subsets by applying some evaluation criteria. Since this feature selection process attempts to find only the best subset, it becomes exhaustive, leading to more cost. Some stopping criteria need to be there for other methods based on heuristic algorithms [28]. The necessary steps for the feature selection algorithm are as follows:

1. generate the next candidate feature subset,
2. evaluate the candidate feature subset using some evaluation function,
3. apply some stopping criteria, and
4. apply some validity check for validation of candidate feature subset [38,39].

In the literature, feature selection is categorized based on different strategies. It can be supervised, semi-supervised, or unsupervised based on the use of training data. According to the relationship between learning methods, feature selection is classified into the filter, wrapper, and hybrid models. Proper feature selection algorithms can improve learning accuracy and reduce the time required to analyze data [3,28]. The filter method uses the relationship between features and the class label. Techniques used for the filter method are information gain, chi-square, and reliefF. Wrapper methods work on the relevance of features and optimal feature subset selection. The embedded/hybrid approach uses a learning algorithm that selects features in the training process [4–6]. These algorithms work well, with different classifiers playing a dominant role in assessing feature reduction algorithms. Feature extraction transforms original data into features with strong pattern-recognition ability. PCA is one of the well-known feature extraction techniques. It generates new features using a linear combination of initial features [7]. Convolutional neural network (CNN) is used to reduce the size of an image without the loss of valuable features. With CNN transformation, the processing of images becomes easy. Here kernel filter is convolved with the input features to generate the feature map. The output features are local and short term. Authors in [51] proposed CNN-based gated recurrent unit model for human activity recognition.

1.3 Motivation

Feature selection, feature extraction, and feature optimization are the different forms of dimensionality reduction. Dimensionality reduction is today's research topic as it plays a vital role in analyzing data in various important fields like sentiment analysis [8]. A lot of research is done on sentiment analysis using machine learning algorithms and different word embedding techniques in a wide range of applications such as massive open online courses (MOOC) course assessment, product reviews, and probable question topic extraction [9–11]. Turkish and bibliometric data analysis is also done using sentiment analysis with feature selection techniques [12,13]. Humanoid robots have a wide application area in today's world. Feature selection plays a vital role in robotics too. The real challenge here is to get push recovery for the humanoid robot and non-linearity associated with the motion. To know the neurological disorders and problem identification of gait abnormality, human motion study is very important [49,51]. Human gait is unique for every person. The human walk is described with different joint trajectories [14–16]. This push recovery data classification is also done in the literature by using deep learning techniques. Features play an essential role in developing a computing module for push and gait recovery [17,18].

Feature reduction is also necessary for natural language processing. Machines do not have the capability like humans to understand the language with proper meaning. For natural language processing,

research is carried out with different dimensions by many authors for the sentiment analysis of the comments, reviews on social media or websites [19], context-based queries [20], teacher assessment reviews [21], and text classification using keyword extraction [22–24]. Thus, feature reduction has become an important part of widely applicable data analytics. One can apply feature reduction to directly impact classifier accuracy, computational time, and space requirements.

1.4 Research contribution

This study contributes toward a feature extraction method that very effectively uses the convolution technique for the first time in the literature for feature reduction.

The highlights of this research are as follows:

1. The proposed method shows a significant reduction in features.
2. We are modeling the convolution technique to reduce the features.
3. Performance analysis: Testing on different dimensional benchmark datasets is done with different classifiers. All combinations give better results than the existing method over accuracy and time.
4. Due to notable reduction in features, reduction in computation time and space is observed.

1.5 Organization of the article

The rest of the article is schematized as follows. The second section describes the literature survey in the dimensionality reduction domain. The section also describes the work done in different domains related to dimensionality reduction. The third section gives the proposed methodology. The detailed algorithms are discussed in this section. The fourth section discusses the experimental setup. The fifth section elaborates on the results obtained with different classifiers and their analysis. The last section presents the concluding remark and future directions of the research.

2 Literature survey

Heart disease is one of the major killers among other diseases, which has caused 7.4 million deaths worldwide in 2015 [25]. Other diseases like cancer, kidney disease, hepatitis are other significant killers worldwide. In the medical field, disease detection is challenging, as the features under consideration may be irrelevant and redundant, causing a reduction in classification accuracy. This irrelevant and redundant feature elimination is essential to reduce classification efforts, reduce the risk of overfitting, and improve classification accuracy. Reduction in the feature set is inversely proportional to classification accuracy and directly proportional to classification time. In the literature, preprocessing steps before the actual data classification include feature reduction, feature selection, feature extraction, and feature optimization. These are popular techniques used for dimensionality reduction [26].

Vipin Kumar et al. discussed different feature selection approaches, such as filters, wrappers, and embedded methods, and their application in the real world. The features could be highly dependent on one another, or there could be too many features. Hence, different feature reduction techniques are used frequently by researchers in different areas. The author described different application areas of feature selection in the real world, such as remote sensing, text categorization, intrusion detection, and image retrieval. Some of the challenges in feature selection, such as large dimensional data, scalability, and stability, are mentioned in this study [27]. A database of 50 subjects is created for human gait analysis

[49] by Vijay et al. The author represented data using 24 attributes per instance, which is preprocessed by using the Kalman filter. Different combinations of deep learning and hybrid deep learning classifiers are used to experiment, which shows a significant increase in accuracy (99.34%). The author also worked on a biped robot. The data regarding different walking styles are collected through inertial measurement unit sensors and walking patterns are analyzed [50].

