Mutual Information-based Feature Selection Strategy for Speech Emotion Recognition using Machine Learning Algorithms Combined with the Voting Rules Method

Hamza Roubhi
ETA Laboratory, Department of Electronics, University Mohamed El Bachir El Ibrahimi of Bordj Bou Arreridj, Algeria
hamza.roubhi@univ-bba.dz (corresponding author)

Abdenour Hacine Gharbi
LMSE Laboratory, Department of Electronics, University Mohamed El Bachir El Ibrahimi of Bordj Bou Arreridj, Algeria
abdenour.hacinegharbi@univ-bba.dz

Khaled Rouabah
Electronics Department, University of M'sila, University Pole, Road of Bordj Bou Arréridj, M'sila 28000, Algeria
khaled.rouabah@univ-msila.dz

Philippe Ravier
PRISME Laboratory, University of Orleans, France
philippe.ravier@univ-orleans.fr

Received: 22 September 2024 | Revised: 16 October 2024, 20 October 2024, and 21 October 2024 | Accepted: 26 October 2024

ABSTRACT

This study proposes a new approach to Speech Emotion Recognition (SER) that combines a Mutual Information (MI)-based feature selection strategy with simple machine learning classifiers such as K-Nearest Neighbor (KNN), Gaussian Mixture Model (GMM), and Support Vector Machine (SVM), along with a voting rule method. The main contributions of this approach are twofold. First, it significantly reduces the complexity of the SER system by addressing the curse of dimensionality by integrating a focused feature selection process, resulting in considerable savings in both computational time and memory usage. Second, it enhances classification accuracy by using selected features, demonstrating their effectiveness in improving the overall performance of the SER system. Experiments carried out on the EMODB dataset, using various feature descriptors, including Mel-frequency Cepstral Coefficients (MFCC), Perceptual Linear Prediction (PLP), and Linear Prediction Cepstral Coefficients (LPCC), showed that the best performance was achieved by GMM, with an accuracy of 85.27% using 39 MFCC features, compared to an accuracy of 82.55% using a high-dimensional vector with 111 features. Furthermore, applying the Joint Mutual Information (JMI) selection technique to extracted MFCC features reduces the vector size by 23.07% while improving the accuracy to 86.82%. These results highlight the effectiveness of combining the feature selection process with machine learning algorithms and the voting rules method for the SER task.

Keywords-speech emotion recognition; machine learning; voting rules; feature selection; mutual information

I. INTRODUCTION

In recent years, Emotion Recognition (ER) has emerged as a critical component in various domains, including human-computer interaction, gaming, marketing, mental health, and call centers, using different modalities, such as speech and facial expressions [1–4]. In particular, Speech Emotion Recognition (SER) is an important task in various application domains, requiring the development of accurate and less complex systems. Common classification approaches are based on lightweight architectures based on feature vector extraction and must operate quickly to be suitable for real-time applications. These approaches use feature vectors, which are sets of numerical values representing various speech characteristics extracted from audio signals to classify different emotions. The design of an SER system requires a training phase for modeling different emotion classes and a testing phase for classifying test signals and evaluating the system's performance. Each phase involves a feature extraction step that divides each signal into a sequence of overlapping analysis windows, then converted into a sequence of feature vectors. Frequently discussed and referenced in numerous studies, these features are short-term spectral features, including MFCC, PLP, and LPCC [5]. Particularly, MFCC features have proven their relevance for SER in several studies [6].

To improve response time and learning performance, these systems must be used in conjunction with the feature selection process. In this context, several studies have proposed different techniques to reduce the complexity effects in terms of computation time and space memory, even when using complicated classifiers [7, 8]. For instance, in [8], the GMM model was explored as a single-state HMM model implemented using HTK tools. This HMM classifier allows the classification of entire sequences of vectors into specific emotional classes using the Viterbi algorithm. In addition, a Mutual Information (MI)-based feature selection approach was used to select optimal features. The results showed that using MFCC coefficients and energy features, along with their dynamic features, achieved maximum accuracy, employing only 32 selected features, evaluated on the EMODB dataset. However, as demonstrated in [9], an HMM classifier takes more time compared to a KNN classifier combined with a voting rule method. Building on this, this work aims to overcome this limitation and enhance the computation time performance of the SER system proposed in [8]. Specifically, this study proposes extending the voting rule strategy not only to a KNN classifier but to other machine learning algorithms, namely GMM and SVM, along with MI-based feature selection to further reduce the system's complexity.

