

Info_PCA: A Hybrid Technique to Improve Accuracy by Dimensionality Reduction

Surabhi Lingwal¹, Komal Kumar Bhatia², Manjeet Singh³

¹ Govind Ballabh Pant Institute of Engineering and Technology, Uttarakhand

^{2,3} YMCAUST, Faridabad, Haryana, India

Abstract: Principal Component Analysis and Shannon Entropy are some of the most widely used methods for feature extraction and selection. PCA reduces the data to a new subspace with low dimensions by calculating the eigenvectors from eigenvalues out of a covariance matrix and thereby reduces the features to a smaller number capturing the significant information. Shannon entropy is based on probability distribution to calculate the significant information content. Information gain shows the importance of a given attribute in the set of feature vectors. The paper has introduced a hybrid technique Info_PCA which captures the properties of Information gain and PCA that overall reduces the dimensionality and thereby increases the accuracy of the machine learning technique. It also demonstrates the individual implementation of Information gain for feature selection and PCA for dimensionality reduction on two different datasets collected from the UCI machine learning repository. One of the major aims is to determine the important attributes in a given set of training feature vectors to differentiate the classes. The paper has shown a comparative analysis on the classification accuracy obtained by the application of Information Gain, PCA and Info_PCA applied individually on the two different datasets for feature extraction followed by ANN classifier where the results of hybrid technique Info_PCA achieves maximum accuracy and minimum loss in comparison to other feature extraction techniques.

Keywords: Principal Component Analysis • Shannon Entropy • Information gain • Artificial Neural Network • Classification • Feature Vector

1. Introduction

Feature extraction and feature selection are considered an important preprocessing step for the classification and regression process. Feature selection[1] is the process of selecting a subset of the original set of features while feature extraction[2] is the process of retrieving important features from existing data. The process helps in converting the raw data into suitable input formats required to solve a problem using Machine learning techniques. Sometimes the presence of irrelevant features degrades the performance of an algorithm[3], on the other hand, a poor model with a good set of features results in better performance. Feature extraction works by

transforming, combining, and formatting basic features into new derived features while feature selection emphasizes discarding redundant or irrelevant features to prevent overfitting. Feature selection is one of the important data preprocessing steps to derive a minimal set of features from the raw data to obtain a high accuracy while maintaining the originality of the features[4]. The feature selection process can be carried through chi-Square and Shannon Entropy formula[5]. Erick Odhiambo Omuya et. al.[6] has applied PCA and Information Gain for feature selection used for classification. Kai Zeng et. al.[7] has applied the Shannon entropy to select significant features from high dimensional data to produce strong classification output. Shifei Ding et. al[2] explains the linear and non-linear feature extraction. Pasi Luukka[8] perform feature selection using fuzzy entropy together with a similarity classifier to reduce the noise and enhance the classification accuracy. The model is tested on four medical datasets where it gave better classification results on a reduced set of data than the original dataset. Iman Khosravi et. al[9] performs feature selection to improve the accuracy of cropland classification. It works on maximum separability and minimum dependency for the feature selection process. Feng Jiang et. al[10] proposed a relative decision entropy feature selection algorithm based on Shannon entropy for rough sets. Dimitrios Effrosynidis et. al[11] has made a comparison of various feature selection techniques employed for environmental data. Due to a large number of features available in a dataset affecting its classification and prediction, different machine learning techniques have been employed. Information gain are based on rough set theory[12,13] while feature extraction technique like PCA are based on transformation based technique[14]. A.M. Jade et. al[14] has adopted Kernel PCA to extract significant non-linear features by transforming the data to a high dimensional feature space. Ji Ma et. al[15] has applied PCA for feature extraction from image data for deep learning. Rough set techniques are better in comparison to maintaining the originality of the feature set on the other hand PCA compressed the individual characteristics of the feature set. Farzana Anowar et. al[16] has discussed several feature extraction techniques which could be adopted to avoid the curse of dimensionality in different machine learning techniques.

