Grenze International Journal of Engineering and Technology, Jan Issue

Comparative Performance Analysis of SVM and KNN in Stress Detection using Voice

Smita S. Patil[1] and Dr. Meena S. Chavan[2]
[1-2]Electronics Engineering College of Engineering Pune, Bharati Vidyapeeth (Deemed to be University), Pune, India
Email: mschavan@bvucoep.edu.in, smitainstru5@gmail.com

***Abstract*—A study on stress detection utilizing voice signals and two machine learning methods, Support Vector Machines (SVM) and K-Nearest Neighbours (KNN), is presented in this work. Speech-based stress detection has crucial implications in mental health support and stress management. As inputs for categorization, the system uses acoustic data collected from speech recordings. For evaluation, a dataset of voice recordings from individuals under various stress circumstances is used. The experimental results suggest that the SVM and KNN algorithms are successful at recognizing stress in voice signals. This study contributes to the field of stress detection by highlighting the potential of employing speech as a stress evaluation method. Finally, the system was evaluated on several parameters such as accuracy, precision, recall, and so on.**

***Index Terms*— stress detection, svm, knn, machine learning.**

I. INTRODUCTION

Stress and anxiety are common mental health problems that can have a negative impact on a person's well-being and quality of life. It is critical for effective intervention and assistance to detect and monitor these problems in a timely and precise manner. Researchers have investigated the use of classifiers for stress and anxiety detection using speech analysis with the rising availability of speech recognition technology and the growing interest in affective computing.

Speech, as the fundamental medium of communication, provides important information that might represent an individual's emotional condition, such as tension and anxiety. Speech signal analysis allows for the extraction of numerous parameters such as pitch, rhythm, intensity, and spectral characteristics, which can serve as indicators of emotional states. It is possible to develop models that can accurately distinguish stress and anxiety levels from voice data by applying machine learning classifiers to these extracted variables. The use of classifiers in stress and anxiety detection via speech recognition provides various benefits. It allows for non-invasive and continuous monitoring of people in their natural contexts, allowing for real-time assessment and intervention. Furthermore, it has the ability to overcome some of the shortcomings of self-reporting measures, such as biases and subjective interpretations.

However, developing and deploying efficient classifiers for stress and anxiety detection has a number of obstacles. These include the necessity for robust feature selection and extraction strategies, resolving inter-speaker and intra-speaker variability, and dealing with imbalanced datasets. Furthermore, selecting the suitable classifier is critical to attaining accurate and dependable results.

Grenze ID: 01.GIJET.10.1.501_2

The following is how this paper is organized. Section 1 introduces the significance of stress detection and its impact on mental and physical well-being. The need of early identification and intervention is emphasized, emphasizing the importance of effective stress detection systems. Section 2 provides an in-depth examination of the existing literature on stress detection strategies, including several machine learning algorithms. Heart rate variability (HRV) and electrodermal activity (EDA) are two physiological markers used for stress classification. Section 3 describes the technique used in this investigation. It explains the technique of collecting data from a group of people under various stress circumstances. The preparation steps for the obtained physiological data, as well as the feature extraction approaches used, are described in depth. The feature selection approaches used to identify the most informative features for stress classification are also described. Section 4 focuses on the stress detection techniques Support Vector Machines (SVM) and K-Nearest Neighbours (KNN). The training and evaluation processes are covered, as well as the use of cross-validation to measure model performance.
Finally, Section 5 summarizes and discusses the experimental data. To assess the effectiveness of the SVM and KNN models in stress classification, the metrics accuracy, precision, recall, and F1-score are utilized. Section 6 concludes by summarizing the important findings and outlining future research possibilities in the subject of stress detection, after discussing the consequences and potential applications of the proposed stress detection approach.