The main contributions of this study are as follows:

- Expands the work in [8] by including additional classifiers to develop a simplified and less complicated SER system.
- Carries out a comparative study of machine learning classifiers, including GMM, KNN, and SVM, demonstrating reduced complexity compared to the HMM classifier [9].
- Implements an MI-based feature selection strategy to enhance system performance in terms of accuracy, memory space, and computational time.

Since most machine learning classifiers cannot simultaneously classify a sequence of feature vectors, this study proposes classifying each feature vector individually and then applying a voting rules method to determine the emotional class of the feature vector sequence corresponding to the input signal. To the best of our knowledge, the combination of GMM, KNN, and SVM classifiers with voting rules-based decisions represents a new approach in SER systems. Therefore, careful selection of the most relevant features is recommended to avoid the curse of dimensionality, which can occur when adding more features. This work ultimately aims to minimize the SER system's complexity by reducing the number of features, thereby improving performance. This operation represents the second objective of this work and involves the application of feature selection algorithms based on maximizing MI.

II. PROPOSED SER SYSTEM USING MACHINE LEARNING CLASSIFIERS COMBINED WITH A VOTING RULE STRATEGY

This study aims to improve the performance of the SER system in terms of both complexity and accuracy using simplified classifiers. The latter, including GMM, KNN, and SVM, allow the classification of each feature vector into an emotion class. Then, a voting rule is applied to the sequence of the previously obtained class indices to determine the emotion class of the signal. The proposed model utilizes the MFCC feature extraction method and combines these classifiers with a voting rule strategy. Figure 1 illustrates the diagram of the automatic classification system of speech occurrences into emotion classes, where the classifier block can be the GMM, KNN, or SVM. In this figure, the red rectangle indicates the novel aspects with respect to [8]. It is worth noting that the structure of this scheme has been applied in several domains, as mentioned in [9, 10].

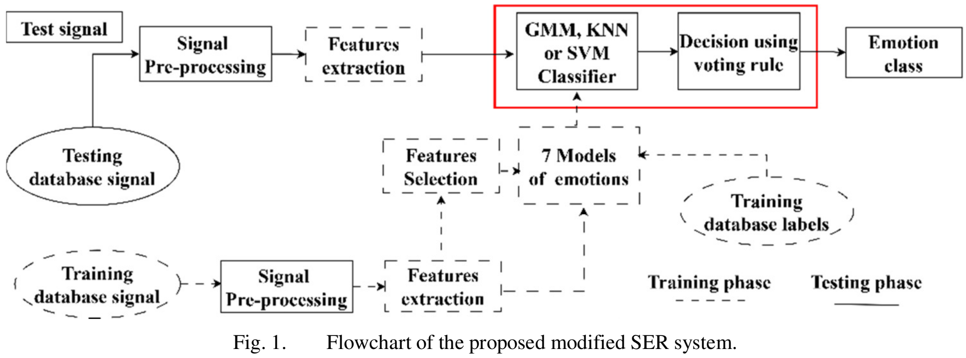

Fig. 1. Flowchart of the proposed modified SER system.

The diagram includes a training phase to model various emotion classes and a testing phase to identify test signals and evaluate system performance. Each phase involves a feature extraction step that divides each signal into a sequence of overlapping analysis windows. These windows are then transformed into a sequence of feature vectors, computed every 10 ms using 30 ms Hamming-windowed speech frames, with the 'Hcopy' command from the HTK toolkit [8]. In the training phase, these feature vectors are fed into a machine-learning classifier to model the emotion classes. In the testing phase, each feature vector within the sequence is classified into an emotion class using the trained classifier. A voting rule is then applied to the sequence of obtained class indices to determine the class of the input signal. The diagram also presents the feature selection process for dimensionality reduction.

III. MATERIALS AND METHODS

A. Theoretical Background

1) GMM Classifier

A GMM [11] is a sophisticated probability density function represented as a weighted combination of Gaussian component densities and is widely utilized in biometric systems, especially in speech recognition. The model is characterized by parameters such as mean vectors, covariance matrices, and mixture weights. Although there are multiple GMM variants, the choice often depends on the available data and the specific type of application. Maximum likelihood estimation is predominantly employed to determine the model's parameters with the iterative Expectation-Maximization (EM) algorithm. Alternatively, Maximum A Posteriori (MAP) estimation can adapt parameters from a universal background model. The Gaussian distribution formula is given as follows:

$$N(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} exp\left(-\frac{1}{2}(x-\mu)^T \overset{-1}{\Sigma}(x-\mu)\right) \quad (1)$$

where d represents the dimension of the feature vector, μ is the mean, and Σ denotes the covariance matrix.

