# Speech Emotion Recognition with SVM, KNN and DSVM

**Hadhami Aouani †,†† and Yassine Ben Ayed ††**

**† National School of Engineers, ENIS University of Sfax, Tunisia**
**†† Multimedia InfoRmation systems and Advanced Computing Laboratory, MIRACL University of Sfax, Tunisia**

**Summary**
Speech Emotions recognition has become the active research theme in speech processing and in applications based on human-machine interaction. In this work, our system is a two-stage approach, namely feature extraction and classification engine. Firstly, two sets of feature are investigated which are: the first one is extracting only 13 Mel-frequency Cepstral Coefficient (MFCC) from emotional speech samples and the second one is applying features fusions between the three features: Zero Crossing Rate (ZCR), Teager Energy Operator (TEO), and Harmonic to Noise Rate (HNR) and MFCC features. Secondly, we use two types of classification techniques which are: the Support Vector Machines (SVM) and the k-Nearest Neighbor (k-NN) to show the performance between them. Besides that, we investigate the importance of the recent advances in machine learning including the deep kernel learning. A large set of experiments are conducted on Surrey Audio-Visual Expressed Emotion (SAVEE) dataset for seven emotions. The results of our experiments showed given good accuracy compared with the previous studies.

***Keywords:***
*Emotion recognition, MFCC, ZCR, TEO, HNR, KNN, SVM, Deep SVM.*

## 1. Introduction

Emotions color our language and it can make its meaning more complex. The listener interacts with the emotional state of the speaker and adapts his behavior to any kind of emotion transmitted by the speaker.

Speech recognition of emotions (SER) aims to identify the emotional or physical state of a human being from his or her voice. This is a relatively recent research topic in the field of speech processing. The automatic recognition of emotions by analyzing the human voice and facial expressions has become the subject of numerous researches and studies in recent years [1]. The fact that automatic emotion recognition systems can be used for different purposes in many areas has led to a significant increase in the number of studies on this subject. The following systems can be cited as an example of the areas in which these studies are used and their intended use:

- Education: a course system for distance education can detect bored users so that they can change the style or level of material provided in addition, provide emotional incentives or compromises.
- Automobile: driving performance and the emotional state of the driver are often linked internally. Therefore, these systems can be used to promote the driving experience and to improve driving performance.
- Security: They can be used as support systems in public spaces by detecting extreme feelings such as fear and anxiety.
- Communication: in call centers, when the automatic emotion recognition system is integrated with the interactive voice response system, it can help improve customer service.
- Health: It can be beneficial for people with autism who can use portable devices to understand their own feelings and emotions and possibly adjust their social behavior accordingly [26].

SER can be realized using automatic learning methods including the extraction and classification of vocal functions [2]. For a better generalization, the characteristics must be well defined [3].

This document deals with the extraction of feasible characteristics to indicate whether it is doable to offer a good precision to the SER and to determine the relevant characteristics. Different classification techniques namely support vector machines (SVM), k-nearest neighbor (k-NN) and the use of Deep SVM methods.

The rest of this article is planned as follows: Section 2 presents the literature survey. Section 3 describes the architecture of our system. Section 4 presents the results of the experimental and comparative studies. The results of the SVM method will be compared to the KNN method, as well as the Deep SVM methods. Finally, in section 5, we present some conclusions.

## 2. Related Work

The acoustic features principally classified as prosody, spectral and voice quality features [4]. Features such as

Manuscript received August 5, 2023
Manuscript revised August 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.8.6

energy, pitch, and Zero Crossing Rate (ZCR) are considered prosodic features, and Linear Predictive Coding (LPC) and Mel Frequency Cepstral Coefficient (MFCC) are considered spectral characteristics.

Lots researchers have done different models of speech recognition using different combinations of features.

Ya Li et al. [5] extract two categories of characteristics which are: the audio characteristics obtaining by the software 'OpenSmile' such as energy level and low spectral descriptors, the sum of the auditory spectrum, the slope, MFCC, the spectral flow and the triggering of low-level descriptors that are a fundamental frequency F0, Formant (F1, F2, F3). And the video features like membership, shape characteristics and face detection by the tracking algorithm (Viola and Jaunes) using the Random Forest classification method (RF) to identify the eight emotions of the base Chinese Natural Audio-Visual Emotion database (CHEAVD) for multimodal recognition [11].

