ZANCO Journal of Pure and Applied Sciences

TYPE Original Research
DOI http://dx.doi.org/10.21271/ZJPAS.36.5.10

OPEN ACCESS

*Corresponding author
Omar Shirko Mustafa
Omar.mustafa@epu.edu.iq

RECEIVED :13 /07/ 2024
ACCEPTED :15/09/ 2024
PUBLISHED :31/ 10/ 2024

KEYWORDS:
Emotion recognition;
Speech analysis;
Sorani Kurdish.

Emotion Recognition in Kurdish Speech from the Sorani Dialect Corpus

Omar Shirko Mustafa[1*], Shavan Kamal Askar[1], Shahab Wahab Karim[1,2], Bzar Khidir Hussein[1]

[1]Department of Information System Engineering, Erbil Technical Engineering College, Erbil, Polytechnic University, Erbil, Kurdistan Region, Iraq
[2] Department of Computer Engineering, College of Engineering, Knowledge University, Erbil, Kurdistan Region, Iraq

ABSTRACT

Given the increasing need for interactive human-computer applications, the field of employing machine learning algorithms to discern emotions from speech has seen a substantial surge in interest. While emotion recognition systems have made substantial progress in languages like German, English, Spanish, Dutch, and Danish, the availability of comprehensive datasets for the Kurdish language remains notably limited. This paper addresses this gap by focusing on emotion recognition in Sorani Kurdish dialect speech data, which was carefully gathered from openly available videos from the YouTube platform and categorized into four clear supposed emotions: neutral, sadness, happiness, and anger. The study applied both natural Mel Spectrogram and Mel-Frequency Cepstral Coefficient (MFCC) features for various spectrals, followed by the classification models K-Nearest Neighbor (KNN), Multi-Layer Perceptron (MLP), and Support Vector Machine (SVM) to evaluate the results. By closely examining and contrasting the results of using several methods for feature extraction, it was found that SVM obtained a higher accuracy, reaching as much as 85.57%. This is so much more than the first Kurdish emotion classification technique for the recognition of the emotion of the words.

1.Introduction

Human communication can be described as verbal and nonverbal messages that interact to build communication (Scholl, 2013, Hazmoune and Bougamouza, 2024). These signposts are, however, not exclusive to the vocal aspect in that they consist of intonation, pitch, facial expression, and writing. Add a depth of context that surpasses the literal interpretation of words. This intricacy underscores the importance of the mind getting the emotional signs right to maintain effective communication (Tushnet et al., 2020). Adding emotion recognition to speech recognition is called speech emotion recognition (SER), which is capable of detecting and understanding the non-verbal dimensions of language on its path (Al-Talabani et al., 2015, Martinez-Lucas et al., 2024, Garcia-Cuesta et al., 2024, Gurowiec and Nissim, 2024). Though SER has the potential to change the world of communication significantly, the mechanism still has inherent challenges (Al-Talabani et al., 2015). These range from the interpretation of emotions using facial expressions to the solution of cultural problems and coping with the sounds of the environment, such as background noise (Al-Talabani et al., 2015, Garcia-Cuesta et al., 2024). Yet, it is the field that explores the potential of SER in providing a detailed analysis of emotional components in human speech, and so it is the area of study that has a great impact on the linguistics and artificial intelligence. Implementing SER systems extends to various domains necessitating human interaction. For example, the importance of SER in automotive systems that monitor driver well-being (Can et al., 2023). Call centers also benefit from their ability to recognize and respond to callers' emotional states. In addition, SER has proven invaluable in the field of psychiatric diagnosis, providing therapists with a supportive tool (Ramakrishnan and El Emary, 2013).