2) KNN Classifier

KNN [12] is a non-parametric method for pattern classification that classifies data points by comparing them to their k nearest neighbors in the dataset. It does not make assumptions about the data's structure but instead relies on the features and labels of the stored data. When classifying a new point, it looks at the categories of its k neighbor points, often measured using the Euclidean (as used here), and assigns the category that appears most frequently among them. The value of k determines how many neighboring points influence the classification. The fundamental workflow of the K-NN classifier is outlined as follows:

1. Calculate the Euclidean distance between the unknown feature vectors $A\ (A_1, A_2, \ldots, A_d)$ and all the known feature vectors $B\ (B_1, B_2, \ldots, B_d)$ using the following formula:

$$distance(A, B) = \sqrt{\sum_{i=1}^{d} (A_i - B_i)^2} \quad (2)$$

where d denotes the number of features.

2. Select the shortest k distances from the unknown feature vectors.
3. Determine the most frequent class label among these k neighbors through majority voting.

3) SVM Classifier

The SVM classifier [13] is a refined learning mechanism primarily designed for binary classification. This study extends its application to multiclass classification, specifically for SER encompassing seven distinct classes. Its core principle is to identify an optimal hyperplane that effectively segregates the data. For non-linearly separable data, kernel functions become essential by adeptly transforming the inherently non-linear problem into a linear one in an augmented dimensional space. Among the various kernel functions, the Radial Basis Function (RBF) was used, which is mathematically expressed as follows:

$$K(x, x') = exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right) \quad (3)$$

where x and x' represent the input data points, and the parameter σ determines the kernel's width and affects the flexibility of the decision boundary.

4) Feature Selection

Feature selection is a crucial preprocessing approach aimed at finding the most important features. It represents an essential step in machine learning and pattern recognition processes, significantly improving the performance in terms of accuracy and computational time by simply selecting the most relevant features [14]. There are three main approaches to feature selection: Wrapper approaches, Filter approaches, and Embedded approaches [15-17]. The first approach depends on the precision of the classification system in measuring relevance, which requires training and testing for each set of possible features. As a result, this method becomes costly in terms of computational resources when handling features with a large number of dimensions [18]. The second approach selects a subset of features using a relevance measurement that is independent of the classification system's performance. While this method is not as precise as the wrapper approach, it is faster and more convenient when working with a large number of dimensions [19].

This study used the Filter approach to classify speech emotions, based on the relevance criterion of maximizing the MI. MI indicates how much uncertainty in one variable decreases when the other is known. The MI between two random discrete variables $X$and Y is given as follows:

$$I(X;Y) = \sum_y \sum_x p(x,y) log_2 \frac{p(x,y)}{p(x)p(y)} \quad (4)$$

where x and y denote the samples of variables X and Y, and $p(x)$ and $p(y)$ indicate the edge probabilities of X and Y, respectively. Additionally, $p(x,y)$ represents their joint probability density. For continuous variables, the MI is described as:

$$\mathcal{I}(X;Y) = \iint_{-\infty}^{+\infty} f(x,y) log_2 \left(\frac{f(x,y)}{f(y)f(y)}\right) dxdy \quad (5)$$

The joint distribution function, denoted as $f(x,y)$, shows how X and Y relate to each other, while $f(x)$ and $f(y)$ represent the probabilities of X and Y individually.

Estimating MI for continuous variables requires the discretized $I(X;Y)$ formula. This estimate uses Sturges's rule [20]. The main objective of feature selection is to identify a subset S of k relevant features from a set F of n features $F = \{X_1, X_2, \ldots, X_n\}$ that preserves essential information for classification. This is realized by finding the maximum MI with target classes via the following formula [18].

$$S_{opt} = arg\max_{S\in\mathcal{F}} I(C;S) \quad (6)$$

However, exploring all feature combinations in a large dataset is computationally challenging due to exponential growth in possibilities. To address this issue, iterative greedy algorithms, such as sequential forward selection, are used. These algorithms select features one at a time based on predefined criteria, thus reducing the computational load. When using a greedy forward selection method with MI as the relevance criterion, the approach is highly effective [21]. Most algorithms rely primarily on measurements derived from a maximum of three variables, which include two features and the class index. The first strategy, called MI Maximization (MIM), is defined as the simplest criterion for selecting features in step $j+1$.

$$X_{p_{j+1}} = arg\max_{X_i\in\mathcal{F}-S_j} I(C;X_i) \quad (7)$$

where $S_j = S_{j-1} \cup \{X_{pj}\}$ represents the subset of features selected at step j.