Song et al. [6] propose a new method of non-negative matrix transfer factorization (TNMF) use the two databases the first is Berlin database that contains seven emotions: Anger, boredom, disgust, fear, happiness, sadness and neutrality. The second eNTERFACE'05 with six emotions Anger, disgust, fear, happiness, sadness and surprise.

In the work presented by Papakostas et al. [7] aimed to analyze the emotions of the speakers on the basis of paralinguistic information. They exploit two automatic learning approaches: a support vector machine (SVM), consisting of a set of 34 extracted features, and a convolutional neuron network (CNN). The datasets used were EMOVO, SAVEE and EMOtional speech-DataBase (EMO-DB). The emotions represented from the EMOVO and SAVEE datasets were the six basic emotions. Seven emotions used at EMO-DB are disgust, anger, joy, fear, sadness, boredom and neutrality.

Ramdinmawii et al. extracted the features such as energy E0, zero crossing rate (ZCR), formant (F1, F2, F3) and fundamental frequency F0, which serve to detect the four types of emotions (joy, fear, anger and neutral) of two German databases (German Emotion Database) and Telugu (Telugu Emotion Database) [8].

Shi uses the Multimodal Emotion Recognition Competition Database (MEC2017), collected in movies and television, to rank eight emotions (neutral, pissed, sad, happy, anxious, worried, surprised and ashamed) with two classifiers which are the Vector Support Machines (SVM) and Artificial Neural Networks (ANN) using two categories of characteristics: the original characteristics which are the Mel frequency cepstral coefficient , the fundamental frequency, the length of the audio, the length of silence in the audio sequence and their mean values, and new features extracted from deep belief networks (DBNs). The results of the classification of the new feature are greater than the original functionality by at least 5%. It can also be noted that the result obtained with DBN-SVM was slightly improved compared to the result obtained with DBN-DNN, which is due to a better classification capacity in a small sample [9].

Siddique Latif et al. [10] made use of transfer learning technique to improve the performance of SER systems by using the eGeMAPS feature set containing 88 features. Evaluations of five different data sets in three different languages reveal that the Deep Belief Network (DBN) offers greater precision than previous approaches.

The main contribution to this work is to use two conventional classification methods such as SVM and KNN, and to compare the results. We have proposed a deep learning, this is the deep SVM using the SAVEE database.

## 3. Proposed system

In this work, we propose a system that addresses speech recognition and its objectives improving outcomes using MFCC and ZCR, TEO and Harmonic to Noise Ratio (HNR) features using two classifiers SVM and KNN firstly and secondly the use of deep learning. The proposed architecture is shown below Fig. 1.

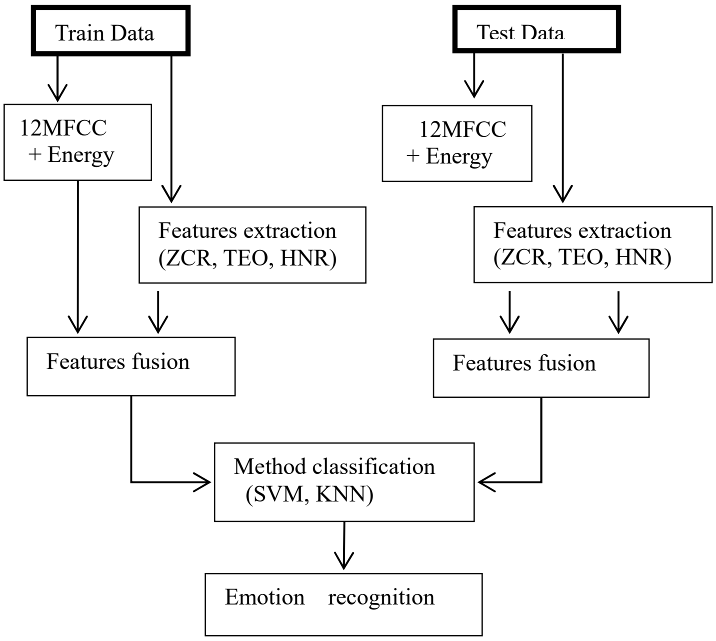
Train Data
Test Data
12MFCC + Energy
12MFCC + Energy
Features extraction (ZCR, TEO, HNR)
Features extraction (ZCR, TEO, HNR)
Features fusion
Features fusion
Method classification (SVM, KNN)
Emotion recognition

Fig.1. Architecture of our system of emotion recognition.