Speech datasets are of paramount importance for training and testing SER systems, for which a varied audio collection having corresponding emotion states should be annotated (Sajjad and Kwon, 2020, El Ayadi et al., 2011). These datasets are context sensitive, covering scripted dialogue, spontaneous conversations, and simulated customer service, hence the application area extending (Sajjad and Kwon, 2020, Pulatov et al., 2023). Notable datasets in this field include the Berlin Database of Emotional Speech, the Interactive Emotional Dyadic Motion Capture (IEMOCAP) data-base, and the Ryerson Audio-Visual Database of Emotional Speech and Song (RAV-DESS), among others (Swain et al., 2018). Researchers keep increasing and diversifying those data sets in order to improve the precision, reliability, and competence of SER systems for diverse circumstances. However, many speech emotion datasets are available for such languages as German, English, and Spanish, which are widely spoken. Some of the re-search works in the SER are broadened to include Dutch, Danish, Mandarin, etc. As well as several languages from Asia and Europe (Kasuriya et al., 2019). Conversely, however, there are very few Kurdish language datasets available in the field (Al-Talabani, 2015). Kurdish, a language of Indo-Iranian origin, is the language most widely spoken by Kurds, an ethnic group that spreads across Turkey, Syria, Iran, and Iraq, as shown in (Figure1) (Kurda, 2022). It is filled with longs decades of literature writing. Among various dialects of Kurdish, Sorani and Kurmanji Kurdish are the ones of the highest prevalence. The most distinctive of Kurdish languages is Sorani which carries the highest significance in the Kurdish linguistic area (Dulz, 2016). It occupies the position as the majority dialect spoken by millions people across the Kurdistan region of Iraq as well as its neighboring areas. Be-sides being one of the entitled languages of Kurdish Regional Government (KRG) both in practice and theory, Sorani Kurdish language carries great important in governance, education and mutual communication (Sheyholislami, 2017). Beyond that, it has two kinds of usage; written and spoken, to refer to literature and formal documents. Also, its central position and importance on the other side, there is no adequate data sets of Sorani Kurdish speech emotion, making it an urgent issue for more research and development focus on refining SER systems to the suitability of the Sorani Kurdish language.

Figure 1: Kurdish-Inhabited Area

In this paper, the focus is on detailing the endeavor to compile a dataset centered on the Sorani Kurdish dialect, as it is the formal dialect in KRG, encompassing four prevalent emotional states: sad, glad, mad, and angry. Spectral characteristics were aimed for perfect feature extraction in SER. Additionally; we applied three prominent machine learning classification methods known for their efficacy in discerning emotions from speech: K-nearest neighbor (KNN), Multi-Layer Perceptron (MLP), and Support Vector Machine (SVM). SVM, an algorithm of extreme power and efficiency, is being used extensively in classification and pattern recognition applications (Chavhan et al., 2010). KNN, being mostly used in SER systems, has proven to be a reliable classifier (Meftah et al., 2020). Be-sides, MLP is a neural network that has input and output layers that take incoming signals and predict the target output based on them (Joy et al., 2020).

The structure of this paper unfolds as follows: Section 2 introduces the relevant studies and their outcomes within this domain. Section 3 outlines the methodology employed in constructing the dataset, elaborates on the process of feature extraction and the proposed classification models. Section 4 displays and deliberates upon the findings. Finally, encapsulates the paper with a conclusion and outlines avenues for future research.

2. Literature Review

Emotion recognition systems in speech show their variability as a function of factors: language, dataset characteristics, feature extraction methods, and the classifiers used to distinguish emotions from spoken language. Researchers have utilized machine learning algorithms to understand emotions in different languages, for instance, the study of (Alamri, 2023) was on Arabic emotion detection, particularly in the Saudi dialect, by using data from the YouTube channel Telfaz11. This investigation touches on emotions like neutral, sadness, happiness, and anger. The experiments are based on the extraction features like MFCC and Zero-Crossing Rate (ZCR) from audio signals and the use of classifiers like SVM, KNN, Convolutional Neural Networks (CNN), and Long Short-Term Memory networks (LSTM).The results indicate the usability of MFCC and CNN reaching 95% accuracy in deciphering Arabic emotions. Furthermore, (Shahin et al., 2023) introduces the GWO-KNN method, combining the Grey Wolf Optimizer (GWO) with K-Nearest Neighbor (KNN) classification. This novel approach selects features emulating wolf hunting behavior in an imitative way that largely improves emotion identification systems performance. In addition, the evaluations was on diverse databases such as Arabic Emirati-accented speech, the Ryerson Audio-Visual Database of Emotional Speech and Song dataset (RAVDESS), and the Surrey Audio-Visual Expressed Emotion dataset (SAVEE), reveal superior classification compared to conventional methods such as the bat algorithm (BAT), cuckoo search (CS), White Shark Optimizer (WSH), arithmetic optimization algorithm (AOA), and recent approaches on the same datasets. Moreover, (El Seknedy and Fawzi, 2023) illustrates a bilingual Arabic-English SER with the utilization of the EYASE and RAVDESS datasets. It offers a new feature set that performs both spectral and prosodic features more efficiently and gainfully. Different ma-chine-learning classifiers were tested, and the SVM version turned out to be the most efficient with the highest emotion recognition accuracy. The suggested feature set out-performed the challenging "InterSpeech 2009" benchmark, achieving accuracies of 85% for RAVDESS and 64% for EYASE. Ensemble learning was successful in valence emotion detection where it performed 90% on RAVDESS and 87.6% on EYASE.