In this approach, the relevance of each feature X_i is evaluated individually, without considering redundancy with the subset of features S_j used to interpret the class index C in step j. Many algorithms, based on various criteria, are designed to optimize relevance and reduce redundancy in feature selection [18]. This study specifically focused on examining and implementing four strategies: Maximum-Relevance Minimum Redundancy (MRMR), Joint MI (JMI), Conditional Infomax Feature Extraction strategy (CIFE), and Interaction Capping (ICAP) [22, 23]. For these four strategies, $X_{p_{j+1}}$ is given as:

$$X_{p_{j+1}} = \begin{cases} arg\max_{X_i\in\mathcal{F}-S_j}\left[MI(C;X_i) - \frac{1}{|S|}\sum_{k=1}^{j} MI\left(X_i;X_{p_k}\right)\right] & \text{for } MRMR \text{ [24]} \\ arg\max_{X_i\in\mathcal{F}-S_j}\left[MI(C;X_i) - \frac{1}{|S|}\sum_{k=1}^{j}\begin{bmatrix} MI(X_i;X_{p_k}) - \\ MI(X_i;X_{p_k} \mid C)\end{bmatrix}\right] & \text{for } JMI \text{ [25]} \\ arg\max_{X_i\in\mathcal{F}-S_j}\left[MI(C;X_i) - \sum_{k=1}^{j} \max\begin{bmatrix} MI(X_i;X_{p_k}) - \\ MI(X_i;X_{p_k} \mid C), 0\end{bmatrix}\right] & \text{for } ICAP \text{[26]} \\ arg\max_{X_i\in\mathcal{F}-S_j}\left[MI(C;X_i) - \sum_{k=1}^{j}\begin{bmatrix} MI(X_i;X_{p_k}) - \\ MI(X_i;X_{p_k} \mid C)\end{bmatrix}\right] & \text{for } CIFE \text{ [27]} \end{cases} \quad (8)$$

Several studies have proposed stopping criteria to estimate the optimal number of relevant features in the selection process [28]. This study applied a stopping criterion considered under two cases: Case 1 and Case 2. In Case 1, a feature subset is considered the best if it achieves an accuracy equal to or greater than the accuracy obtained when using all features. This approach is efficient because it stops as soon as the target accuracy is reached, thus avoiding further selection with all features and typically resulting in a smaller subset. In Case 2, the focus is on achieving the highest possible accuracy, even if it matches the accuracy of the full feature set. This method usually takes more time because it continues to compare different subsets until it finds the best accuracy, which might lead to a larger subset of features compared to Case 1. Both cases provide a valuable understanding of feature selection by balancing the trade-off between efficiency and maximizing classification performance.

B. EMO-DB Speech Database

The EMO-DB database comprises 10 distinct German statements obtained from ordinary talks categorized into two groups: Set A consists of 5 short statements and Set B consists of 5 longer ones. Ten voice actors, with an equal distribution of males and females, spoke these statements while imitating seven basic emotions: anger, boredom, disgust, fear, happiness, sadness, and a neutral state. The database, consisting of 535 utterances, was first recorded at a sampling rate of 48 kHz and subsequently downsampled to 16 kHz. The study utilized Set A, consisting of 277 utterances, as the training dataset, and Set B, comprising 258 utterances, for testing purposes. It is important to note that test sentences have different content from the training ones, leading to an SER system that operates in a text-independent mode [29]. Table I summarizes the emotion distribution of the EMO-DB dataset.

TABLE I. DISTRIBUTION OF EMO-DB SENTENCES ACROSS 7 EMOTIONAL STATES, CATEGORIZED BY STATE FOR TEST/TRAIN

Emotions	**Anger**	**Boredom**	**Disgust**	**Fear**	**Happi-ness**	**Sad-ness**	**Neutral**
Number	127	81	46	69	71	62	79
Test/Train	62/65	40/41	21/25	34/35	33/38	30/32	38/41

IV. EXPERIMENTS, RESULTS, AND DISCUSSION

This section presents the results and discussion of the proposed SER system. Remember that the novelty of the proposed work lies in developing an SER system using MI-based feature selection with simple machine learning classifiers such as KNN, GMM, and SVM, combined with a voting rule strategy. The primary contributions of this study are as follows:

- Implements a simple SER system using machine learning classifiers combined with a voting rule strategy.
- Applies MI-based feature selection strategies, including CIFE, mRMR, JMI, and ICAP to vectors of 39 MFCC features.
- Combines MFCC, LPCC, and PLP features using the aforementioned feature selection strategies.
- Estimates the optimal number of features achieved using the proposed stopping criterion.