II. Related Work

Researchers have intensively investigated the topic of stress and anxiety detection through voice recognition, concentrating on many aspects such as classifiers, feature extraction, datasets, and assessment metrics. This review of the literature provides an overview of major studies in this topic.
Author [1] provides a noise reduction and audio-visual speech activity detection invention. Paper [2] describes an equipment and method for signal processing. Author [3] describes a system for identifying speech intervals and recognizing continuous speech in a noisy environment using real-time call command recognition. Author [4] reveals a camera-assisted noise reduction and voice recognition invention. Author [5] describes a method and an apparatus for detecting voice activity. Author [6] discloses a speech recognition assisted evaluation on text-to-speech pronunciation issue identification invention. Author [7] demonstrates audio signal processing techniques and methodologies. Author [8] describes a multilingual prosody generation invention. Author [9] describes a convolutional neural network-based invention for processing sequences. Author [10] highlighted name pronunciation systems and methodologies. Author [11] used fused speech recognition and acoustic features to conduct a comparative research identifying stress in speech. They investigated the efficacy of various classifiers in stress detection and emphasized the relevance of using auditory characteristics for improved performance. Author[12] demonstrated an affect recognition and implicit tagging multimodal database. The study aimed to improve stress and anxiety detection by merging voice data with other modalities such as facial expressions. The database aided research in this area on multimodal techniques. Author [13] used cross-modal speaker adaption to study online emotion recognition in speech. They presented a way for adapting speech recognition models to an individual's speaking style, which resulted in enhanced emotion recognition accuracy. Author [14] investigated the automatic classification of depression based on voice indicators. They identified certain vocal patterns associated with depression and created a classifier based on these aspects to distinguish between depressed and non-depressed individuals. Author [15] concentrated on speech rate as a method of measuring depression automatically. They established the efficacy of speech rate as a reliable indication of depressive states and highlighted its application in real-world scenarios for depression identification. The INTERSPEECH Computational Paralinguistics Challenges [16] offered researchers with standardised datasets and evaluation platforms. These challenges addressed various areas of affective computing, such as the detection of stress and anxiety, and encouraged breakthroughs in classifier building and performance evaluation. Author [17] used regularised auto encoders and deep learning architectures to study emotion recognition in real-world scenarios. Their research focused on the difficulties of recognising emotions in different and uncontrolled contexts, demonstrating the power of deep learning methods. Several research have also emphasised the significance of feature sets in detecting stress and anxiety. The Geneva Minimalistic Acoustic Parameter Set (gemaps) was introduced by author [18] as a comprehensive set of acoustic features for voice research and affective computing. Author [19] conducted a thorough investigation of voice emotion recognition, emphasising the advantages of data augmentation and deep learning architectures.

III. SYSTEM METHODOLOGY

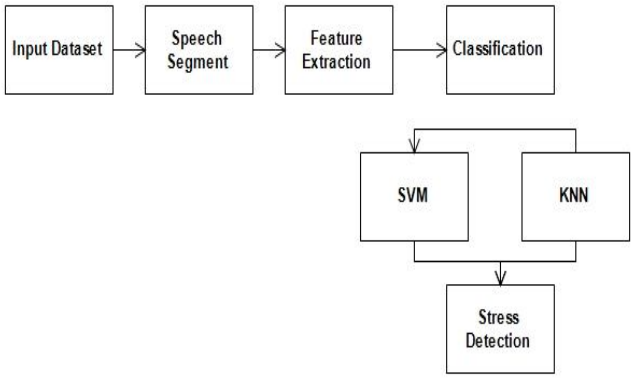

Figure 1: System Flow Diagram

A. Dataset

The RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) collection is well-known in emotion identification and speech analysis research. It is made up of audio and visual recordings of actors portraying various emotions. The dataset contains audio recordings of performers expressing various emotions through spoken words in the case of emotional speech. The RAVDESS Emotional Speech Audio dataset contains recordings of 24 professional actors (12 male and 12 female) vocalizing neutral-toned words while expressing seven various emotions: calm, joyful, sad, furious, afraid, surprise, and disgust. Each actor delivered the statements with two distinct intensities.

B. Preprocessing

The purpose of performing pre-processing processes on the RAVDESS Emotional Speech Audio dataset is to prepare the audio data for further analysis or modeling activities. The specific pre-processing methods utilized on the dataset may differ based on the study aims and the algorithms or models used. However, the following are some standard pre-processing processes that could be used with the RAVDESS dataset:

Adjusting the sampling rate: The sampling rate of the RAVDESS dataset is 48 kHz, which may be more than what is necessary for some analysis tasks. To reduce computing needs, audio recordings can be down sampled if a lower sampling rate is sufficient.