In this article, feature extraction used MFCC. After extracting the MFCC entities, save them as feature vectors.

The characteristic vectors are entered in the two classifiers SVM and KNN are used for the classification of the emotions. Then, extract the prosodic features ZCR, TEO and HNR and save them as feature vectors by performing feature merge with MFCC coefficients to gain 16 characteristics in order to proceed with the classification with both methods. This brings the need to select characteristics in the recognition of emotions In order to improve the results of SVM and KNN, we have proposed the deep learning. To increase the performance of this system, we proposed DSVM.

### 3.1 Feature Extraction

In this work, we use 12 MFCC coefficients+Energy [11], Zero Crossing Rate (ZCR), Teager Energy Operator (TEO) and Harmonic to noise Ratio (HNR).

- Mel-frequency Cepstral Coefficient

To calculate the MFCC coefficients, the inverse Fast Fourier Transform (IFFT) is applied to the logarithm of the Fast Fourier Transform (FFT) module of the signal, filtered according to the Mel scale [18].

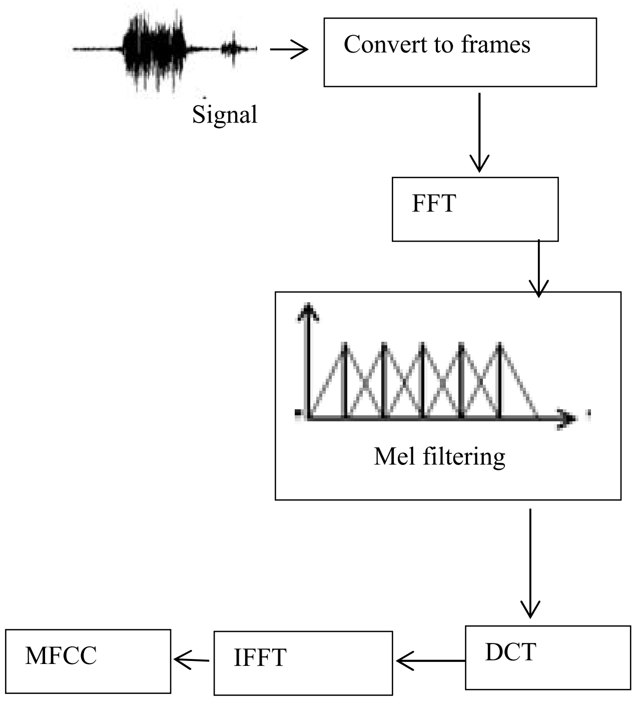

Fig.2. Steps for calculating MFCC coefficients.

The method to find MFCCs is generally with the following steps. These steps are illustrated in the figure (Fig.2.)

The initial step, apply the Fast Fourier Transform on input signal. In next step, map the power of the spectrum obtained in above step to the Mel scale. In next step take the logs of powers at each of the Mel frequencies of speech signal. Then take Discrete Cosine Transform on bank of Mel log powers. In this final step, we convert the log Mel spectrum back to time.

The result is called the Mel frequency cepstrum coefficients (MFCC).

- Zero Crossing Rate

The Zero Crossing Rate (ZCR) is an interesting parameter that has been used in many speech recognition systems. As the name suggests, it is defined by the number of zero crossings in a defined region of the signal, divided by the number of samples in that region [12].

$$ZCR = \frac{1}{N-1}\sum_{n=1}^{N-1} sign\big(s(n)s(n-1)\big)$$

Where $sign\big(s(n)s(n-1)\big) = \begin{cases} 1, & s(n)s(n-1) \geq 0 \\ 0, & s(n)s(n-1) < 0 \end{cases}$

Zero crossing points for a defined signal region are shown in this figure Fig. 3.