Some of the nature-inspired algorithms for feature selection, such as the binary bat algorithm, particle swarm optimization, and modified cuckoo search algorithms, are discussed by Shrivastava et al. [28,29]. An integrated filter and wrapper method for a sequential search procedure improves classifier performance to tackle the overfitting problem and chances of getting a local optimal solution [30]. Xie and Wu proposed a feature selection algorithm based on association rules, which discovers the class attribute features as per the association analysis theory. Yet, this algorithm's time complexity is relatively high due to the Apriori algorithm [31]. Alessio Ferone proposed a novel approach of feature selection based on rough set theory [32]. Jinghua Liua et al. proposed a feature selection method based on the distinguishing ability of features. He used the maximum nearest neighbor concept to discriminate against the nearest neighbor of samples and evaluated the quality of features [33]. Tajanpure and Jena [38] put forth a multistage classifier system approach, where the feature decimation according to their level of processing need is done. The first-level features are processed first, and as per the first classifier's output, further level processing decision is taken. Dua et al. worked on human activity recognition with a proposed deep neural network-based CNN and gated recurrent unit. The proposed method performs features extraction and classification too [51].

Saul Solorio-Fernandez et al. proposed a new unsupervised spectral feature selection method. In many practical problems, the dataset under study is described through numerical and nonnumerical features, that is, mixed datasets. The spectral feature selection-based proposed method uses a kernel and a new spectrum-based feature evaluation measure for deciding the relevance of features. K-nearest neighbor (KNN), Naïve Bayes, and support vector machine (SVM) classifiers are used to find the proposed algorithm's performance, and their accuracy is compared [39]. Much research is done on the disease diagnosis system based on different approaches, such as learning vector quantization and artificial neural networks [40] and classification algorithms [41,42]. Many dimensionality reduction algorithms are developed by researchers for different applications to work with some advantages and some limitations like overfitting, higher time complexity, and so on. The common base to evaluate these techniques is classifier accuracy and other parameters like specificity, sensitivity, F-measure, and so on [43].

Wei Song proposed an effective content-based feature selection approach to improve clustering performance for genetic algorithms. The conventional genetic algorithms face the problem of slow learning and local minimum due to a high-dimensional exploration space. The proposed approach is a parametric and a nonparametric algorithm to adjust the genetic algorithm operators properly [44].

In the brute-force feature selection method, all possible combinations of the input features are evaluated to find the best subset. Here the computational cost is high, with the considerable danger of overfitting. An important aspect of feature selection techniques is the evaluation of a candidate feature subset and searching through the feature space. If there exist at least two instances with the same feature values but with different class labels, the feature subset is classified as inconsistent [45].

3 Proposed feature reduction system design

Convolution is a way of converting two sequences into another sequence. Each value in the input sequence is viewed as scaling and shifting of unit impulse or delta function in digital signal processing ($\delta(n)$). The output of convolution is simply expressed as shifted and scaled version of the input impulse. From this, it is very clear that the dimensions that are dominant in the input will result in output with the same dominance so the accuracy of the output is improved despite feature reduction [52]. This research proposes a feature extraction method based on convolution. Convolution is one of the basic operations of analog signals. There are two forms of convolution: linear convolution (LC) and circular convolution (CC). LC gives a linear

Table 1: Comparison between LC and CC

	LC	CC
Shifting of samples	Linear shifting	Circular shifting
No. of samples in the convolution result for inputs $x_1(n)$ and $x_2(n)$	$L = \text{length}(x_1(n)) + \text{length}(x_2(n)) - 1$	$N = \text{nearest } 2^n \text{ of } [\max(\text{length}(x_1(n)), \text{length}(x_2(n)))]$

overlapping result of the two sequences, which is nothing but the system's output when triggered with input $x_1(n)$ having transformation function $x_2(n)$. LC relates an output sequence with a given input sequence and impulse response, as shown in equation (1). LC is computed over all relevant values of n, that is, from $-\infty$ to $+\infty$. In mathematical form, if k is any positive integer, convolution is expressed [35] as

$$Y(n) = x_1(n) * x_2(n) = \sum_{n=-\infty}^{\infty} x_1(k)x_2(n-k), \tag{1}$$

whereas circular convolution gives the output when two sequences are circularly overlapped. Table 1 shows the comparison between LC and CC.

One can get the output of CC same as LC [34] if we follow

$$N = L = \text{length}(x_1(n)) + \text{length}(x_2(n)) - 1. \tag{2}$$

Here we focus more on the CC since we can reduce more features with it, as shown in the second row of Table 1. From the table, one can reduce only one feature per LC application, whereas one can find the nearest 2^n value concerning the maximum size of the first or second sequence for CC [34]. For getting the same number of values in both the input sequences, that is, N, both sequences are zero-padded as per need.

CC is one of the important properties of discrete Fourier transform (DFT). Equation (3) represents the mathematical form. If L is number of values in $Y(L)$, that is, output sequence, n is number of values in $x_1(n)$ and $x_2(n)$, and N is the nearest 2^n of $[\max(\text{length}(x_1(n)), \text{length}(x_2(n)))]$, then CC is mathematically expressed [35] as

$$Y(L) = \sum_{n=0}^{N-1} x_1(n)x_2(L-n)_N \quad \text{for } L = 0, 1, 2, \dots, N-1. \tag{3}$$

The above convolution equation involves index $((L-n))_N$, known as CC. It is an important property of DFT. The multiplication of a DFT of two sequences corresponds to the CC of two sequences in the time domain [17].