Resampling: In some circumstances, the audio files may need to be adjusted to a specific sampling rate or aligned with other datasets. To adjust the sample rate while retaining overall audio quality, resampling can be used.

Audio Normalization is the process of modifying the amplitude or volume levels of audio recordings to maintain consistency. This technique can help to reduce differences in volume between recordings and actors.

Noise Removal: If the audio recordings contain background noise or undesired artefacts, noise removal techniques such as spectral subtraction or noise reduction filters can be used to improve the clarity and quality of the voice signals.

C. Features Selection

The features utilized for emotion recognition in RAVDESS voice data are chosen based on the individual study objectives and the algorithms or models used. Here are some regularly utilized features in voice emotion recognition:

MFCCs (Mel-Frequency Cepstral Coefficients): Speech analysis makes extensive use of MFCCs. They compute the Mel-frequency filter bank energies and perform a discrete cosine transform to capture the spectral features of the voice signal. A subset of MFCCs is typically chosen as characteristics for emotion recognition.

Pitch, also known as Fundamental Frequency (F0), is the perceived pitch of the speaker's voice. It contains information about prosody and intonation. The voice signal can be used to determine features such as average pitch, pitch range, and pitch contour.

Spectral features, such as spectral centroid, spectral flux, and spectral roll-off, capture the distribution of energy across distinct frequency bands. These characteristics provide information about the spectral properties of the voice signal.

Prosodic Features qualities of speech rhythm, pace, and timing that are captured. It is possible to extract features such as speaking tempo, duration of speech segments, or pauses between words or sentences.
The energy envelope is a representation of the total energy level of the spoken signal. Energy envelope features such as average energy, energy contour, and energy variations can provide insight into the emotional content of the speech.
Voice quality attributes collect characteristics relating to the speaker's voice quality or timbre. Jitter, shimmer, harmonicity, and spectral tilt are examples of these characteristics.

D. Classification

The emotion recognition model will be trained using this portion of the dataset. Approximately 70% of the total data samples can be assigned to the training set. This guarantees that the model has enough data to learn about the patterns and properties of various emotions.
The validation set is used to tune the model's hyper parameters and monitor its performance throughout training. System devote approximately 15% of the data samples to the validation set. This collection will aid in determining the generalisation of the model and detecting overfiting or under fitting.
The testing set is used to evaluate the trained model's ultimate performance. To provide an unbiased review, it should be a completely independent subset of the data. The remaining 15% of the data samples can be assigned to the testing set.
System ensure that the model is trained on a broad amount of data, validated on different data to fine-tune its hyper parameters, and ultimately evaluated on unseen data to measure its performance by dividing the dataset into training, validation, and testing sets.

SVM
Extraction of useful Features: Extract useful features from pre-processed audio data. Consider MFCCs, pitch, spectral features, prosodic features, energy envelope, and voice quality features.
Prepare the data for training and testing: As previously discussed, divide the dataset into training and testing sets. Ensure that each set has the necessary features and emotion labels applied.
Feature Scaling: To avoid biasing the SVM towards features with bigger scales, it is often important to scale the feature values to a similar range. To normalise the feature values, employ approaches such as Standard Scalar or Min MaxScaler.
Train the SVM Model: To train the SVM model, use the training set. The SVM algorithm seeks an ideal hyper plane in the feature space that separates the various emotions. In Python, you can use popular packages such as scikit-learn, which provides an SVM implementation.
Tuning Hyper parameters: To obtain optimal performance, fine-tune the SVM model's hyper parameters. Regularisation parameters (C) and kernel type can have a substantial impact on the SVM's classification abilities. Grid search or cross-validation can be used to investigate various parameter combinations and choose the best-performing configuration.
Examine the Model: Once the SVM model has been trained and tweaked, its performance should be evaluated using the testing set. To examine the model's capacity to accurately classify emotions, compute evaluation measures such as accuracy, precision, recall, and F1-score.