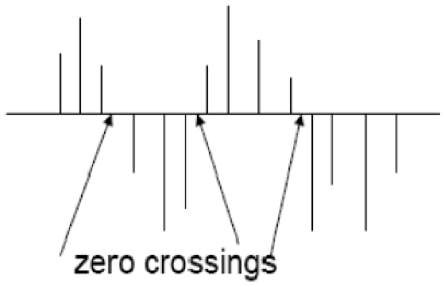

Fig. 3. Definition of Zero Crossing Rate.

- Teager Energy Operator

Teager Energy Operator (TEO) functions check the characteristics of the speech when the utterance presents a certain stress. TEO functions measure the non-proximity of the utterance by treating its behavior in the frequency and time domain.

For estimation of the TEO, each output of the M signal is segmented into frames of equal length (for example, 25 milliseconds with a frame offset of 10 milliseconds); where M is the number of critical bands and f is the number of the frame for which the TEO is extracted. In our work we extract the TEO from the total of signal.

$$\Psi M\,[xf\,[t]] = (xf\,[t])^2 - (xf\,[t-1]\,xf\,[t+1]))$$

- Harmonic to Noise Ratio

The harmonic-to-noise ratio (HNR) is a measure of the proportion of harmonic noise in the voice measured in decibels [13].It describes the distribution of the acoustic energy between the harmonic part and the inharmonic part of the radiated vocal spectrum.

## 3.2 Classification

The extracted features were introduced in two machine-learning models in order to choose the most efficient. In this work, experiments conducted on SVM and k-NN. Then the use of the deep learning which is Deep SVM [11].

- Support Vector Machines (SVM)

At first, SVM is proposed to distinguish between 2 classes. Several approaches have been proposed in order to extend the binary classifier to multi-class classification tasks. In fact, the multi-class SVMs are used in several fields and have proved their effectiveness in identifying the different classes of data presented to it [19].

Indeed, the data is represented by S = {(xi, yi) with xi Rn, i = 1, m and yi {1... k}} where k is the number of classes used.

The resolution of this problem by the SVMs is done initially considering a decomposition that combines several binary classifiers. In this case we find three types of methods: one-against-all [20], one-on-one [21] and DAGSVM [22].

Finally the problem is transformed by Vapnik [23] and Weston [24] to a simple optimization problem aimed at seeking the minimization of the following quantity:

$$\phi(\omega,\xi) = \frac{1}{2}\sum_{j=1}^{k}(\omega_j.\omega_j) + C\sum_{i=1}^{m}\sum_{j\neq y_i}\xi_i^j \qquad (3)$$

*Or*

$$(\omega_{yi}.x_i) + b_{yi} \geq (\omega_j.x_i) + b_j + 2 - \xi_i^j$$

*And* $\xi_i^j \geq 0$ , $j = 1,\ldots, m,$ $j \in \{1,\ldots, k\}¥\, yi$

The principle of SVMs is based on the construction of the optimal hyperplane that is the best separating of the training data projected in the feature space by a kernel function K. the most used kernel functions, such as linear, polynomial, RBF. Therefore, the use of this classification method essentially consists of selecting good kernel functions and adjusting the parameters to obtain a maximum identification rate.

We will use the SVM with these three kernel functions, so that:

— **Linear:** K($x_i$,$x_j$)=$x_i^T x_j$

— **Polynomial:** K($x_i$,$x_j$)=($\gamma\ x_i^T x_j$+$r$)$^d$,$\gamma$>0

— **RBF:** K($x_i$,$x_j$)=exp(−$\gamma$‖$x_i$−$x_j$‖$^2$),$\gamma$>0

With:

d: Degree of polynomial,

r: weighting parameter (used to control weights)

$\gamma$: kernel flexibility control parameter.

Then, the adjustment of the various parameters of the SVM classifier is done in an empirical way, each time we modify the type of the SVM kernel in order to determine the values $\gamma$, r, d and c which are values chosen by the user, in order to find the most suitable kernel parameters for our search.

- k-Nearest Neighbour (k-NN)

The k-Nearest Neighbour algorithm (k-NN) is a widely used classifier for classifying objects based on closest training examples in the feature space. It is the simplest classifier of all machine learning algorithms [25].
The k is a constant defined by the user; it is the nearest neighbors for whom we wish to vote.
In our work, the kNN algorithm based on Euclidean distances.