As the proposed algorithm uses CC, we have different methods to find CC.

Consider $x_1(n)$ and $x_2(n)$ are the two sequences used to convolve linearly with the number of samples "m" and "n," respectively, to get the output sequence $y(l)$, containing "l" number of samples.

3.1 Method 1: CC by linear convolution equivalence (LCE) method

One can get the result of CC equivalent to that of LC by applying zero paddings to make the number of elements in the input sequence equal to $(m + n - 1)$. According to the principle of LC, after convolution, one value will be less in the output sequence than the addition of the number of values in the input sequence. It is the base of the proposed attribute reduction concept.

3.2 Method 2: CC by DFT/inverse discrete fourier transform (IDFT) method

To find the CC of two sequences of length "m" and "n," condition $m = n$ should meet, and the length of two sequences should be close to the upper 2^n value. To meet this condition, one has to apply a zero-padding technique to the input sequence and then find the CC of the input sequences.
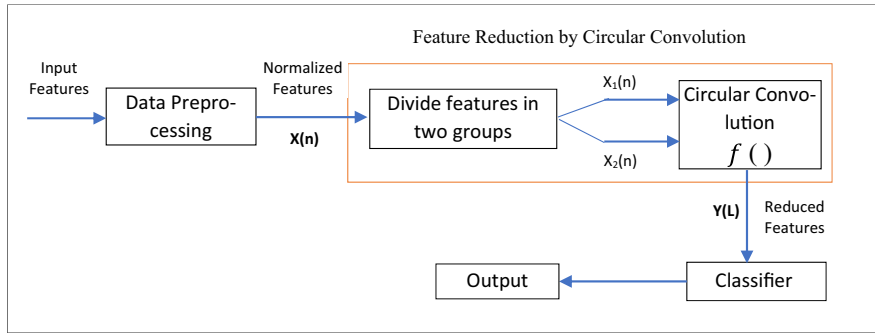
Figure 1: Proposed feature reduction system using FrbyCC.

The architecture of the proposed system based on CC is as follows:

As shown in Figure 1, the input dataset undergoes data-preprocessing techniques. Then normalized data is divided into two groups, with elements nearly equal to the closest 2^n in each group. These two sets of features act as input to CC to extract a reduced set of features at the output. Afterwards, this reduced set of features is tested with a classifier to judge the proposed feature reduction algorithm's performance.

According to the above-mentioned method 1 and method 2 for finding CC, there are two variants of the FrbyCC method.

3.2.1 LCE method (FrbyLCE)

In this method, zero padding is applied to $x_1(n)$ and $x_2(n)$ so that each sequence will contain L ($L = m + n - 1$) number of elements. With this method, at a time, one feature is reduced. Here to reduce the number of attributes, we have to apply FrbyLCE repeatedly on the input.

Algorithm 1. Feature reduction using the LC equivalence method.

　　Input: $F(n)$ – *set of input features.*
　　Output: $Y(L)$ – reduced set of features.
　　Begin
　　　　1. Divide the number of input features $F(n)$ in two parts $x_1(n_1)$ and $x_2(n_2)$ ... $n_1 \approx n_2$
　　　　2. Find the number of elements in output, that is,
　　　　　　$L = n_1 + n_2 - 1.$
　　　　3. Perform zero padding
　　　　　　$x_1(L) = [x_1(n_1), \text{zeropadding} (1, L - n_1)]$
　　　　　　$x_2(L) = [x_2(n_2), \text{zeropadding} (1, L - n_2)]$
　　　　4. Find CC of $x_1(L)$ and $x_2(L)$,
　　　　　　　　$X_1(K) = \text{DFT}(x_1(L)),$
　　　　　　　　$X_2(K) = \text{DFT}(x_2(L)),$
　　　　　　　　$Y(K) = X_1(K) * X_2(K),$ and
　　　　　　　　$Y(L) = \text{IDFT} (Y(K))$
　　End

3.2.2 DFT IDFT circular convolution method (FrbyCC)

Here, depending upon the number of points/samples, that is, N expected in DFT, zero padding is applied. The number of points in DFT could be 2, 4, 8, 16, 32, 64,…

Algorithm 2. Feature reduction using the DFT and IDFT CC method.

    Input: $F(n)$ – *set of input features*.

    Output: $Y(L)$ – *reduced set of features*.

    Begin

        1. Divide the number of input features $F(n)$ in two parts $x_1(n_1)$ and $x_2(n_2)$ … $n_1 \approx n_2$

        2. Select number of elements in output, that is, N

            N = nearest higher DFT point in the range of 2^n, which is 2, 4, 8, 16, 32, 64,…

        3. Perform zero padding

            $x_1(n) = [x_1(n_1), \text{zeropadding } (1, N - n_1)]$ and

            $x_2(n) = [x_2(n_2), \text{zeropadding } (1, N - n_2)]$.

        4. Find CC of $x_1(n)$ and $x_2(n)$,

                $X_1(K) = \text{DFT}(x_1(n))$,

                $X_2(K) = \text{DFT}(x_2(n))$,

                $Y(K) = X_1(K) * X_2(K)$, and

                $Y(L) = \text{IDFT } (Y(K))$

    End

4 Experimental setup

The experiment is carried out on Intel core i5-7200U CPU @ 2.50 GHz with 8 GB RAM and 64-bit Windows 10 Operating System laptop. The feature reduction and classification algorithms are implemented in Matlab 2015b.