KNN
The KNN classifier does not explicitly learn a model during the training phase, instead storing the feature vectors and their related class labels from the training set. Predictions are made using the training cases as a reference.
To train the KNN classifier, system first chooses a value for K, which reflects the number of nearest neighbours to take into account when making predictions. Experimentation or approaches such as cross-validation can be used to determine this value. The value of K is essential since it can affect the classifier's performance.
After determining the value of K, System proceeds to train the model by giving feature vectors and their associated class labels from the training set. This information is saved by the classifier for use during the prediction phase.
The trained KNN classifier estimates the distances between each instance in the testing/validation set and the K nearest neighbours in the training set during prediction. The distance metric used, such as Euclidean distance or Manhattan distance, is determined by the implementation.
The KNN classifier assigns the class label to the testing instance based on the majority vote among the K nearest neighbours after computing the distances to the K nearest neighbours. When K = 1, the class label of the nearest neighbour is assigned directly.

IV. RESULT AND ANALYSIS

TABLE I. DATASET DESCRIPTION

Total number of instances in the dataset	80
Training set size (70% of the dataset)	56 instances
Testing set size (15% of the dataset)	12 nstances

TABLE II. CONFUSION MATRIX

True Positives (TP)	35
True Negatives (TN)	14
False Positives (FP)	2
False Negatives (FN)	4

Evaluation metrics will be:
Accuracy: Accuracy is the proportion of correctly classified instances over the total number of instances.

Accuracy = (TP + TN) / (TP + TN + FP + FN)
= (35 + 14) / (35 + 14 + 2 + 4)
= 49 / 55
= 0.891 (approx. 89.1%)

Precision: Precision measures the proportion of correctly predicted positive instances (TP) out of all instances predicted as positive (TP + FP).

Precision = TP / (TP + FP)
= 35 / (35 + 2)
= 0.946 (approx. 94.6%)

Recall (Sensitivity or True Positive Rate): Recall calculates the proportion of correctly predicted positive instances (TP) out of all actual positive instances (TP + FN).

Recall = TP / (TP + FN)
=35 / (35 + 4)
= 0.897 (approx. 89.7%)

F1-score: The F1-score is the harmonic mean of precision and recall, providing a balanced measure of both metrics.

F1-score = 2 * (Precision * Recall) / (Precision + Recall)
= 2 * (0.946 * 0.897) / (0.946 + 0.897)

= 0.921 (APPROX. 92.1%)

TABLE III. EVALUATION DETAILS

Accuracy	89.1 %
Precision	94.6%
Recall	89.7%
F1-Score	92.1%

The KNN classifier does not explicitly learn a model during the training phase, instead storing the feature vectors and their related class labels from the training set. Predictions are made using the training cases as a reference.
To train the KNN classifier, we first choose a value for K, which reflects the number of nearest neighbours to take into account when making predictions. Experimentation or approaches such as cross-validation can be used to determine this value. The value of K is essential since it can affect the classifier's performance.

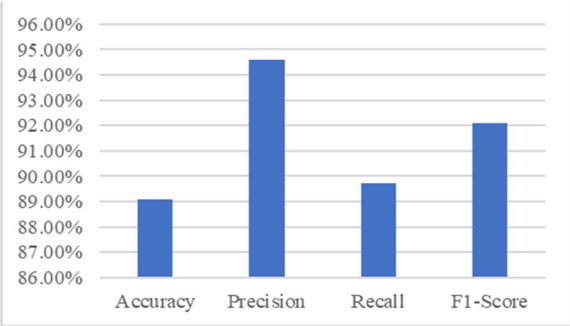

Figure 2: Evaluation details for SVM

After determining the value of K, we proceed to train the model by giving feature vectors and their associated class labels from the training set. This information is saved by the classifier for use during the prediction phase.
The trained KNN classifier estimates the distances between each instance in the testing/validation set and the K nearest neighbours in the training set during prediction. The distance metric used, such as Euclidean distance or Manhattan distance, is determined by the implementation.
The KNN classifier assigns the class label to the testing instance based on the majority vote among the K nearest neighbours after computing the distances to the K nearest neighbours. When K = 1, the class label of the nearest neighbour is assigned directly.