- Deep Support Vector Machine (DSVM)

Deep Support Vector Machine is an -level multiple kernel architecture with h sets of m kernels at each layer [16] (Fig.4.).

Our deep SVM forms an SVM in the standard way, and then uses the kernel activa-tions of the support vectors as inputs to form another SVM to the next layer. We will use four unique base kernels for each layer: linear kernel, kernel polynomial with degree 2, kernel polynomial with degree 3 and kernel RBF (Fig.5.).

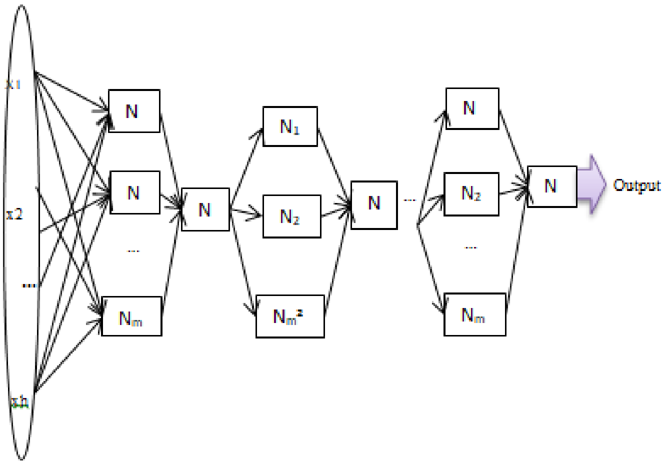

Fig. 4. Architecture of Deep Support Vector Machine.

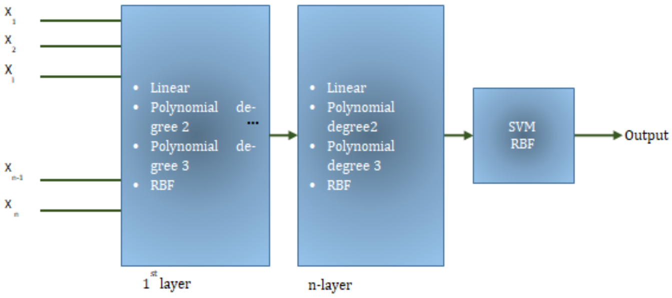

Fig. 5. Forms of our Deep SVM.

## 4. Experiments and results

In our work, we used the Surrey Audio-Visual Expressed Emotion Database (SAVEE) is widely used in the emotional recognition of speech [17]. It is easily ac-cessible and well annotated.

This database includes seven different emotions, namely anger, disgust, fear, neutral, happy, surprise and sad. The experiments are performed on four subjects called DC, JE, JK and KL. Each emotion class contains 15 samples except the neutral class contains 30 samples and a total of 120 samples per class available, or 480 sentences in total, whose audio are wave type and have a sampling rate of 44.1 kHz. Each audio sample is represented on 32 bits.

We divided each subject earlier each subject class of our corpus in two parts: 80% dedicated for learning and 20% for the test phase.

The table 1 presents a summary of the best recognition rate found for the three SVM kernels as a function of features.

Table 1.The recognition rates on the test corpus obtained with the SVM as a function of features.

| Features<br>Kernel of SVM | 12 MFCC +Energie | 12 MFCC +Energie, ZCR,TEO,HNR |
|---|---|---|
| Kernel Linear | 60.86 | 65.22 |
| Kernel Polynomial | 66.67 | 73.91 |
| Kernel RBF | 72.43 | 74.29 |

The results obtained show that the improvement of the recognition rate is proportional to the increase in characteristics.