Numerous researches have been conducted to identify emotions in speech across a wide range of languages, but there is a lack of studies for SER in the Kurdish Language, which we can encompass with a group of researchers at Buckingham University. In reference (Al-Talabani et al., 2013), a study proposed a method for improving emotion recognition by combining Prediction Cepstral Cepstral Coefficients (LPCC) and MFCC features with Wavelet Octave Coefficient of Residual (WOCOR) as excitation source features. These were integrated with 6552 Low-Level Descriptors (LLDs) using SVM and ANN classifiers. The experiments are tested on a newly acquired emotional speech database in the Kurdish language, the Berlin emotional speech database, and the Aibo database. The Kurdish emotional speech database includes 7 emotions acted by 6 male and 6 female actors, with 3360 recordings. The actors utter 10 Kurdish sentences for each emotion, with different acting experiences and ages. The data files are recorded in WAV format with a 32 KHz sample rate. A subjective test was conducted to determine the perceptibility of the emotions, with an average correct labeling of 41%. The same researchers referenced in (Al-Talabani et al., 2013) later devised a method, as outlined in (Al-Talabani et al., 2015), with the specific aim of refining emotion recognition from speech. Their strategy entails extracting Meta-features through techniques like PCA or random projection, focusing on reducing the dimensionality of speech data. These gleaned features are then integrated to enhance the performance of the recognition model. Notably, their approach outperforms existing methods when applied to the same speech databases described in (Al-Talabani et al., 2013). Furthermore, researchers advocate for a broader perspective on emotion recognition, suggesting that it should not be treated exclusively as a classification problem, given the wide variety of emotions evident in speech.

In terms of evaluating the effectiveness of a speech recognition system on emotional speech data, it is crucial to create an appropriate database. There are four essential criteria in database preparation: scope, physical structure, contents, and chosen language. The scope of database design encompasses various facets of variation, including the number of speakers, gender of speakers, range of emotions, dialects, language type, and age group (Swain et al., 2018). However, there is no consensus regarding the specific criteria for the minimum or maximum number of speakers, units, and labelers that should be included in building the speech recognition database. Notably, there are several widely recognized databases with fewer than 10 speakers and some with just one or two labelers (Swain et al., 2018). Interestingly, unable to find any online databases specifically designed for recognizing emotions in Kurdish speech, particularly Sorani Kurdish. (Table 1) shows different characteristics of some existing corpora.

Table 1: Exemplars of Widely Used Emotional Speech Databases

Corpora	Emotion Types	Speakers	Language
IITKGP-SESC(Koolagudi et al., 2009)	Anger, Compassion, Disgust, Fear, Happy, Neutral, Sarcastic and Surprise	10 professional artists from All India Radio (5 male and 5 female)	Telugu
EMOVO (Costantini et al., 2014)	disgust, fear, anger, joy, surprise, sadness	6 professional actors (3 males and 3 females)	Italian
EMODB (Burkhardt et al., 2005)	Joy, Anger, Disgust, Fear, Boredom, Sadness, Neutral	10 actors (5 males, 5 females)	German
SAVEE (Jackson and Haq, 2014)	Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral	4 actors (All male)	English
DES (Engberg and Hansen, 1996)	Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral	4 actors, two of each gender	Danish

3. Methodology

3.1 Dataset Design

This study focused on Sorani Kurdish dialect to construct a dataset of semi-natural emotional speech. based on videos from the YouTube platform originating from the renowned Kurdish channel, Rudaw Media Network (Rudaw Media Network, 2013). Rudaw Media Network holds a prominent position as a major media outlet in the Kurdistan Region of Iraq. With its headquarters located in Erbil, the capital of Iraq's Kurdistan Region, the majority of its produced content is primarily in the Sorani Kurdish language.