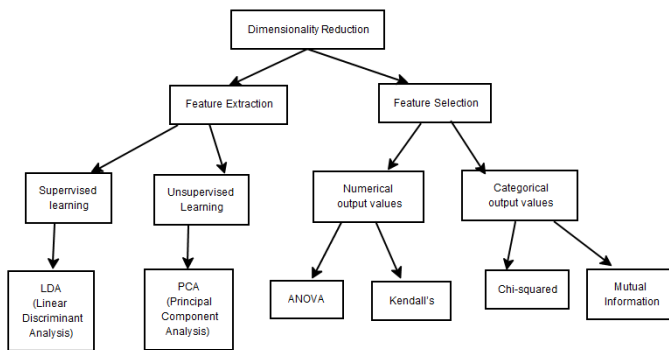


Fig. 1: Dimensionality reduction techniques

Majorly, two methods are used for feature selection in supervised learning for large datasets such as wrapper and filter methods[11]. The paper has demonstrated the application of chi-square and mutual information for feature selection. This is followed by the process of feature extraction[11] via the technique as PCA meant for statistical data. PCA is a dimensionality reduction technique required for different fields such as artificial intelligence, machine learning, pattern recognition, and data mining. The PCA tends to preserve the original data information while presenting the high dimensional data to low dimensional space with the linear transformation. Like in statistical datasets there are a lot of attributes that are considered where some are contributing and some are irrelevant to the dataset. Such main features are needed to be identified which are later replaced with a linear combination of original features[17]. PCA reduces the dimensions of the dataset by storing the features with larger information and variance and rejecting the attributes with less variance and information. PCA is used for the optimization process. PCA deals with high dimensional data to save from the curse of dimensionality. Ji Ma et.al[15] focuses on the reduction of dimension to extract image features using PCA where the images are further classified using CNN. The dataset of our work is composed of categorical input and categorical output. The paper focused on the process of feature selection followed by feature extraction to reduce the dimensionality of the dataset for better classification by machine learning techniques. The process of dimensionality reduction is classified as shown in Fig. 1 where the technique for feature selection can be opted based upon the data chosen either numerical or categorical.

In this paper, we have applied a hybrid technique named Info_PCA to improve the classification accuracy deployed on two different standard datasets collected from UCI machine learning repository[18] namely iris and leaf data set[19]. This both dataset belongs to the category of plants and vegetation related to the agriculture field. The iris dataset has three class labels classifying three different iris plants such as Setosa, Versicolor and, Virginia. The leaf dataset comprises the

collection of shape and texture features extracted from digital images of leaf specimens belonging to 30 different plant species. Initially, the data is classified with ANN classifier, the accuracy thus achieved is compared with the classifier preceded by feature selection and feature extraction process. The paper thus demonstrated that our approach Info_PCA followed by ANN classifier attained better results than the individual application of Information gain or PCA or simple ANN classifier. The rest of the paper is organized as follows: Section 2 covers the methodology. Section 3 describes the Experimentation and Section 4 deals with Result Analysis. Section 5 forms the Conclusion.

2. Methodology

The methodology of the proposed work is divided into four sub-sections where the purpose is to improve the classification accuracy. The data collected from the UCI machine learning repository undergoes data preprocessing as shown in Fig. 2. Section 2.1 deals with the direct implementation of Machine learning classifiers like Artificial Neural Network (ANN) on the dataset, section 2.2 deals with the process of feature selection followed by ANN, Section 2.3 deals with feature extraction followed by ANN, and Section 2.4 deals with hybrid classifier Info_PCA. The classification accuracy thus produced by the different classifiers under different sections is compared for best classification results.