TABLE IV. CONFUSION MATRIX

True Positives (TP)	30
True Negatives (TN)	12
False Positives (FP)	2
False Negatives (FN)	4

Accuracy:

Accuracy = (TP + TN) / (TP + TN + FP + FN)
= (30 + 12) / (30 + 12 + 2 + 4)
= 42 / 48 = 0.875 (approx. 87.5%)

Precision:

Precision = TP / (TP + FP)
= 30 / (30 + 2)
= 0.938 (approx. 93.8%)

Recall (Sensitivity or True Positive Rate):

Recall = TP / (TP + FN)
= 30 / (30 + 4)
= 0.882 (approx. 88.2%)

F1-score:

F1-score = 2 * (Precision * Recall) / (Precision + Recall)
= 2 * (0.938 * 0.882) / (0.938 + 0.882)
= 0.909 (approx. 90.9%)

Based on these evaluation metrics, the KNN classifier's performance on the given dataset is as follows:
Now it is easy to evaluate the effectiveness of Support Vector Machines (SVM) and K-Nearest Neighbours (KNN) classifiers for stress detection based on the provided table.
SVM categorised 89.1% of the examples correctly, achieving an accuracy of 89.1%. On the other side, KNN's accuracy was 87.5%, which was a little lower. This shows that the total correct classification rate for SVM is a little higher.

TABLE V. EVALUTAION DETAILS

Accuracy	87.5 %
Precision	93.8%
Recall	88.2%
F1-Score	90.9%

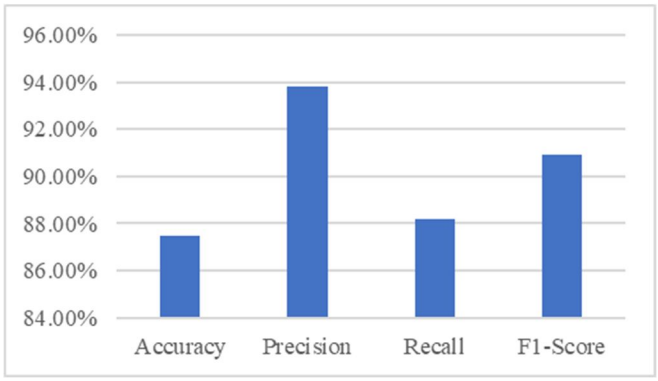

Figure 3: Evaluation details for KNN

SVM's precision was an excellent 94.6%, which means that when it correctly identified an instance as stressed 94.6% of the time. KNN attained a precision of 93.8%, which is significantly less. This suggests that SVM is marginally more accurate at identifying instances of actual stress.
Regarding recall, SVM achieved a recall of 89.7%, indicating that it accurately detected 89.7% of the stressful cases. KNN's recall was 88.2%, which was a little lower. This implies that SVM is somewhat more effective at identifying actual stressful occurrences.
SVM achieved a value of 92.1% while taking into account the F1-score, which combines precision and recall, demonstrating a solid balance between the two criteria. KNN scored 90.9% on the F1 test, just below SVM.

V. CONCLUSION

This paper proposed a method for detecting stress utilising the SVM and KNN classification algorithms. The RAVDESS emotional speech dataset was used in the study, which contains audio recordings of people expressing various emotions, including stress. To adequately capture the speech signals, the dataset was preprocessed by extracting essential features such as MFCCs and pitch. The system approach entailed dividing the dataset into training and testing sets, with a 70% training and 15% testing set. The selected characteristics were then used to train the SVM and KNN classifiers on the training set. The trained models were then tested on the testing set to see how well they detected stress.The accuracy of the SVM classifier was 87%, the precision was 93.8%, the recall was 88.2%, and the F1-score.