4.1 Dataset description and preprocessing

From the UC irvine machine learning repository, nine datasets are used to evaluate the FrbyCC algorithm [36]. Out of these nine datasets, a good combination of less than 50 features, around 200 features, and more than 700 features dataset is selected. The datasets are preprocessed first. The missing values are replaced by the average of the column values and normalization is applied. FrbyLCE is assessed on the Cleveland heart disease dataset, whereas large attribute datasets evaluate FrbyCC. The purpose of selecting a large attribute dataset is to dominantly show the reduction in features. The output of FrbyCC is given to different classifiers for evaluation. Here tenfold cross-validation is applied to evaluate the classifier performance. Table 2 shows the basic information of the datasets.

5 Experimental evaluation

The proposed algorithm is applied to the dataset for feature reduction. The percent feature reduction is calculated to get the effectiveness of the proposed algorithm. Equation (4) finds percent feature reduction, which is the ratio of dimensions reduced after application of proposed algorithm to the original dimensions present in the dataset.

$$\text{PFR}_{\text{FrbyCC}} = \frac{f_{\text{Original}} - f_{\text{FrbyCC}}}{f_{\text{Original}}}. \tag{4}$$

Table 2: Datasets under consideration

Dataset	Feature reduction algorithm	No. of tuples	No. of features
Parkinson's disease	FrbyCC	756	754
Arrhythmia	FrbyCC	452	280
Internet ads	FrbyCC	3,279	782
QSAR androgen receptor data set	FrbyCC	1,687	1,024
Self-care activities dataset (SCADI)	FrbyCC	70	206
Heart disease	FrbyLCE	303	13
Kidney	FrbyCC	402	25
Hepatitis	FrbyCC	132	19
Breast cancer	FrbyCC	699	9

The reduced dataset is classified using SVM, K-nearest neighbor, and decision tree (DT) classifiers. A summary of an effect of the proposed feature reduction algorithm on different datasets is given in Table 3.

Table 3 shows that the FrbyCC algorithm has >30% reduction for large attribute datasets. Feature reduction also reduces the storage space and execution time of an algorithm. It is always useful to process a dataset containing a large volume of attributes. Metrics, such as accuracy, specificity, sensitivity, F-measure, and so on, are used to assess imbalanced datasets [37]. The evaluation of the proposed system is done by using the most frequently used accuracy metric on different datasets.

FrbyLCE method reduces one attribute per execution of the routine. One has to apply the FrbyLCE five times with the output of the previous stage applied as input to the next stage for reduction by five features. Figure 2 shows a two-stage FrbyLCE in a cascade fashion.

After the desired reduction in attributes, a classifier evaluates the results. Here two classifiers such as DT and SVM are used to evaluate the proposed method. Table 4 shows the results of the LCE–FrbyLCE method with DT classifier.

Comparing Tables 4 and 5, it is clear that the FrbyLCE method's use shows improvement in the accuracy of the DT classifier. The LCE method combines the two sets of input attributes to produce a new set of attributes, making DT classifier to classify more correctly on it. Hence, as we reduce features one by one, the DT classifier's accuracy increases, and from stage III, accuracy remains constant for further reduction in features as observed from Table 4. Also, it is clear in Table 5 that the SVM classifier cannot select decision

Table 3: Datasets under consideration and impact of FrbyCC method

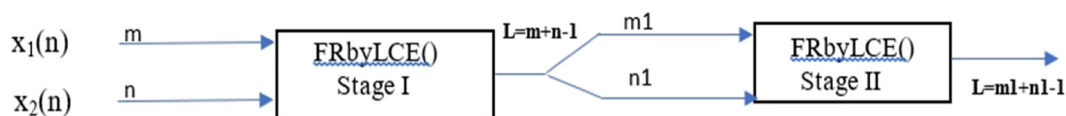| Dataset under consideration	No. of tuples	No. of features	Extracted features with FrbyCC	Percent feature reduction
Parkinson's disease	756	754	512	32.10
Arrhythmia	452	280	128	54.29
Internet ads	3,279	782	512	34.53
QSAR	1,687	1,025	512	50.05
SCADI	70	206	64	68.93
Kidney	402	25	16	36
Heart dataset	303	13	8	38.46
Hepatitis	132	19	16	15.78
Breast cancer	699	9	8	11.11

Figure 2: Feature reduction in a two-stage FrbyLCE method.

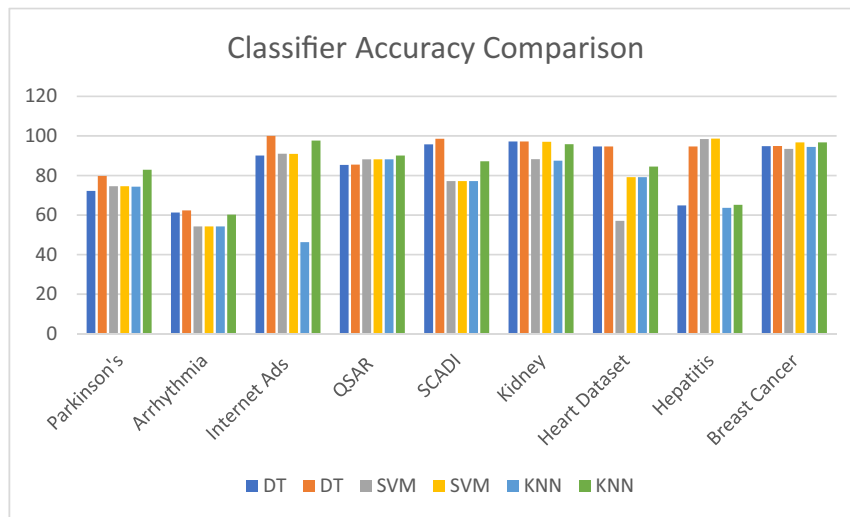**Table 4:** Results of FrbyLCE method with DT classifier on Cleveland dataset