2.1 Machine learning: ANN

Artificial Neural Network[20] is a widely used tool for classification in machine learning. It classifies the samples into discrete classes as a function of the input data. The model is a brain-inspired system which is intended to learn as human learns through the repetitive process. The datasets are initially trained with the Artificial Neural Network model. ANN is the architecture that consists of multiple fully connected layers with a different set of neurons. It is based on a connected input-output network associated with different weights related to different connections. It consists of one input layer, different hidden layers involving transfer functions, and one output layer. The network learns by adjusting the weights of the connections[21]. The work in this paper utilizes a feed-forward neural network where the connections between different layers do not form a cycle. The network passes the activation signal via the transfer function that will develop the non-linearity to produce the output where the neurons of the network are activated by the weighted sum of the input. These weights of the inter-connection are optimized to achieve the desired level of accuracy. ANN has been applied due to various advantages like deep and effective learning, less noise affected, etc.

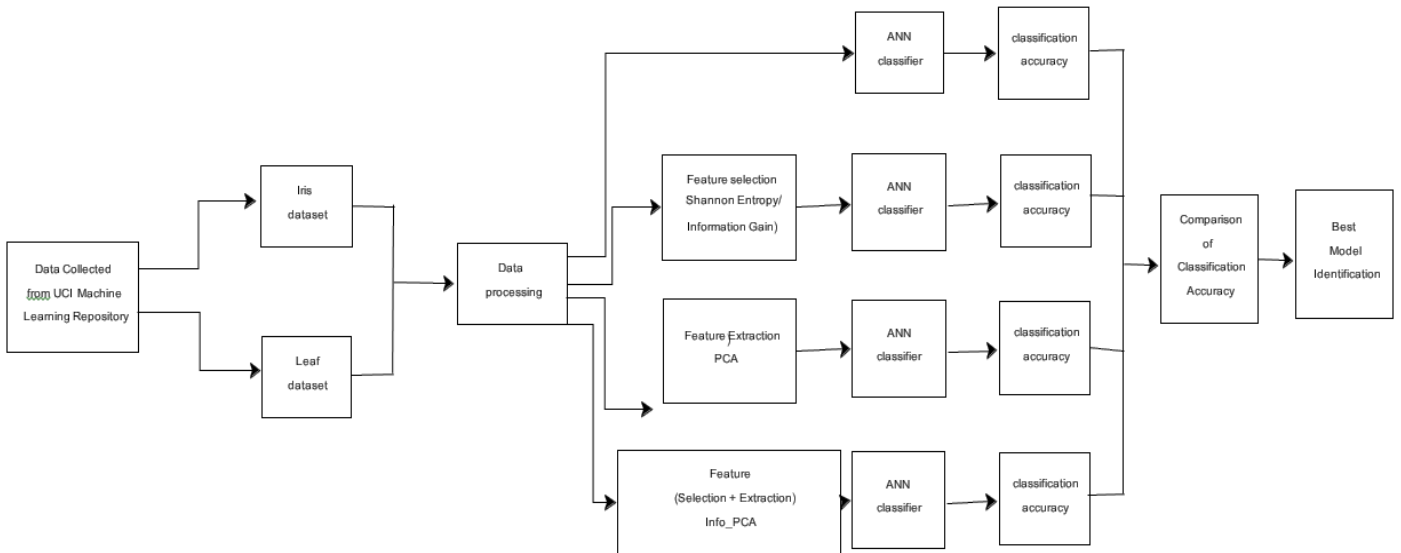


Fig.2: Procedure for dimensionality reduction process

2.2 Feature Selection

Feature selection is the process of selecting the most relevant features which are significantly impacting the dependent variable. The task is to select the best set of features to build a suitable model[22]. More the features are positively correlated less is the individual impact of the independent feature on the dependent variable. Feature selection techniques are broadly classified as supervised techniques and unsupervised techniques. Supervised techniques are applied for labeled data for feature selection in classification and regression models while Unsupervised techniques are used for unlabeled data. These techniques can be classified as Filter, wrapper, embedded, and hybrid methods[23]. Filter methods are generally preferred for high dimensional data as it is faster and computationally less expensive. It collects the intrinsic properties of the features through univariate statistics. Two widely used techniques of filter method for the categorical dataset are Chi-square and Information gain[24]. The Chi-square method is used to calculate the Chi-square value between each independent and dependent variable to find the suitable attributes with the best chi-square score. Shannon entropy-based information gain has been applied to the work to select the high impact independent features by calculating the information gain of each variable concerning the dependent variable. It is used to transform the data to calculate the reduction in entropy. On applying information gain to leaf dataset, the relevant features that are selected are Eccentricity, Aspect Ratio, Elongation, Solidity, Stochastic Convexity, Isoperimetric Factor, Maximal Indentation Depth, Average Intensity, Average Contrast, Smoothness and, Third moment while on applying information gain to iris dataset, the important features found are petal length, petal width and, sepal length.