A video set was meticulously reviewed and assessed to select scenes featuring perfect emotional cues. This stage is required to train and evaluate a number of ma-chine learning algorithms that specialize in accurate emotion recognition. Following this, audio files were retrieved from the mapped videos. To guarantee practicality and efficiency, these audio files are broken into smaller, more manageable chunks. This di-vision depended on natural pauses and gaps in speech, yielding pieces ranging from 1 to 13 seconds in duration. This method enhances the analysis and processing of emotional speech data, making it more suitable for algorithmic interpretation and application.

Four distinct emotions (neutral, sadness, happiness, and anger), which are commonly assessed in studies related to SER, were extracted from the dataset when examining the audio files. As a result, these emotion labels are assigned for chunks of audio. However, before processing and cleaning the data for classification purposes, a pre-processing step was performed to remove any noise and background music, as well as periods of silence.

The final dataset had 287 records, both male and female speakers' words being represented. There were a 189 chunks from men and 98 form women. The total length of the data loaded was around 16 minutes. A detailed breakdown of the data distribution is shown in (Table 2).

Table 2: Emotion Distribution in Speech Dataset

Emotion	Total Chunks	Chunks from Male	Chunks from Female	Total Length (minutes)
Anger	68	45	23	3:35
Happiness	85	55	30	4:32
Sadness	65	40	25	2:51
Neutral	69	49	20	2:48

3.2 Feature Extraction

The preliminary and most fundamental steps of SER include the extraction of required essential features from speech signals. This preliminary process quantifies various parameters relating to audio data. The features extracted serve as input to a classifier that can be used to differentiate between finer emotional states. The well-designed flow of this SER system has been encapsulated in the schematic representation in Figure 2.

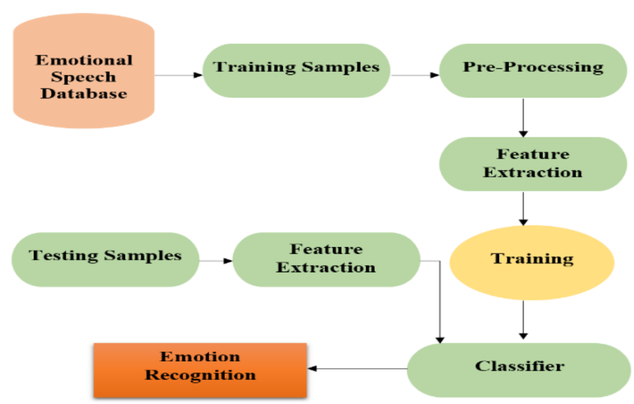

Figure 2: Kurdish SER Flow Diagram

This study is majoring in the extraction of spectral features from the dataset provided, transforming time-based signals into frequency-based characteristics. For this purpose, the Librosa Python library (McFee et al., 2015) which is able in audio analysis and feature extraction, was used. In feature extraction, the spectral feature employed several methods:

- MFCC, mainly adopted in SER systems. It captures subtle emotions that have been represented in acoustic features due to the nonlinear spacing of lower- and higher-frequency components. In the present study, 13 coefficients are extracted for each frame of audio data, which consequently results in a feature vector of size 13 × N, where N stands for the number of frames.

• Mel Spectrogram: In this process, Mel-scaled spectrograms are calculated further in order to make the representation of the audio data more analyzable. The resultant feature vector depends on the length of the audio file and the sampling rate. Generally, it results in a 40 × N-sized feature vector, where 40 is the count of the Mel bands.

• Spectral Contrast: It divides each frame of a spectrogram into sub-bands and calculates the contrast in energy by subtracting peak (top quantile) energies from valley (bottom quantile) energies within each sub-band (Jiang et al., 2002). The spectral contrast gives 7 contrasts of subbands for each frame, and thereby the length of the resulting feature vector will be of size 7 × N.