A. Performance Results with Machine Learning Classifiers Using MFCC Descriptor

This subsection details the results obtained using machine learning classifiers, including KNN, SVM, and GMM, with the MFCC descriptor. Here, feature selection strategies are applied to the classifier that yielded the best performance, and its dimension will be extended.

1) KNN Classifier

Table II shows the results of a KNN classifier with k ranging from 1 to 50. Only the results corresponding to k ranging from 1 to 10 are illustrated, as no improvement was observed for values of k greater than 10. The default Euclidean distance was used, and the best result was obtained with a k equal to 2.

TABLE II. ACCURACY OF THE SER SYSTEM USING THE KNN CLASSIFIER AS A FUNCTION OF NUMBER OF NEIGHBORS K

Value of k	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Accuracy (%)	76.36	**76.53**	74.80	75.19	75.19	75.58	74.03	72.88	72.48	70.54

2) SVM Classifier

For the SVM model with the RBF kernel function, two crucial parameters were examined: the Box Constraint (BC) and the kernel scale. The BC parameter controls the tolerance to classification errors, while the kernel scale affects the model's flexibility. The kernel scale was set to auto, allowing it to be optimized automatically by the function, and the BC was adjusted from 2 to 18. As shown in Table III, the optimal combination was obtained with a BC equal to 8.

TABLE III. ACCURACY OF THE SER SYSTEM USING THE SVM CLASSIFIER AS A FUNCTION OF BC

SVM BC	**2**	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**18**
Accuracy (%)	82.17	83.33	83.33	**83.72**	83.33	82.55	82.17	81.78	82.17

3) GMM Classifier

In the case of the GMM classifier, experiments were carried out to determine the optimal number of Gaussian components of each emotion class to achieve the best accuracy. As shown in Table IV, the results indicated that the best accuracy was obtained with 14 Gaussian components and 39 MFCC features. The experimental results also showed that the GMM classifier achieved the highest accuracy of 85.27%, surpassing KNN (76.53%) and SVM (83.72%). It is also worth noting that the accuracy of the proposed system using the same MFCC descriptor combined with the voting rule method was superior compared to the system in [8], which achieved 84.5%.

TABLE IV. ACCURACY OF THE SER SYSTEM USING THE GMM CLASSIFIER AS A FUNCTION OF GMM COMPONENTS

GMM components	**2**	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**18**
Accuracy (%)	77.51	79.84	78.68	82.17	80.23	82.55	**85.27**	77.90	78.68

To extend this study, additional features were integrated, such as PLP and LPCC, followed by implementing feature selection techniques on the winning classifier, which is GMM.

B. Feature Selection Results with GMM Classifier using MFCC Descriptor

Table V shows the relevance of the selected features along with their corresponding accuracy. Note that the best feature selection strategy is the one that provides better accuracy using a reduced number of selected features. It was observed that each selection strategy provided the same accuracy and number of features for both cases of stopping criterion. The results indicate that the JMI strategy improved the system performance by using a reduced subset of features (30 features from 39), achieving a maximum accuracy of 86.82%. Meanwhile, the ICAP strategy maintained the same initial accuracy of 85.27% with only 30 features.

TABLE V. ACCURACY AND NUMBER OF RELEVANT FEATURES USING CIFE, JMI, MRMR, AND ICAP STRATEGIES WITH THE MFCC DESCRIPTOR

	Case 1: Accuracy (Selected features) ≥ Accuracy (All features)		**Case 2: Maximum Accuracy**	
	No. of selected features	**Accuracy (%)**	**No. of selected features**	**Accuracy (%)**
CIFE	39	85.27	39	85.27
JMI	30	86.82	30	86.82
mRMR	39	85.27	39	85.27
ICAP	30	85.27	30	85.27

The impact of the dimensionality curse phenomenon was examined on the results of the SER system. Figure 2 presents the graphical representation of the feature selection process, showing the curves of variation of accuracy (Recognition Rate) as a function of the number of selected features using CIFE, JMI, mRMR, and ICAP strategies. It can be observed that all curves reach a plateau with minor fluctuations when the number of selected features exceeds around 15. In addition, the system reaches a maximum accuracy of 86.82% with 30 features.