2.3 Feature Extraction

Feature extraction is the process of extracting the relevant attributes of multiple features that can significantly affect the dependent variable[14]. It is needed to speed up the training and avoid overfitting of data. Feature extraction or thus dimensionality reduction technique helps in improving the accuracy and data visualization. It emphasizes reducing the feature set by creating a new set of features out of the original by extracting the relevant attributes of every feature. It summarizes the maximum of information contained in the original dataset and thus discards the original data. Two important feature extraction techniques that are widely popular are PCA and LDA[16,15]. LDA is a supervised dimensionality reduction technique used for multi-class classification[25]. It separates the samples of the training dataset by their class values that focus on maximizing the separation of samples between classes but minimizes the separation within each class.

The paper has deployed Principal Component Analysis for feature extraction process[15]. PCA is one of the widely used linear dimensionality reduction techniques that extract the important subfeatures from the different features and thus compresses them together into several principal components. It follows the pair wise distances to maximize the variance and minimize the reconstruction error. The data is projected to a set of orthogonal axes where each axis is scored based on its importance. It is an unsupervised learning technique that focuses on variance. The paper thus applied the PCA technique to the original dataset to derive the level of accuracy on the training and test data.

2.4 Info_PCA

The paper has developed a hybrid technique Info_PCA to improve the classification accuracy of the training and test dataset. The method has combined the capabilities of both the

techniques- information gain and Principal Component Analysis which improves the performance in terms of accuracy and loss. The algorithm and architecture of the proposed hybrid model are given below in Algorithm 1 and Fig 3. The level of accuracy derived on the original dataset with the simple

application of ANN model, application of information gain feature selection technique and application of PCA feature extraction technique was less as compared to the Info_PCA hybrid model as can be seen from Table 1 (a) and (b).

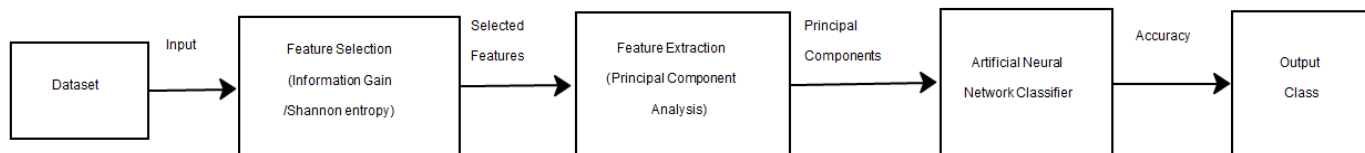


Fig. 3: Block diagram of Hybrid model: Info_PCA

Algorithm 1 procedure for Info_PCA model

1. Start.
2. Read the dataset.
3. Input the data to feature selection process through Information gain.
4. Pass the selected features to next phase.
5. Extract the compressed attributes from selected features through PCA.
6. Generate principal components.
7. Pass the principal components to ANN classifier.
 - (a) Input the principal components to the ANN model
 - (b) Set the learning rate.
 - (c) Set the number of hidden layers.
 - (d) Set the number of epochs.
 - (e) Set the bias and weights of the input layer.
 - (f) Calculate the weighted sum of the output layer.
 - (g) Calculate the error.
 - (h) Move to step 7(e).
 - (i) Back propagate followed by feed forward until the error is minimized.
8. Generate the result.
9. Calculate the accuracy.
10. End.