Algorithm DT	No. of features for evaluation	Accuracy (%)	Sensitivity	Specificity	Precision	Recall	F-measure	Confusion matrix	
Without FrbyLCE	13	86.13	85.92	86.31	83.45	85.92	84.67	145	19
								23	116
FrbyLCE reduction by one feature	12	89.769	91.538	88.439	85.612	91.538	88.476	153	11
								20	119
FrbyLCE reduction by two features	11	90.099	93.6	87.64	84.173	93.6	88.636	156	8
								22	117
FrbyLCE reduction by three features	10	91.419	90.647	92.073	90.647	90.647	90.647	151	13
								13	126
FrbyLCE reduction by four features	9	90.429	88.194	92.453	91.367	88.194	89.753	147	17
								12	127
FrbyLCE reduction by five features	8	90.429	92.308	89.017	86.331	92.308	89.219	154	10
								19	120

Table 5: Results of FrbyLCE method with SVM classifier on Cleveland dataset

Algorithm SVM	No. of features for evaluation	Accuracy (%)	Sensitivity	Specificity	Precision	Recall	F-measure	Confusion matrix	
Without FrbyLCE	13	93.72	93.47	93.93	92.80	93.47	93.14	155	9
								10	129
FrbyLCE reduction by one feature	12	93.06	94.03	92.308	90.647	94.03	92.308	156	8
								13	126
FrbyLCE reduction by two features	11	88.45	91.935	86.034	82.014	91.935	86.692	154	10
								25	114
FrbyLCE reduction by three features	10	73.59	86.42	68.919	50.36	86.42	63.636	153	11
								69	70
FrbyLCE reduction by four features	9	79.21	75.676	82.581	80.576	75.676	78.049	128	36
								27	112
FrbyLCE reduction by five features	8	78.54	75.342	81.529	79.137	75.342	77.19	128	36
								29	110

Table 6: Accuracy of FrbyCC and PCA method with DT, KNN, and SVM classifier

Sr. No.	Dataset	Original features	Extracted attributes with FrbyCC	DT			SVM			KNN		
				PCA	FrbyCC	Δ_{DT}	PCA	FrbyCC	Δ_{SVM}	PCA	FrbyCC	Δ_{KNN}
1	Parkinson's disease	754	512	72.2	79.76	10.47	74.6	74.6	0.00	74.34	82.9	11.51
2	Arrhythmia	280	128	61.28	62.39	1.81	54.2	54.2	0.00	54.2	60.18	11.03
3	Internet ads	782	512	90.11	100	10.98	91	90.88	−0.13	46.35	97.62	110.61
4	QSAR	1,025	512	85.36	85.54	0.21	88.14	88.14	0.00	88.2	90.1	2.15
5	SCADI	206	64	95.71	98.57	2.99	77.14	77.14	0.00	77.14	87.14	12.96
6	Kidney	25	16	97.25	97.25	0.00	88.25	97	9.92	87.5	95.75	9.43
7	Heart dataset	13	8	94.71	94.73	0.02	57.1	79.21	38.72	79.21	84.48	6.65
8	Hepatitis	19	16	64.88	94.69	45.95	98.48	98.69	0.21	63.64	65.15	2.37
9	Breast cancer	9	8	94.85	94.92	0.07	93.42	96.71	3.52	94.42	96.71	2.43
Average Δ						8.06			5.80			18.80

Figure 3: Comparison of accuracy of FrbyCC and PCA with DT KNN SVM classifier.

Table 7: Execution time comparison of FrbyCC and PCA method with DT, KNN, and SVM classifier (time measured in seconds)

Name of dataset	FrbyCC			PCA		
	DT time	KNN time	SVM time	DT time	KNN time	SVM time
Parkinson's disease	5.84	2.04	3.42	9.57	2.23	3.62
Arrhryathmia	2.76	1.39	2.17	3.8	1.44	2.37
Internet ads	2.68	12.69	25.21	7.62	16.71	45.79
QSAR	8.83	5.25	13.22	9.35	8.03	24.69
SCADI	0.23	1.54	1.79	0.26	1.61	1.94

boundaries as features generated after FrbyLCE are a combination of two sets of input attributes. Here the accuracy decreases as we reduce the number of attributes.