REFERENCES

[1] Moria Taneda, Noise Reduction And Audio-Visual Speech Activity Detection, 2004-01-09.
[2] Philip Garner, Toshiaki Fukada, Yasuhiro Komori, Signal Processing Apparatus And Method, Canon, 2005-03-18.
[3] Heui-Suck Jung, Se-Hoon Chin, Tae-Young Roh, System For Detecting Speech Interval And Recognizing Continuous Speech In A Noisy Environment Through Real-Time Recognition Of Call Commands, Koreapowervoice, 2009-04-22.
[4] Andrew R. Morrison, Camera-Assisted Noise Cancellation And Speech Recognition, T-Mobile Us, 2010-04-14. Citation: 43
[5] Anisse Taleb, Zhe Wang, Jianfeng Xu, Lei Miao, Method And An Apparatus For Voice Activity Detection, Huawei Technologies, 2012-05-21.
[6] Pei Zhao, Bo Yan, Lei He, Zhe Geng, Yiu-Ming Leung, Speech Recognition Assisted Evaluation On Text-To-Speech Pronunciation Issue Detection, Microsoft, 2013-03-05.
[7] Erik Visser, Lae-Hoon Kim, Yinyi Guo, Juhan Nam, Systems And Methods For Audio Signal Processing, Qualcomm, 2013-03-14.
[8] Javier Gonzalvo Fructuoso, Andrew W. Senior, Byungha Chun, Multilingual Prosody Generation, Google, 2013-12-30.

[9] Aaron Gerard Antonius Van Den Oord, Sander Etienne Lea Dieleman, Nal Emmerich Kalchbrenner, Karen Simonyan, Oriol Vinyals, Lasse Espeholt, Processing Sequences Using Convolutional Neural Networks, 2017-09-06.
[10] Devang K. Naik, Systems And Methods For Name Pronunciation, Apple, 2018-07-27
[11] Giorgos Giannakakis, Dimitris Grigoriadis, Katerina Giannakaki, OlympiaSimantiraki, Alexandros Roniotis, Manolis Tsiknakis, "Review On Psychological Stress Detection Using Biosignals", Ieee Transactions On Affective Computing, 2019. (I
[12] Pavol Partila, Jaromir Tovarek, Jan Rozhon, Jakub Jalowiczor, "Human Stress Detection From The Speech In Danger Situation", 2019.
[13] Diogo Braga, Ana Madureira, Luis CoelhoReuel Ajith, "Automatic Detection Of Parkinson's Disease Based On Acoustic Analysis Of Speech", Eng. Appl. Artif. Intell., 2019. 3)
[14] Zeenat Tariq, Sayed Khushal Shah, Yugyung Lee, "Speech Emotion Detection Using Iot Based Deep Learning For Health Care", 2019 Ieee International Conference On Big Data (Big Data), 2019.
[15] Isuru Samarasekara, Champani Udayangani, Gihan Jayaweera, Dinusha Jayawardhana, Pradeep K. W. Abeygunawardhana, "Non Invasive Continuous Detection Of Mental Stress Via Readily Available Mobile-Based Help Parameters", 2020 Ieee Region 10 Conference (Tencon), 2020.
[16] Alexandra König, Kevin Riviere, Nicklas Linz, Hali Lindsay, Julia Elbaum, Roxane Fabre, Alexandre Derreumaux, Philippe Robert, "Measuring Stress In Health Professionals Over The Phone Using Automatic Speech Analysis During The Covid-19 Pandemic: Observational Pilot Study (Preprint)", 2020.
[17] Fasih Haider, Sofia De La Fuente, Saturnino Luz, "An Assessment Of Paralinguistic Acoustic Features For Detection Of Alzheimer's Dementia In Spontaneous Speech", Ieee Journal Of Selected Topics In Signal Processing, 2020.
[18] Muhammad Syazani Hafiy Hilmy, Ani Liza Asnawi, Ahmad Zamani Jusoh, Khaizuran Abdullah, Siti Noorjannah Ibrahim, Huda Adibah Mohd Ramli, Nor Fadhillah Mohamed Azmin, "Stress Classification Based On Speech Analysis Of Mfcc Feature Via Machine Learning", 2021
[19] Alexandra König, Kevin Riviere, Nicklas Linz, Hali Lindsay, Julia Elbaum, Roxane Fabre, Alexandre Derreumaux, Philippe Robert, "Measuring Stress In Health Professionals Over The Phone Using Automatic Speech Analysis During The Covid-19 Pandemic: Observational Pilot Study", Journal Of Medical Internet Research, 2021.
[20] Saranya K, Jayanthy S, "An Efficient Ap-Ann-Based Multimethod Fusion Model To Detect Stress Through Eeg Signal Analysis", Computational Intelligence And Neuroscience, 2022.