3. Experimentation

The experimentation is performed on two different datasets collected from the UCI machine learning repository which belongs to the category of plants and vegetation. These are the leaf dataset and iris dataset.

Leaf dataset: The dataset consists of leaf specimens of 40 different plant species including their shape and texture features. The dataset consists of in total 16 attributes namely Class (Species), Specimen Number, Eccentricity, Aspect Ratio, Elongation, Solidity, Stochastic Convexity, Isoperimetric Factor, Maximal Indentation Depth, Lobedness, Average Intensity, Average Contrast, Smoothness, Third moment, Uniformity and, Entropy. The dataset was created by Pedro F.B. Silva, Andre R.S. Marcal, Rubim M. Almeida da Silva. The

dataset is suitable for the classification of data containing real values.

Iris dataset: This is one of the most widely used pattern recognition and classification dataset. The dataset consists of 3 classes named as- Iris Setosa, Versicolour and, Virginica. Each class contains 50 instances. The dataset is characterized by four features namely petal length, petal width, sepal length and, sepal width. Each class represents a type of iris plant where one class is separable from the other. The data was created by R.A. Fisher.

The two datasets undergo four different models of training. The first model is the simple application of the ANN on the dataset for training. The second model consists of the process of feature selection followed by classification through ANN. The third model consists of the process of feature extraction through PCA followed by ANN classification. The fourth model is the application of Info_PCA for dimensionality reduction followed by classification through ANN. The accuracy and loss obtained on the two datasets during training are given in Tables 1 (a) and (b) for the four different models. Thus, it can be seen from the two tables that the accuracy and loss obtained during training under Info_PCA is better than the other three methods.

Table 1: accuracy and loss obtained during training of the two datasets

(a) leaf dataset

Techniques	Training Accuracy	Training loss
ANN	0.7721	0.6522
Feature Selection	0.7878	0.5095
Feature Extraction	0.7868	0.6410
Info_PCA	0.8088	0.4834

(b) iris dataset

Techniques	Training Accuracy	Training loss
ANN	0.8833	0.2875
Feature Selection	0.9250	0.2384
Feature Extraction	0.9333	0.2202
Info_PCA	0.9834	0.1356

4. Result Analysis

The four different models are applied to the iris and leaf test dataset. The results are analysed with respect to the

confusion matrix, accuracy score, mean absolute error, and classification report including precision, recall, and F1-score. The results of iris data which is trained for 50 epochs as it contains only three classes are given in Table 2 and Table 3. Table 2 illustrates the confusion matrix between the actual class and the predicted class of 30 test samples under the four different models as given in sub-table (a), (b), (c) and, (d). Table 3 (a),(b), (c) and, (d) explains the precision, recall, and F1-score value obtained under the three different classes of iris data, along with the overall accuracy score and mean absolute error produced on the test data under different models.

Table 2: Confusion Matrix of test samples of iris data using different models

(a) Confusion Matrix of test data using ANN

Predicted_class	actual_class		
	Setosa	versicolor	virginica
Setosa	11	0	0
versicolor	1	8	4
virginica	0	1	5

(b) Confusion Matrix of test data using feature selection

Predicted_class	actual_class		
	Setosa	versicolor	virginica
Setosa	11	0	0
versicolor	1	9	3
virginica	0	1	5

(c) Confusion Matrix of test data using feature extraction

Predicted_class	actual_class		
	Setosa	versicolor	virginica
Setosa	11	0	0
versicolor	1	10	2
virginica	1	0	5

(d) Confusion Matrix of test data using Info_PCA

Predicted_class	actual_class		
	Setosa	versicolor	virginica
Setosa	11	0	0
versicolor	0	13	0
virginica	0	0	6

The results of leaf data which is trained for 300 epochs due to its large number of classes are given in Table 4, Table 5, and Table 6. Table 4 (a) and (b) illustrates the confusion matrix between the actual class and the predicted class of 68 test samples under the ANN and feature selection model while Table 5(a) and (b) describes the confusion matrix between the actual class and predicted class of the test samples under the feature extraction and Info_PCA model. Table 6 (a),(b), (c) and, (d) gives the precision, recall, and F1-score measure of the different classes of leaf data, along with the overall accuracy

score and mean absolute error generated on the test data under different models.