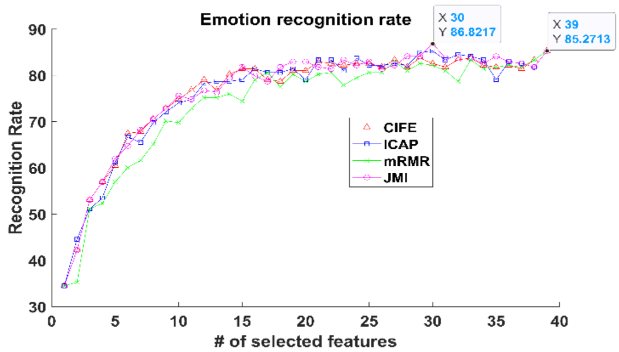

Fig. 2. Variation in emotion recognition accuracy as a function of the number of selected features, using CIFE, JMI, mRMR, and ICAP strategies with the GMM classifier and MFCC descriptor.

C. Feature Selection Results with GMM Classifier using a Large Dimension of Vectors

To evaluate the performance of the enhanced system, PLP and LPCC descriptors were added to MFCC, making a total of 111 features. The GMM components were also varied from 2 to 18 and the results are illustrated in Table VI.

TABLE VI. ACCURACY OF THE SER SYSTEM USING THE GMM CLASSIFIER AS A FUNCTION OF GMM COMPONENTS

GMM components	**2**	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**18**
Accuracy (%)	78.29	82.17	**82.55**	74.80	73.64	72.48	74.03	73.64	73.25

The best accuracy of 82.55% was achieved with 6 components, which is still superior compared to the system in [8], which achieved an accuracy of 81.01%. However, it was observed that adding LPCC and PLP features affects the accuracy of the SER system, which can probably be explained by the curse of dimensionality phenomenon. Table VII highlights the results of the feature selection strategies after adding the PLP and LPCC descriptors to the MFCC and their impact on accuracy. The following observations summarize the performance across strategies and feature sets:

First case:

- The accuracy remains consistent at 82.55% for the CIFE, JMI, and mRMR strategies, despite variations in the number of selected features.
- The ICAP strategy shows a slight improvement in accuracy to 82.94% with 42 selected features.

Second case:

- CIFE maintains its accuracy of 82.55% with 111 features, suggesting no improvement with additional features.
- JMI slightly improves the accuracy to 82.94% with 88 features, indicating that it achieves higher accuracy with fewer features compared to CIFE.
- mRMR achieves a notable accuracy of 84.10% with 104 features, demonstrating its effectiveness in selecting a slightly larger subset of features for better performance.
- ICAP achieves the highest accuracy of 84.49% with 95 features, highlighting its ability to identify an optimal set of features that significantly improves accuracy.

TABLE VII. ACCURACY AND NUMBER OF RELEVANT FEATURES USING CIFE, JMI, MRMR, AND ICAP STRATEGIES WITH HIGH-DIMENSIONAL VECTORS.

	First case: Accuracy (Selected features) ≥ Accuracy (All features)		**Second case: Maximum Accuracy**	
	No. of selected features	**Accuracy (%)**	**No. of selected features**	**Accuracy (%)**
CIFE	111	82.55	111	82.55
JMI	68	82.55	88	82.94
mRMR	48	82.55	104	84.10
ICAP	42	82.94	95	84.49

The impact of the dimensionality curse phenomenon was examined on the results of the SER system. Figure 3 presents the results of applying the four feature selection strategies (CIFE, JMI, mRMR, and ICAP) to high-dimensional vectors. The curves in this figure demonstrate that all strategies reach a plateau at about 40 selected features, which are sufficient to clearly explain the classes. Consequently, they improve the performance of the SER system in terms of processing time and memory usage.

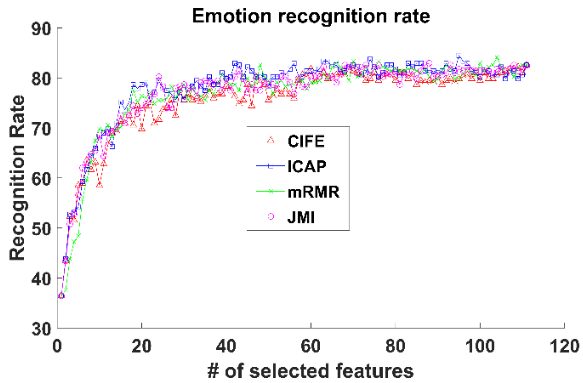

Fig. 3. Variation in emotion recognition accuracy as a function of the number of selected features using CIFE, JMI, mRMR, and ICAP strategies with the GMM classifier for high-dimensional vectors.