Now, Table 6 shows FrbyCC accuracy evaluated with DT, KNN, and SVM classifiers. The results are taken on nine benchmark datasets from the UCI repository. Accuracy is taken as the performance measure for evaluation. The results are compared with the existing PCA feature reduction algorithm. If the accuracy
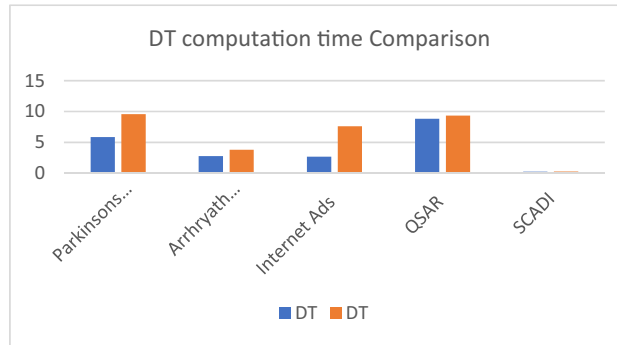
Figure 4: Comparison of the execution time of FrbyCC and PCA with DT classifier.
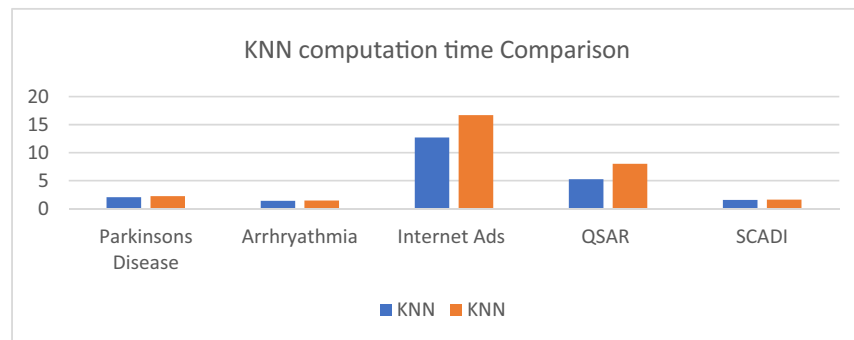

Figure 5: Comparison of the execution time of FrbyCC and PCA with KNN classifier.
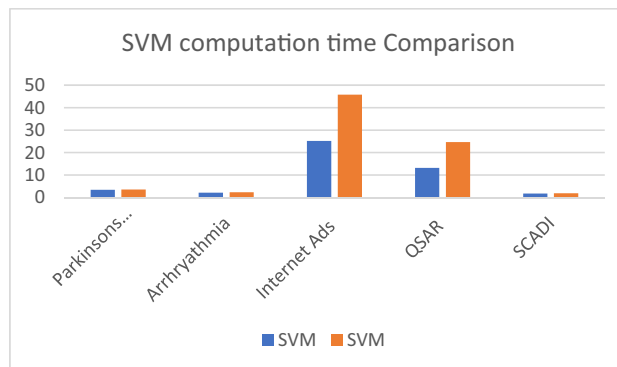

Figure 6: Comparison of the execution time of FrbyCC and PCA with SVM classifier.

of a system is represented by A, the accuracy improvement with the proposed FrbyCC method is represented by Δ [53] and calculated as shown in equation (5),

$$\Delta_{\text{DT}} = \frac{A_{\text{FrbyCC}} - A_{\text{PCA}}}{A_{\text{PCA}}} \times 100 \tag{5}$$

The assessment of Table 6 leads to the fact that classification performance is improved by an average 8% with DT classifier and 5% for SVM classifier. The 18% average accuracy improvement is shown with KNN classifier. The accuracy improvement achieved is purely due to feature extraction done by FrbyCC algorithm. FrbyCC multiplies and adds the results of multiplication; that is, it maintains the importance of features in extracted features too. The features extracted are convolved output of the input features with

important features influencing the extracted features. Despite of feature reduction, the impact of important features remains in the extracted output. Also, the difference in the improvement of FrbyCC with different classifiers is observed due to the impact of individual classifier behavior on the respective dataset. This proves that the FrbyCC method was very effective in accuracy terms. FrbyCC reduces more features in a single application on the dataset under consideration.

Table 6 shows the results with tenfold cross-validation. Observations show no drop in the accuracy of the classifier, despite feature reduction. In comparison with PCA, the FrbyCC method gives improved accuracy.

When we compare the FrbyLCE with FrbyCC, FrbyCC seems to be the most effective method. It reduces a significant number of features in one application, whereas FrbyLCE is to be applied in cascade fashion n times to reduce the "n" number of features. To verify accuracy in different datasets, FrbyCC is applied to different datasets, such as Parkinson's dataset, Arrhythmia, Internet ads, QSAR, and SCADI dataset, and compared with the PCA algorithm. As seen in Table 3, the first five datasets are high dimensional. We get a reduction in features up to 60% with the FrbyCC algorithm. The rest of the datasets have fewer dimensions as compared to the first five. Our focus is on high-dimensional datasets, in which the reduction in features leads to a decrease in execution time and storage space. The best reduction is observed in features if half of the total number of features is near any 2^n.

The following graphs show each dataset's accuracy with DT, KNN, and SVM, which implies that all the classifiers work differently on each dataset with improved accuracy.

It is observed from Figure 3 that DT and KNN work well, showing increased accuracy with the FrbyCC method compared to PCA. SVM gives accuracy in line with the PCA algorithm. Since PCA is also a feature extraction technique based on a linear combination of input attributes like FrbyCC, both algorithms work with similar accuracy for SVM.

In Matlab, time is measured using function tic and toc, which gives execution time in seconds. It is clear from Table 7 that the execution time of the FrbyCC algorithm is less than that of the PCA algorithm as FrbyCC is implemented using a DFT algorithm. Figures 4–6 show the graphical execution time comparison of FrbyCC and PCA feature reduction algorithm for different classifiers.

FrbyCC is also tested with Naïve Bayes classifier, which is a probabilistic classifier. Naïve Bayes purely works on the independence assumption between the input features. Hence Naïve Bayes algorithm shows poor accuracy on FrbyCC results. Here FrbyCC algorithm extracts the reduced features by overlapping of original features. Hence Naïve Bayes does not classify with good accuracy.