The best results for different emotions using the 16 functions (13 MFCCs with ZCR, TEO and HNR) with RBF core SVM are summarized in the following Table 2:

Table 2. The recognition rates on the test corpus obtained using SVM with RBF kernels as a function of 13 MFCC and fusion features.

| Features<br>Emotions | 12 MFCC +Energie | 12 MFCC +Energie, ZCR,TEO,HNR |
|---|---|---|
| Angry | 77.77 | 77.78 |
| Disgust | 77.77 | 77.77 |
| Fear | 55.55 | 55.55 |
| Happy | 76.66 | 66.66 |
| Neutral | 79.88 | 86.66 |
| Sad | 55.55 | 88.88 |
| Surprise | 44.44 | 66.77 |

The two figures below (Fig.6, Fig.7) show the variation of the KNN identification rates for each change in the value of k on the two systems respectively: the system using the 13 MFCCs and the one with the fusion features (13MFCC, ZCR, TEO and HNR).

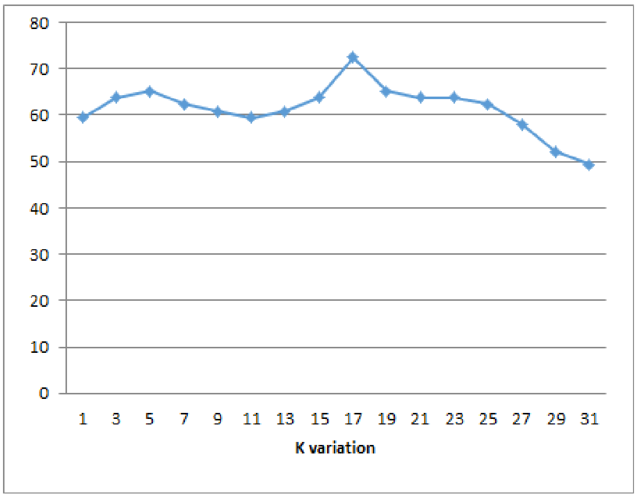

Fig.6. The identification rates found of KNN method for each change in the value of k on the system with 13 MFCC.

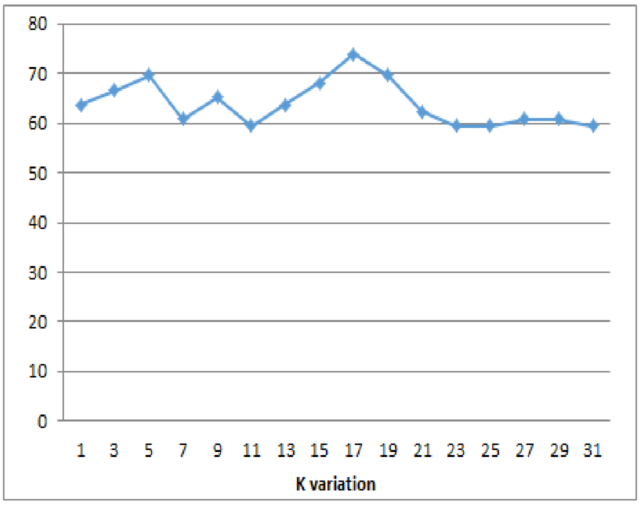

Fig. 7. The identification rates found of KNN method for each change in the value of k on the system with fusion features

After a series of experiments by varying the value k, we conclude that for both systems when the value k is 17, we have identified the best recognition rates which are equal to 72.46% and 73.91% respectively for the two systems: with 13 MFCC and the system using fusion features.

The table below represents a summary of the best results found for each emotion with k = 17 for both systems: 13 MFCC and fusion features (13MFCC, ZCR, TEO and HNR).

Table 3. The recognition rates on the test corpus obtained using KNN as a function of 13 MFCC and fusion features.

| Features<br>Emotions | 12 MFCC +Energie | 12 MFCC +Energie, ZCR,TEO,HNR |
|---|---|---|
| Angry | 88.88 | 77.77 |
| Disgust | 55.55 | 89 |
| Fear | 44.44 | 56 |
| Happy | 66.66 | 44.44 |
| Neutral | 86.66 | 80 |
| Sad | 77.77 | 88.88 |
| Surprise | 77.77 | 78 |

The tables below summarize the best recognition rates found for the seven emotions across the characteristics used and the different systems used.