V. CONCLUSIONS

This study proposed feature selection strategies based on MI combined with machine learning classifiers and a voting rule technique to improve classification accuracy while reducing the number of features, thus lowering computational costs and memory requirements. First, the GMM, SVM, and KNN classifiers were applied, combined with a voting rule strategy, to implement a simple and less complicated SER system using the MFCC descriptor. Second, the best classifier was identified and feature selection strategies were evaluated on low-dimensional vectors with 39 MFCC features and high-dimensional vectors with 111 features. The results demonstrate that the GMM classifier combined with the voting rule method yielded an accuracy of 85.27% with MFCC descriptors and 82.55% with a larger feature set, surpassing the accuracy of a related work. Furthermore, the use of feature selection strategies improved performance to 86.82% using the JMI strategy with just 30 features. Future work will explore a broader range of feature selection techniques, including Principal Component Analysis (PCA) and Independent Component Analysis (ICA), and provide a comparison of JMI with other feature selection methods. These analyses aim to further optimize system performance, contributing to the development of improved SER systems.

REFERENCES

[1] R. Cowie *et al.*, "Emotion recognition in human-computer interaction," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 32–80, Jan. 2001, https://doi.org/10.1109/79.911197.

[2] C. J. Huang *et al.*, "Intelligent feature extraction and classification of anuran vocalizations," *Applied Soft Computing*, vol. 19, pp. 1–7, Jun. 2014, https://doi.org/10.1016/j.asoc.2014.01.030.

[3] A. Dey and K. Dasgupta, "Emotion Recognition Using Deep Learning in Pandemic with Real-time Email Alert," in *Proceedings of Third International Conference on Communication, Computing and Electronics Systems*, Coimbatore, India, 2022, pp. 175–190, https://doi.org/10.1007/978-981-16-8862-1_13.

[4] A. Dey and K. Dasgupta, "Mood Recognition in Online Sessions using Machine Learning in Realtime," in *2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP)*, Chennai, India, May 2021, pp. 1–6, https://doi.org/10.1109/ICCCSP52374.2021.9465495.

[5] S. Basu, J. Chakraborty, A. Bag, and Md. Aftabuddin, "A review on emotion recognition using speech," in *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, Coimbatore, India, Mar. 2017, pp. 109–114, https://doi.org/10.1109/ICICCT.2017.7975169.

[6] I. Trabelsi and M. S. Bouhlel, "Comparison of Several Acoustic Modeling Techniques for Speech Emotion Recognition:," *International Journal of Synthetic Emotions*, vol. 7, no. 1, pp. 58–68, Jan. 2016, https://doi.org/10.4018/IJSE.2016010105.

[7] R. Subramanian and P. Aruchamy, "An Effective Speech Emotion Recognition Model for Multi-Regional Languages Using Threshold-based Feature Selection Algorithm," *Circuits, Systems, and Signal Processing*, vol. 43, no. 4, pp. 2477–2506, Apr. 2024, https://doi.org/10.1007/s00034-023-02571-4.

[8] A. Hacine-Gharbi and P. Ravier, "On the optimal number estimation of selected features using joint histogram based mutual information for speech emotion recognition," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 9, pp. 1074–1083, Nov. 2021, https://doi.org/10.1016/j.jksuci.2019.07.008.

[9] F. Ghazali, A. Hacine-Gharbi, P. Ravier, and T. Mohamadi, "Extraction and selection of statistical harmonics features for electrical appliances identification using k-NN classifier combined with voting rules method," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 27, no. 4, pp. 2980–2997, Jul. 2019, https://doi.org/10.3906/elk-1812-80.

[10] F. Ghazali, A. Hacine-Gharbi, and P. Ravier, "Statistical features extraction based on the discrete wavelet transform for electrical appliances identification," in *Proceedings of the 1st International Conference on Intelligent Systems and Pattern Recognition*, Oct. 2020, pp. 22–26, https://doi.org/10.1145/3432867.3432900.

[11] D. Reynolds, "Gaussian Mixture Models," in *Encyclopedia of Biometrics*, Boston, MA, USA: Springer US, 2009, pp. 659–663.