6 Conclusion

Feature reduction and accuracy are crucial concerns in data classification. High data dimensionality is the main issue in data classification. The problem of high-dimensional data handling is addressed in this research. We proposed a feature extraction method based on convolution to reduce features without loss of classification accuracy. In the first technique described in this article, feature reduction by LCE reduces one feature per application. It works well with the DT algorithm. The second method, FrbyCC, proves very effective in dimensionality reduction. Experiements show that it works well with DT and KNN. For SVM, accuracy is seen inline with the PCA for most of the datasets. It is observed that the feature reduction in each of the dataset depends on its proximity with 2^n value. The average increase in accuray (Δ) achieved with DT, SVM, and KNN is 8.06, 5.80, and 18.80, respectively, on different benchmark datasets. The proposed algorithm reduces execution time with the use of the DFT/IDFT [52]. Overall in FrbyCC algorithm, with feature reduction, we relish the accuracy with less storage space and execution time.

Conflict of interest: The authors do not have any type of conflict of interest.

Future research direction: The future scope of this research is to explore and evaluate the proposed algorithm using deep learning and hybrid deep learning techniques.

References

[1] Flach P. Index. Machine learning: the art and science of algorithms that make sense of data. Cambridge: Cambridge University Press; 2012.

[2] Han J, Kamber M. Data mining: concepts and techniques. 3rd edn. Waltham: Morgan Kaufmann Publishers; 2006.

[3] Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: a new perspective. Neurocomputing. 2018 July;300:70–9.

[4] Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. Bioinformatics. 2007 Oct 1; 23(19):2507–17. doi: 10.1093/bioinformatics/btm344.

[5] Lu Q, Li X, Dong Y. Structure preserving unsupervised feature selection. Neurocomputing. 2018;301:36–45.

[6] Jain D, Singh V. Feature selection and classification systems for chronic disease prediction: a review. Egypt Inform J. 2018;19:179–89.

[7] Keerthi Vasan K, Surendiran B. Dimensionality reduction using principal component analysis for network intrusion detection. Perspect Sci. 2016;8:510–2.

[8] Onan A, Korukoglu S. A feature selection model based on genetic rank aggregation for text sentiment classification. J Inf Sci. 2015;43(1):25–38. doi: 10.1177/0165551515613226.

[9] Onan A. Sentiment analysis on massive open online course evaluations: a text mining and deep learning approach. Comput Appl Eng Educ. 2020;29:572–89. doi: 10.1002/cae.22253.

[10] Onan A. Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. Concurr Comput Pract Exp. 2020 June 29. doi: 10.1002/cpe.5909.

[11] Onan A, Tocoglu MA. Weighted word embeddings and clustering-based identification of question topics in MOOC discussion forum posts. Comput Appl Eng Educ. 2020;29:675–89. doi: 10.1002/cae.22252.

[12] Onan A. Sentiment analysis in Turkish based on weighted word embeddings. 2020 28th Signal Processing and Communications Applications Conference (SIU), Gaziantep, Turkey: IEEE; 2020. p. 1–4. doi: 10.1109/SIU49456.2020.9302182.

[13] Onan A. Two-stage topic extraction model for bibliometric data analysis based on word embeddings and clustering. IEEE Access. 2019;7:145614–33. doi: 10.1109/ACCESS.2019.2945911.

[14] Semwal VB, Singha J, Sharma P, Chauhan A, Behera B. An optimized feature selection technique based on incremental feature analysis for bio-metric gait data classification. Multimed Tools Appl. 2017;76:24457–75. doi: 10.1007/s11042-016-4110-y.

[15] Gupta A, Semwal VB. Multiple task human gait analysis and identification: ensemble learning approach. In: Mohanty SN, editor. Emotion and information processing. Cham: Springer; 2020. doi: 10.1007/978-3-030-48849-9_12.

[16] Raj M, Semwal VB, Nandi GC. Bidirectional association of joint angle trajectories for humanoid locomotion: the restricted Boltzmann machine approach. Neural Comput Appl. 2018;30:1747–55. doi: 10.1007/s00521-016-2744-3.

[17] Semwal VB, Mondal K, Nandi GC. Robust and accurate feature selection for humanoid push recovery and classification: deep learning approach. Neural Comput Appl. 2017;28:565–74. doi: 10.1007/s00521-015-2089-3.

[18] Semwal VB, Gaud N, Nandi GC. Human gait state prediction using cellular automata and classification using ELM. In: Tanveer M, Pachori R, editors. Machine intelligence and signal analysis. Advances in intelligent systems and computing. Vol. 748. Singapore: Springer; 2019. doi: 10.1007/978-981-13-0923-6_12.

[19] Onan A, Toçoğlu MA. A term weighted neural language model and stacked bidirectional LSTM based framework for sarcasm identification. IEEE Access. 2021;9:7701–22. doi: 10.1109/ACCESS.2021.3049734.

[20] Singh U, Kedas S, Prasanth S, Kumar A, Semwal VB, Tikkiwal VA. Design of a recurrent neural network model for machine reading comprehension. Proc Comput Sci. 2020;167:1791–800. doi: 10.1016/j.procs.2020.03.388. ISSN 1877-0509.

[21] Onan A. Mining opinions from instructor evaluation reviews: a deep learning approach. Comput Appl Eng Educ. 2020;28:117–38.

[22] Onan A, Korukoglu S, Bulut H. Ensemble of keyword extraction methods and classifiers in text classification. Expert Syst Appl. 2016;57:232–47. doi: 10.1016/j.eswa.2016.03.045. ISSN 0957-4174.