Table 4. Results obtained by the different systems for the system with 13 MFCC.

| Methods<br>Emotions | SVM | DSVM |
|---|---|---|
| Angry | 77.77 | 100 |
| Disgust | 77.77 | 88.88 |
| Fear | 55.55 | 44.44 |
| Happy | 76.66 | 66.66 |
| Neutral | 79.88 | 100 |
| Sad | 76.55 | 77.77 |
| Surprise | 55.55 | 66.66 |

Table 5. Results obtained by the different systems for the system with fusion features.

| Methods<br>Emotions | SVM | DSVM |
|---|---|---|
| Angry | 77.78 | 88.88 |
| Disgust | 44.44 | 100 |
| Fear | 77.77 | 88.88 |
| Happy | 88.88 | 44.44 |
| Neutral | 86.66 | 93.33 |
| Sad | 66.66 | 66.66 |
| Surprise | 66.77 | 88.88 |

These tables show the effectiveness of the DSVM algorithm for the two systems: systems of 13 MFCC and the system using fusion features which give an overall rate equal to 79.71% and 82.60% respectively.

The table below summarizes the best recognition rates found for our two systems using the different methods of classifications.

Table 6. The recognition rates obtained using different classifiers for both systems.

| Systems<br>Classifiers | 13 MFCC | Fusion features (13MFCC,ZCR,TEO,HNR) |
|---|---|---|
| KNN | 72.46 | 73.91 |
| SVM | 72.43 | 74.29 |
| DSVM | 79.71 | 82.60 |

For the use of KNN method, we find after a series of experiments that the best recognition rates are of the order of 73.91% fusion features systems and 72.46% for the system using 13 MFCC.

Using the RBF kernel SVM model, the best result is 74.29% for the system that combines the features that is superior to the system with 13 MFCC which gives a rate equal to 72.43%.

Table 7 shows a comparison of the proposed SER system with some of the recent published studies that used SAVEE dataset with different classifiers and feature.

Table 7. Comparison of our proposed system with some previous studies that used the SAVEE database.

| **Related work** | **Methods and No. of features** | | **Recognition Accuracy (%)** | **No. of emotions** |
|---|---|---|---|---|
| **Noroozi et al. [15]** | Random Forest (RF)<br>13 features | | 66.28 | 6 emotions |
| **Papakostes et al. [7]** | SVM<br>CNN<br>34 features | | 30 with SVM<br>25 with CNN | 4 emotions |
| **Siddique Latif et al. [10]** | DBN<br>88 features | | 56.76 | 7 emotions |
| **Current study** | 13 MFCC | KNN | 72.46 | 7 emotions |
| | | SVM | 72.43 | |
| | | DSVM | 79.71 | |
| **Current study** | Fusion features<br>(13MFCC,ZCR,TEO,HNR) | KNN | 73.91 | 7 emotions |
| | | SVM | 74.29 | |
| | | DSVM | 82.60 | |

## 5. Conclusion

In this paper, we propose two systems, one that use 13 MFCC only and the other that merges the features: 13 MFCC, Zero Crossing Rate (ZCR), Teager Energy Operator (TEO), and Harmonic noise ratio (HNR) by testing them with different methods of classifications. The first method of classification is the Support Vector Machines (SVM) which gives the best result only for the merger system features. The second method is the k-Nearest Neighbor (KNN), after a series of experience on the change in value of k, we obtain a better recognition rate equal to 73.91% with the system of fusion features compared to the system using 13 MFCC that gives recognition rate equal to 72.46%. In order to improve the results, we have proposed the Deep SVM (DSVM) which shows its efficiency in the two systems: system using 13 MFCC and system with fusion features with a rate respectively equal to 79.71% and 82.60%.

Our comparative study of emotion classification systems demonstrates the effectiveness of using the fusion

features system with the SVM method compared to system with 13 MFCC as well as the use of Deep SVM (DSVM) for the two systems.
This work achieves better accuracy using the two systems in comparison with recent previous studies that used the same dataset.

Two speech emotion recognition systems based on seven emotions were proposed in this article using different classifiers and compared their performance. Future work should try to think of other descriptors. We can also consider performing the recognition of emotions using an audiovisual base (image and speech) and in this case to benefit from the descriptors from speech and others from image. This allows us to improve the recognition rate of each emotion.

## References

[1]. Simina Emerich, Eugen Lupu — Improving Speech Emotion Recognition using Frequency and Time Domain Acoustic features, EURSAIP 2011.