The number of these features taken together at the input to the classifier contributes to the total size of the feature vector. For example, concatenating all three features (MFCC, Mel Spectrogram, and Spectral Contrast) creates a feature vector size of 60 × N. This process has been done for classification tasks like SVM or MLP models. Figure 3 shows the extraction of audio spectral features.

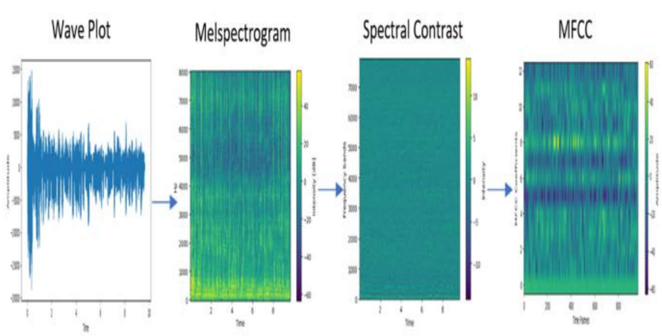

Figure 3: Spectral Signal Features Visualized

3.3 Classification

The Scikit-Learn machine learning library, version 0.24.1 (Pedregosa et al., 2011), is used to implement the classification of emotions presented in this paper. This library provides efficient tools for predictive data analysis, making it an optimal choice in the data science community. During data preprocessing, extracted audio features were standardized to rescale the distribution of features so that the mean of observed features is 0 and the standard deviation is 1. Then, the dataset was split into 85% as a training set and 15% as a test set.

It is worth noting that the impact of standardization scaling on model performance can be algorithm-specific. In this specific case, we observed a significant improvement in classification accuracy for models utilizing MLP and KNN. However, for the SVM, model standardization resulted in a notable decrease in accuracy, by approximately 30%. Moreover, among the three classification methodologies employed, the SVM model demonstrated the highest accuracy without standardization, achieving an accuracy rate of 85.57%. For a comprehensive comparison of classifier performance, please refer to Table 3. This tabulated data provides a detailed breakdown of accuracy scores, facilitating an informed selection of the most suitable model for the given task.

Table 2: Emotion Distribution in Speech Dataset

Classifier	Results	
	Unstandardized Accuracy	**Standardized Accuracy**
KNN	65.45%	77.30%
MLP	71.14%	82.37%
SVM	85.57%	59.90%

3.4 Mathematical Model for the Proposed System

The mathematical model for the system involves the following steps:

• Feature Extraction: Let the audio signal be represented as x(t). The MFCC feature extraction process applies the discrete cosine transform (DCT) on the logarithmic energy of Mel-filtered spectrograms:

$$\text{MFCC}=\text{DCT}(\log(\text{Mel-Spectrogram}(x(t)))) \qquad (1)$$

- Classification Model: The classification task can be formulated to minimize the loss function. For the SVM, the primal optimization problem can be expressed as:

$$min_{w,b,\delta} \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i \quad (2)$$

Subject to:

$$y_i(w^T \emptyset(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad (3)$$

where w is the weight vector, b is the bias term, and ξi are slack variables allowing for misclassifications.

4. Results and analysis

In this section, different experiment results were conducted to predict emotions from Sorani Kurdish dialect speech. The results of the various experiments utilizing both single and combination features are shown in (Table 4). In general, the most effective results were achieved by integrating three characteristics: MFCC, mel spectrogram, and spectral contrast when using SVM and MLP. Conversely, KNN demonstrated the maximum accuracy when combining MFCC and spectral contrast. Therefore, the utilization of feature combinations predominantly improved the classifier performance and provided more accuracy in comparison to single features.

Table 4: Rates of Emotion Recognition across Various Features

Extracted Features	SVM	MLP	KNN
MFCC	62.27%	65.45%	68.44%
MFCC + Mel spectrogram	73.72%	74.02	63.25%
MFCC + spectral contrast	62.40%	71.14%	77.30%
Mel spectrogram	48.40%	62.31%	51.50%
MFCC + Mel spectrogram + spectral contrast	85.57%	82.37%	65.18%

To examine the results in deeper detail, we created confusion matrices. (Figures 4, 5, and 6) show how different classifiers performed in terms of predicting each emotion. On the horizontal axis, find the labels of predicted, while the vertical axis represents the labels of true. In general, anger was recognized by all classifiers with the overall highest recognition rate. Notably, MLP demonstrated the most robust performance, achieving a prediction rate of 94.3%, surpassing all other classifiers. In contrast, for the other negative emotion, sadness, both SVM and KNN outperformed MLP. Achieved estimate rates of 87.6% and 76.5%, respectively, compared to MLP's 67.8%.