[23] Onan A, Korukoğlu S, Bulut H. A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification. Inf Process Manag. 2017;53(4):814–33. doi: 10.1016/j.ipm.2017.02.008. ISSN 0306-4573.

[24] Kontonatsios G, Spencer S, Matthew P, Korkontzelos I. Using a neural network-based feature extraction method to facilitate citation screening for systematic reviews. Expert Syst Appl. 2020;6:100030.

[25] Key facts about heart disease, the World Health Organization (WHO). Cardiovascular disease; June 2017 [Online]. Available: http://www.who.int/mediacentre/factsheets/fs317/en/.

[26] Vivekanandan T, Ch Sriman Narayana Iyengar N. Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease. Comput Biol Med. 2017;90:125–36.

[27] Kumar V. Feature selection: a literature review. Smart Comput Rev. 2014;4:211–29. doi: 10.6029/smartcr.2014.03.007.

[28] Shrivastava P, Shukla A, Vepakomma P, Bhansali N, Verma K. A survey of nature-inspired algorithms for feature selection to identify Parkinson's disease. Comput Methods Prog Biomed. 2017;139:171–9.

[29] Sudarson J, Balasaheb T. Improved artificial neural network (ANN) with aid of artificial bee colony (ABC) for medical data classification. Int J Bus Intell Data Min. 2017;1:1. doi: 10.1504/IJBIDM.2017.10010713.

[30] Peng Y, Wu Z, Jiang J. A novel feature selection approach for biomedical data classification. J Biomed Inform. 2010;43:15–23.

[31] Xie J, Wu J. Feature selection algorithm based on association rule mining method. Eighth IEEE/ACIS ICCIS; 2009.

[32] Ferone A. Feature selection based on the composition of rough sets induced by feature granulation. Int J Approx Reason. 2018;101:276–92.

[33] Liua J, Lin Y. Feature selection based on the quality of information. Neurocomputing. 2017;225:11–22.

[34] Oppenheim AV, Schafer RW. Digital signal processing. 1st edn. The University of Michigan, Pearson; Jan 12 1975.

[35] Proakis JG, Manolakis DK. Digital signal processing: principles, algorithms, and applications. 3rd edn. South Asia: Pearson Publications; 1996.

[36] Dua D, Graff C. UCI Machine learning repository. Irvine, CA: The University of California, School of Information and Computer Science; 2019. http://archive.ics.uci.edu/ml.

[37] He H, Garcia EA. Learning from imbalanced data. IEEE Trans Knowl Data Eng. 2009;21:9.

[38] Tajanpure RR, Jena S. Diagnosis of disease using feature decimation with multiple classifier system. In: Dash S, Das S, Panigrahi B, editors. International Conference on Intelligent Computing and Applications. Advances in Intelligent Systems and Computing, 632. Singapore: Springer; 2018.

[39] Solorio-Fernández S, Martínez-Trinidad J, Ariel Carrasco-Ochoa J. A new unsupervised spectral feature selection method for mixed data: a filter approach. Pattern Recognit. 2017;72:314–26.

[40] Alkım E, Gürbüz E, Kılıç E. A fast and adaptive automated disease diagnosis method with an innovative neural network model. Neural Netw. 2012;33:88–96.

[41] Jain D, Singh V. Feature selection and classification systems for chronic disease prediction: a review. Egypt Inf J. 2018;19:179–89.

[42] Cura T. Use of support vector machines with a parallel local search algorithm for data classification and feature selection. Expert Syst Appl. 2020;145:113133.

[43] Yasmin G, Das AK, Nayak J, Pelusi D, Ding W. Graph based feature selection investigating boundary region of rough set for language identification. Expert Syst Appl. 2020;158:113575.

[44] Song W, Wang ST, Li CH. Parametric and nonparametric evolutionary computing with a content-based feature selection approach for parallel categorization. Expert Syst Appl. 2009;36:11934–43.

[45] Dash M, Liu H. Consistency-based search in feature selection. Artif Intell. 2003;151:155–76.

[46] Weitschek E, Felici G, Bertolazzi P. Clinical data mining: problems, pitfalls and solutions. 2013 24th International Workshop on Database and Expert Systems Applications, 1529-4188/13 $2600@2013. IEEE; 2013. doi: 10.1109/DEXA.2013.42.

[47] Dash M, Liub H. Consistency-based search in feature selection. Artif Intell. 2003;151:155–76.

[48] Ferone A. Feature selection based on composition of rough sets induced by feature granulation. Int J Approx Reason. 2018;101:276–92.

[49] Vijay Bhaskar S, Gupta A, Lalwani P. An optimized hybrid deep learning model using ensemble learning approach for human walking activities recognition. J Supercomput. 2021;103:1–24.

[50] Vijay Bhaskar S, Neha G, Praveen L, Vishwanath B, Abhay Kumar A. Pattern identification of different human joints for different human walking styles using inertial measurement unit (IMU) sensor. Artif Intell Rev. 2021;1–21.

[51] Dua N, Singh SN, Semwal VB. Multi-input CNN-GRU based human activity recognition using wearable sensors. Computing. 2021;103:1461–78.

[52] Smith SW. The scientist and engineer's guide to digital signal processing. San Diego, Calif: California Technical Publishing; 1997.

[53] Tarle B. Integrating multiple methods to enhance medical data classification. Evol Syst. 2020;11:133–42. doi: 10.1007/s12530-019-09272-x.