Thus, it can be identified from the above matrices and accuracy measures under the iris and leaf dataset that our proposed approach Info_PCA performed better with improved accuracy and reduced loss as compared to the other models. Regardless of the number of data instances and several classes in the two datasets which are significantly different, our hybrid model Info_PCA performed remarkably well.

Table 3: Accuracy measure of test samples of iris data using different models

(a) Accuracy measures of test data using ANN

	Precision	Recall	F1-score	Accuracy_Score	Mean_Absolute_Error
Setosa	1.00	1.00	1.00	24	0.122
versicolor	0.89	0.62	0.73		
virginica	0.56	0.83	0.67		

(b) Accuracy measures of test data using feature selection

	Precision	Recall	F1-score	Accuracy_Score	Mean_Absolute_Error
Setosa	1.00	1.00	1.00	25	0.099
versicolor	0.90	0.69	0.78		
virginica	0.62	0.83	0.71		

Table 6: Accuracy measures of test samples of leaf data using different model

(a) Accuracy measures of test data using ANN

	Precision	Recall	F1-score	Accuracy_Score	Mean_Absolute_Error
1	0.33	0.50	0.40	33	0.03431373
2	0.50	0.50	0.50		
3	0.20	0.50	0.29		
4	0.00	0.00	0.00		
5	0.00	0.00	0.00		
6	0.75	0.60	0.67		
7	1.00	0.33	0.50		
8	1.00	1.00	1.00		
9	0.50	0.50	0.50		
10	1.00	0.25	0.40		
11	0.00	0.00	0.00		
12	0.80	1.00	0.89		
13	0.60	0.60	0.60		
14	1.00	0.50	0.67		
15	1.00	0.50	0.67		
16	1.00	0.75	0.86		
17	0.00	0.00	0.00		
18	0.67	0.50	0.57		
19	0.00	0.00	0.00		
20	0.00	0.00	0.00		
21	0.00	0.00	0.00		
22	0.00	0.00	0.00		
23	1.00	0.50	0.67		
24	1.00	0.50	0.67		
25	0.67	0.67	0.67		
26	0.50	0.25	0.33		
27	0.00	0.00	0.00		
28	0.50	0.50	0.50		
29	0.00	0.00	0.00		
30	0.00	0.00	0.00		

(b) Accuracy measures of test data using feature selection

	Precision	Recall	F1-score	Accuracy_Score	Mean_Absolute_Error
1	0.00	0.00	0.00	35	0.037450984
2	0.00	0.00	0.00		
3	0.50	0.50	0.50		
4	0.50	1.00	0.67		
5	0.50	1.00	0.67		
6	1.00	0.60	0.75		
7	0.80	0.67	0.73		
8	1.00	1.00	1.00		
9	0.67	1.00	0.80		
10	0.75	0.75	0.75		
11	0.00	0.00	0.00		
12	0.50	0.50	0.50		
13	0.71	1.00	0.83		
14	0.00	0.00	0.00		
15	1.00	0.50	0.67		
16	0.50	0.50	0.50		
17	0.00	0.00	0.00		
18	0.50	0.25	0.33		
19	0.00	0.00	0.00		
20	0.00	0.00	0.00		
21	0.25	1.00	0.40		
22	0.00	0.00	0.00		
23	1.00	1.00	1.00		
24	1.00	0.50	0.67		
25	0.67	0.67	0.67		
26	0.33	0.50	0.40		
27	1.00	1.00	1.00		
28	0.50	0.50	0.50		
29	0.50	1.00	0.67		
30	0.00	0.00	0.00		