Interestingly, in the realm of neutral emotions, SVM emerged as the standout performer, achieving an impressive 92.6% prediction rate, while other classifiers faltered in this category. This discrepancy could be attributed to the inherent challenge of distinguishing neutral speech from other emotional states. On the other side, happiness proved to be a trickier emotion for all classifiers to predict, especially in comparison to other emotional states. This difficulty likely arose from the limited dataset records used for training classifiers in this specific emotion. Among the classifiers, SVM demonstrated a stronger aptitude in identifying happiness compared to the other two.

The performance of MLP stands out in its ability to accurately identify anger, surpassing its performance for the other three emotions. KNN, on the other hand, excels not only in identifying anger but also in recognizing sadness, demonstrating a particular strength in discerning negative emotions. In contrast, SVM exhibits impressive prediction rates for both anger and sadness of negative emotions and, notably, it achieves the highest prediction rate among all methods for neutral emotional states, showcasing its versatility and effectiveness in emotional recognition tasks.

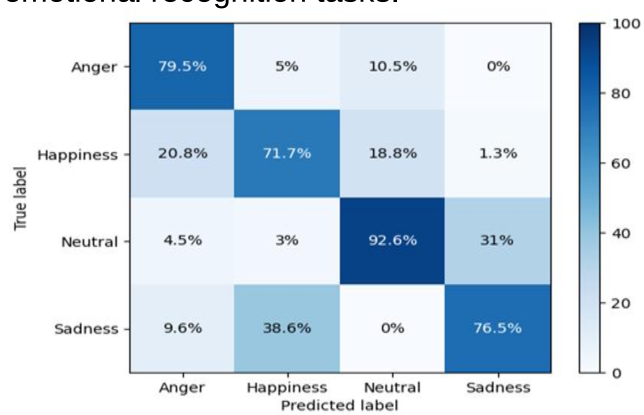

Figure 4: SVM confusion matrix

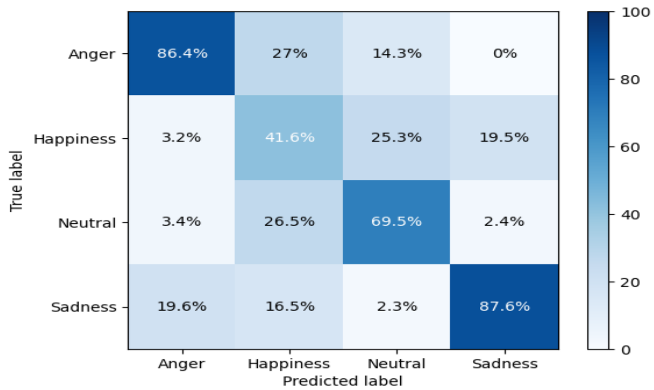

Figure 5: MLP confusion matrix

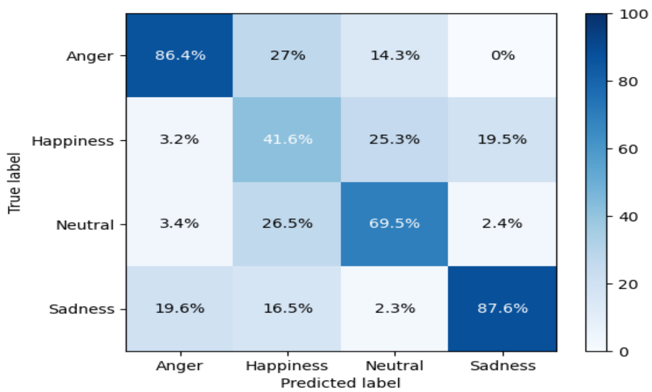

Figure 6: KNN confusion matrix

Conclusions and Future Work

Recent years have seen a rise in interest in SER systems due to their potential to improve human-machine interaction through two-way communication. Various crucial factors influence the performance of these systems, such as the number of classes, the quality of datasets, the preprocessing techniques of the data, feature selection methods, and the algorithms of classification.