[2]. Park, J.-S., J.-H. Kim and Y.-H. Oh, Feature vector classification based speech emotion recognition for service robots. IEEE Transactions on Consumer Electronics, 2009. 55(3).

[3]. A Dictionary of Physics. 7 ed. 2015: Oxford University Press.

[4]. Zhibing, X., Audiovisual Emotion Recognition Using Entropy estimation- based Multimodal Information Fusion. 2015, Ryerson University.

[5]. Hinton, G. E., and Salakhutdinov, R. R.Reducing the dimensionality of data with neural networks. Science 313(5786):504–507, 2006

[6]. P. Song, S. Ou, W.Zheng, Y. Jin, & L. Zhao: "Speech emotion recognition using transfer non-negative matrix factorization". In Proceedings of IEEE international conference ICASSP, pp. 5180–5184, 2016.

[7]. Papakostas, M., et al., Recognizing Emotional States Using Speech Information, in GeNeDis 2016. 2017, Springer. p. 155-164.

[8]. E. Ramdinmawii, A.Mohanta, V.K. Mittal: "Emotion Recognition from Speech Signal ", IEEE 10 Conference (TENCON), Malaysia, November 5-8, 2017.

[9]. P. Shi: "Speech Emotion Recognition Based on Deep Belief Network", IEEE, 2018.

[10]. Siddique Latif, R.R., Shahzad Younis, Junaid Qadir, Julien Epps, Transfer Learning for Improving Speech Emotion Classification Accuracy. ArXiv:1801.06353v3 [cs.CV] 2018.

[11]. Aouani H, Ben Ayed Y: "Emotion recognition in speech using MFCC with SVM, DSVM and auto-encoder",IEEE, 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP) 2018.

[12]. L.X.Hùng : Détection des émotions dans des énoncés audio multilingues. Institut polytechnique de Grenoble, 2009.

[13]. Ferrand, C: Speech science: An integrated approach to theory and clinical practice. Boston, MA: Pearson, 2007.

[14]. Noroozi, F., et al., Vocal-based emotion recognition using random forests and decision tree. International Journal of Speech Technology, 20(2): p. 239-246, 2017.

[15]. M. Swerts and E. Krahmer. Gender-related differences in the production and perception of emotion. In Proc. Interspeech, pages {334,337}, 2008.

[16]. Eric V. Strobl & Shyam Visweswaran:' Deep Multiple Kernel Learning' ICMLA, 2013.

[17]. http:// personal.ee.surrey.ac.uk/Personal/ P Jackson/SAVEE.

[18]. Y. Ben Ayed : Détection de mots clés dans un flux de parole. Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications ENST, 2003.

[19]. DELLAAERT F., POLZIN T., WAIBEL A., "Recognizing Emotion in Speech ", Proc.of ICSLP,Philadelphie , 1996.

[20]. L. Bottou, C. Cortes, J. Drucker, I. Guyon, Y. LeCunn, U. Muller, E. Sackinger, P. Simard et V. Vapnik : "Comparaison of classifier methods : a case study in handwriting digit recognition", dans Proceedings of the International Conference on Pattern Recognition, p.77_87, 1994.

[21]. S. Knerr, L. Personnaz et G. Dreyfus : "Single-layer learning revisited: a stepwise procedure for building and training a neural network", Neurocomputing: Algoritms, Architectures and Applications, p.68, 1990.

[22]. J. C. Platt, N. Cristianini et J. Shawe-Taylor : "Large margin dags for multiclass classification", dans Advances in Neural Information Processing Systems, MIT Press, 12, p.547_553, 2000.

[23]. V. Vapnik : "Statistical learning theory", John Wiley and Sons, 1998.

[24]. J. Weston et C. Watkins : "Support vector machines for multiclass pattern recognition", In Proceedings of the Seventh European Symposium On Artificial Neural Networks, 1999.

[25]. A. Amina, A. Mouhamed, and C. Morad. Identification des personnes par système multimodale.

[26].Sucksmith, E., Allison, C., Baron-Cohen, S., Chakrabarti, B., & Hoekstra, R. A. Empathy and emotion recognition in people with autism, first-degree relatives, and controls. Neuropsychologia, 51(1), 98-105,2013