(c) Accuracy measures of test data using feature extraction

	Precision	Recall	F1-score	Accuracy_Score	Mean_Absolute_Error
1	0.33	0.50	0.40	32	0.03529412
2	0.33	0.50	0.40		
3	0.50	0.50	0.50		
4	0.00	0.00	0.00		
5	0.50	1.00	0.67		
6	0.50	0.20	0.29		
7	1.00	0.33	0.50		
8	1.00	1.00	1.00		
9	0.00	0.00	0.00		
10	0.75	0.75	0.75		
11	0.00	0.00	0.00		
12	0.67	0.50	0.57		
13	1.00	0.40	0.57		
14	0.33	0.50	0.40		
15	1.00	0.50	0.67		
16	0.50	0.50	0.50		
17	0.00	0.00	0.00		
18	0.75	0.75	0.75		
19	0.00	0.00	0.00		
20	0.00	0.00	0.00		
21	0.14	1.00	0.25		
22	0.00	0.00	0.00		
23	1.00	1.00	1.00		
24	0.50	0.50	0.50		
25	0.50	0.33	0.40		
26	0.50	0.25	0.33		
27	0.50	1.00	0.67		
28	0.50	0.50	0.50		
29	0.00	0.00	0.00		
30	0.00	0.00	0.00		

(d) Accuracy measure of test data using Info_PCA

	Precision	Recall	F1-score	Accuracy_Score	Mean_Absolute_Error
1	0.67	1.00	0.80	42	0.025490198
2	0.00	0.00	0.00		
3	0.67	1.00	0.80		
4	0.00	0.00	0.00		
5	1.00	1.00	1.00		
6	1.00	0.80	0.89		
7	1.00	0.83	0.91		
8	1.00	1.00	1.00		
9	1.00	1.00	1.00		
10	1.00	0.50	0.67		
11	0.00	0.00	0.00		
12	0.67	0.50	0.57		
13	0.75	0.60	0.67		
14	0.50	0.50	0.50		
15	1.00	1.00	1.00		
16	0.50	0.50	0.50		
17	0.00	0.00	0.00		
18	0.50	0.50	0.50		
19	0.00	0.00	0.00		
20	0.00	0.00	0.00		
21	0.17	1.00	0.29		
22	0.00	0.00	0.00		
23	1.00	1.00	1.00		
24	0.50	0.50	0.50		
25	0.67	0.67	0.67		
26	0.50	0.50	0.50		
27	0.00	0.00	0.00		
28	0.50	0.50	0.50		
29	0.00	0.00	0.00		
30	0.00	0.00	0.00		

5. Conclusion

In this research, we proposed a hybrid model Info_PCA involving the capabilities of feature selection and feature extraction process followed by the classification of two different datasets using the ANN classification technique. The dataset is collected from the UCI Machine Learning repository. The hybrid model initially selects the features that possess the highest separability and lowest correlation among all features of the classes. The proposed hybrid model is compared with other models involving feature selection or extraction process. The experimental test performed on the dataset indicates that the Info_PCA method has higher accuracy and minimum mean absolute error as compared to other methods as can be seen from Table 3(d) and Table 6(d). Moreover, the Info_PCA correctly predicted 30 out of 30 test samples of the iris dataset while 42

out of 68 test samples of the leaf dataset are correctly predicted. On the other hand, the method also selects the optimal features. The information gain through Shannon entropy can find similar and different optimum features for different classes. Thus, our proposed model is found to be a better model with improved accuracy over the other models.

References

[1] Blum AL, Langley P. Selection of relevant features and examples in machine learning. Artificial intelligence. 1997;97(1-2):245–271.
 [2] Ding S, Zhu H, Jia W, Su C. A survey on feature extraction for pattern recognition. Artificial Intelligence Review. 2012;37(3):169–180.