Three different classifiers (KNN, MLP, and SVM) are used for comparative analysis in this study to predict emotions in speech spoken in the Sorani Kurdish dialect. A newly constructed dataset was employed, focusing on four emotions: neutral, sadness, happiness, and anger. The dataset was sourced from Rudaw Media Network videos, resulting in 287 records after preprocessing to remove noise and silence. Audio segments were labeled based on natural pauses and categorized into the four emotions. Spectral features, specifically MFCC and spectral contrast, demonstrated the highest accuracy for KNN at 77.30%. The inclusion of mel spectrogram features alongside the previous ones significantly improved the predictions for MLP and SVM, with accuracies of 85.57% and 82.37%, respectively. Furthermore, the results showed that, among all classifiers, anger was the emotion that could be predicted with the highest accuracy. Notably, SVM and KNN also exhibited strong performance in recognizing sadness, another negative emotion.

Undoubtedly, a major challenge encountered in this study was the limited size and imbalance of the dataset, particularly in the case of the "happiness" emotion, which led to lower accuracy for this class. Future work can address this issue by expanding the dataset and applying data augmentation techniques like time-stretching and pitch-shifting to balance the data and enhance model generalization. Additionally, dimensionality reduction methods such as Principal Component Analysis (PCA) can be utilized to reduce feature redundancy, improving classification performance while lowering computational complexity. Moreover, this research successfully applied machine learning algorithms (SVM, KNN, and MLP) to a semi-natural Kurdish speech dataset, achieving satisfactory results in this underexplored context, while future studies could explore advanced models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM), to further improve the accuracy and robustness of emotion recognition in Sorani Kurdish speech.

References

AL-TALABANI, A. 2015. Automatic speech emotion recognition-feature space dimensionality and classification challenges. University of Buckingham.

AL-TALABANI, A., SELLAHEWA, H. & JASSIM, S. Excitation source and low level descriptor features fusion for emotion recognition using SVM and ANN. 2013 5th computer science and electronic engineering conference (CEEC), 2013. IEEE, 156-161.

AL-TALABANI, A., SELLAHEWA, H. & JASSIM, S. A. Emotion recognition from speech: tools and challenges. Mobile Multimedia/Image Processing, Security, and Applications 2015, 2015. SPIE, 193-200.

ALAMRI, H. 2023. Emotion recognition in Arabic speech from Saudi dialect corpus using machine learning and deep learning algorithms.

CAN, Y. S., MAHESH, B. & ANDRÉ, E. 2023. Approaches, applications, and challenges in physiological emotion recognition—a tutorial overview. Proceedings of the IEEE.

CHAVHAN, Y., DHORE, M. & YESAWARE, P. 2010. Speech emotion recognition using support vector machine. International Journal of Computer Applications, 1, 6-9.

DULZ, I. 2016. The displacement of the Yezidis after the rise of ISIS in Northern Iraq. Kurdish Studies, 4, 131-147.

EL AYADI, M., KAMEL, M. S. & KARRAY, F. 2011. Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern recognition, 44, 572-587.

EL SEKNEDY, M. & FAWZI, S. Arabic english speech emotion recognition system. 2023 20th Learning and Technology Conference (L&T), 2023. IEEE, 167-170.

GARCIA-CUESTA, E., SALVADOR, A. B. & PÃEZ, D. G. 2024. EmoMatchSpanishDB: study of speech emotion recognition machine learning models in a new Spanish elicited database. Multimedia Tools and Applications, 83, 13093-13112.

GUROWIEC, I. & NISSIM, N. 2024. Speech emotion recognition systems and their security aspects. Artificial Intelligence Review, 57, 1-45.

HAZMOUNE, S. & BOUGAMOUZA, F. 2024. Using transformers for multimodal emotion recognition: Taxonomies and state of the art review. Engineering Applications of Artificial Intelligence, 133, 108339.

JIANG, D.-N., LU, L., ZHANG, H.-J., TAO, J.-H. & CAI, L.-H. Music type classification by spectral contrast feature. Proceedings. IEEE international conference on multimedia and expo, 2002. IEEE, 113-116.