[12] E. Fix and J. L. Hodges, "Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties," *International Statistical Review / Revue Internationale de Statistique*, vol. 57, no. 3, Dec. 1989, Art. no. 238, https://doi.org/10.2307/1403797.

[13] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, https://doi.org/10.1007/BF00994018.

[14] N. Yusup, A. M. Zain, and A. A. Latib, "A review of Harmony Search algorithm-based feature selection method for classification," *Journal of Physics: Conference Series*, vol. 1192, Mar. 2019, Art. no. 012038, https://doi.org/10.1088/1742-6596/1192/1/012038.

[15] F. Z. Boukhobza, A. Hacine Gharbi, and K. Rouabah, "A New Facial Expression Recognition Algorithm Based on DWT Feature Extraction and Selection," *The International Arab Journal of Information Technology*, vol. 21, no. 4, 2024, https://doi.org/10.34028/iajit/21/4/6.

[16] S. Bashir, I. U. Khattak, A. Khan, F. H. Khan, A. Gani, and M. Shiraz, "A Novel Feature Selection Method for Classification of Medical Data Using Filters, Wrappers, and Embedded Approaches," *Complexity*, vol. 2022, no. 1, Jan. 2022, Art. no. 8190814, https://doi.org/10.1155/2022/8190814.

[17] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1–2, pp. 273–324, Dec. 1997, https://doi.org/10.1016/S0004-3702(97)00043-X.

[18] A. H. Gharbi, P. Ravier, and M. N. Meziane, "Relevant harmonics selection based on mutual information for electrical appliances identification," *International Journal of Computer Applications in Technology*, vol. 62, no. 2, 2020, Art. no. 102, https://doi.org/10.1504/IJCAT.2020.104691.

[19] S. Nuanmeesri and W. Sriurai, "Multi-Layer Perceptron Neural Network Model Development for Chili Pepper Disease Diagnosis Using Filter and Wrapper Feature Selection Methods," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7714–7719, Oct. 2021, https://doi.org/10.48084/etasr.4383.

[20] A. Hacine-Gharbi, M. Deriche, P. Ravier, R. Harba, and T. Mohamadi, "A new histogram-based estimation technique of entropy and mutual information using mean squared error minimization," *Computers & Electrical Engineering*, vol. 39, no. 3, pp. 918–933, Apr. 2013, https://doi.org/10.1016/j.compeleceng.2013.02.010.

[21] D. François, F. Rossi, V. Wertz, and M. Verleysen, "Resampling methods for parameter-free and robust feature selection with mutual information," *Neurocomputing*, vol. 70, no. 7–9, pp. 1276–1288, Mar. 2007, https://doi.org/10.1016/j.neucom.2006.11.019.

[22] G. Brown, A. Pocock, M.-J. Zhao, and M. Luján, "Conditional likelihood maximisation: a unifying framework for information theoretic feature selection," *The journal of machine learning research*, vol. 13, no. 1, pp. 27–66, 2012.

[23] S. M. Hameed, W. A. Ahmed, and M. A. Othman, "Leukemia Diagnosis using Machine Learning Classifiers based on MRMR Feature Selection," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 15614–15619, Aug. 2024, https://doi.org/10.48084/etasr.7720.

[24] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005, https://doi.org/10.1109/TPAMI.2005.159.

[25] H. Yang and J. Moody, "Feature selection based on joint mutual information," in *Proceedings of international ICSC symposium on advances in intelligent data analysis*, 1999, vol. 23.

[26] A. Jakulin, "Machine Learning Based on Attribute Interactions," Ph.D. dissertation, Univerza v Ljubljani, Slovenia, 2005.

[27] D. Lin and X. Tang, "Conditional Infomax Learning: An Integrated Framework for Feature Extraction and Fusion," in *Computer Vision – ECCV 2006*, vol. 3951, A. Leonardis, H. Bischof, and A. Pinz, Eds. Springer Berlin Heidelberg, 2006, pp. 68–82.

[28] R. Touahria, A. Hacine-Gharbi, P. Ravier, and M. Mostefai, "Relevant Multi Domain Features Selection Based on Mutual Information for Heart Sound Classification:," in *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods*, Rome, Italy, 2024, pp. 918–923, https://doi.org/10.5220/0012565300003654.

[29] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, "A database of German emotional speech," in *Interspeech 2005*, Sep. 2005, pp. 1517–1520, https://doi.org/10.21437/Interspeech.2005-446.