- [3] Rostami M, Berahmand K, Nasiri E, Forouzande S. Review of swarm intelligencebased feature selection methods. *Engineering Applications of Artificial Intelligence*. 2021;100:104210.
- [4] Pedrycz W, Vukovich G. Feature analysis through information granulation and fuzzy sets. *Pattern Recognition*. 2002;35(4):825–834.
- [5] Shannon CE. A mathematical theory of communication. *The Bell system technical journal*. 1948;27(3):379–423.
- [6] Omuya EO, Okeyo GO, Kimwele MW. Feature Selection for Classification Using Principal Component Analysis and Information Gain. *Expert Systems with Applications*. 2021;p. 114765.
- [7] Zeng K, She K, Niu X. Feature selection with neighborhood entropy-based cooperative game theory. *Computational intelligence and neuroscience*. 2014;2014.
- [8] Luukka P. Feature selection using fuzzy entropy measures with similarity classifier. *Expert Systems with Applications*. 2011;38(4):4600–4607.
- [9] Khosravi I, Safari A, Homayouni S. MSMD: maximum separability and minimum dependency feature selection for cropland classification from optical and radar data. *International Journal of Remote Sensing*. 2018;39(8):2159–2176.
- [10] Jiang F, Sui Y, Zhou L. A relative decision entropy-based feature selection approach. *Pattern Recognition*. 2015;48(7):2151–2163.
- [11] Effrosynidis D, Arampatzis A. An evaluation of feature selection methods for environmental data. *Ecological Informatics*. 2021;61:101224.
- [12] Cekik R, Uysal AK. A novel filter feature selection method using rough set for short text data. *Expert Systems with Applications*. 2020;160:113691.
- [13] Dai J, Xu Q. Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification. *Applied Soft Computing*. 2013;13(1):211–221.
- [14] Jade A, Srikanth B, Jayaraman V, Kulkarni B, Jog J, Priya L. Feature extraction and denoising using kernel PCA. *Chemical Engineering Science*. 2003;58(19):4441–4448.
- [15] Ma J, Yuan Y. Dimension reduction of image deep feature using PCA. *Journal of Visual Communication and Image Representation*. 2019;63:102578.
- [16] Anowar F, Sadaoui S, Selim B. Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, tSNE). *Computer Science Review*. 2021;40:100378.
- [17] Viharos ZJ, Kis KB, Fodor A, Bu'ki' AM. Adaptive, Hybrid Feature Selection (AHFS). *Pattern Recognition*. 2021;p. 107932.
- [18] Dua D, Graff C. UCI Machine Learning Repository; 2017. Available from: <http://archive.ics.uci.edu/ml>.
- [19] Silva PF, Marcal AR, da Silva RMA. Evaluation of features for leaf discrimination. In: *International Conference Image Analysis and Recognition*. Springer; 2013. p. 197–204.
- [20] Schilling FP, Stadelmann T. Artificial neural networks in pattern recognition: proceedings of the 9th IAPR TC3 workshop, ANNPR 2020, Winterthur, Switzerland, September 2–4, 2020. In: *Lecture Notes on Artificial Intelligence*. Springer; 2020. .
- [21] Sajja PS. Examples and Applications on Artificial Neural Networks. In: *Illustrated Computational Intelligence*. Springer; 2021. p. 101–153.
- [22] Gao L, Wu W. Relevance assignment feature selection method based on mutual information for machine learning. *Knowledge-Based Systems*. 2020;209:106439.
- [23] Solorio-Fernández S, Carrasco-Ochoa JA, Martínez-Trinidad JF. A review of unsupervised feature selection methods. *Artificial Intelligence Review*. 2020;53(2):907–948.
- [24] Al-Kasassbeh M, Mohammed S, Alauthman M, Almomani A. Feature selection using a machine learning to classify a malware. In: *Handbook of computer networks and cyber security*. Springer; 2020. p. 889–904.
- [25] Xu J. A weighted linear discriminant analysis framework for multi-label feature extraction. *Neurocomputing*. 2018;275:107–120.