JOY, J., KANNAN, A., RAM, S. & RAMA, S. 2020. Speech emotion recognition using neural network and MLP classifier. Ijesc, 2020, 25170-25172.

KASURIYA, S., THEERAMUNKONG, T., WUTIWIWATCHAI, C. & SUKHUMMEK, P. 2019. Developing a Thai emotional speech corpus from Lakorn (EMOLA). Language Resources and Evaluation, 53, 17-55.

KURDA, Z. A. 2022. The Kurdish Issue: A Transnational Political Conflict. The European Union and the Kurdish Issue: The EU as a Rational and Normative Actor. Springer.

MARTINEZ-LUCAS, L., LIN, W.-C. & BUSSO, C. 2024. Analyzing continuous-time and sentence-level annotations for speech emotion recognition. IEEE Transactions on Affective Computing.

MCFEE, B., RAFFEL, C., LIANG, D., ELLIS, D. P., MCVICAR, M., BATTENBERG, E. & NIETO, O. librosa: Audio and music signal analysis in python. SciPy, 2015. 18-24.

MEFTAH, A., QAMHAN, M., ALOTAIBI, Y. A. & ZAKARIAH, M. 2020. Arabic speech emotion recognition using KNN and KSUEmotions corpus. International Journal of Simulation--Systems, Science & Technology, 21, 1-5.

PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R. & DUBOURG, V. 2011. Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.

PULATOV, I., OTENIYAZOV, R., MAKHMUDOV, F. & CHO, Y.-I. 2023. Enhancing speech emotion recognition using dual feature extraction encoders. Sensors, 23, 6640.

SAJJAD, M. & KWON, S. 2020. Clustering-based speech emotion recognition by incorporating learned features and deep BiLSTM. IEEE access, 8, 79861-79875.

RAMAKRISHNAN, S. & EL EMARY, I. M. 2013. Speech emotion recognition approaches in human computer interaction. Telecommunication Systems, 52, 1467-1478.

SHAHIN, I., ALOMARI, O. A., NASSIF, A. B., AFYOUNI, I., HASHEM, I. A. & ELNAGAR, A. 2023. An efficient feature selection method for arabic and english speech emotion recognition using Grey Wolf Optimizer. Applied Acoustics, 205, 109279.

SHEYHOLISLAMI, J. 2017. Language status and party politics in Kurdistan-Iraq: The case of Badini and Hawrami varieties. Zazaki–yesterday, today and tomorrow. Survival and standardization of a threatened language. Dieter Halwachs: Grazer Plurlingualismus Studien (GPS 04). Graz: GLM.

SCHOLL, W. 2013. The socio-emotional basis of human interaction and communication: How we construct our social world. Social Science Information, 52, 3-33.

SWAIN, M., ROUTRAY, A. & KABISATPATHY, P. 2018. Databases, features and classifiers for speech emotion recognition: a review. *International Journal of Speech Technology,* 21**,** 93-120.

TUSHNET, M. V., CHEN, A. K. & BLOCHER, J. 2020. *Free speech beyond words: the surprising reach of the First Amendment*, NYU Press.

Koolagudi, S.G., Maity, S., Kumar, V.A., Chakrabarti, S. and Rao, K.S., 2009. IITKGP-SESC: speech database for emotion analysis. In Contemporary Computing: Second International Conference, IC3 2009, Noida, India, August 17-19, 2009. Proceedings 2 (pp. 485-492). Springer Berlin Heidelberg.

Costantini, G., Iaderola, I., Paoloni, A. and Todisco, M., 2014. EMOVO corpus: an Italian emotional speech database. In Proceedings of the ninth international conference on language resources and evaluation (LREC'14) (pp. 3501-3504). European Language Resources Association (ELRA).

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W.F. and Weiss, B., 2005, September. A database of German emotional speech. In Interspeech (Vol. 5, pp. 1517-1520).

Jackson, P. and Haq, S., 2014. Surrey audio-visual expressed emotion (savee) database. University of Surrey: Guildford, UK.

Engberg, I.S. and Hansen, A.V., 1996. Documentation of the Emotional Speech Data Base, DES.

RUDAW MEDIA NETWORK is a major media broadcaster in the Kurdistan Region, Iraq. Founded: May 29,2013. [Online]. Available: https://www.youtube.com/@RudawKurdish/vides