Network Modeling Analysis in Health Informatics and Bioinformatics (2024) 13:49
https://doi.org/10.1007/s13721-024-00485-y

ORIGINAL ARTICLE

Identifying PTSD sex-based patterns through explainable artificial intelligence in biometric data

Andrea Alejandra García-Valdez[1] · Israel Román-Godínez[1] · Ricardo A. Salido-Ruiz[1] · Sulema Torres-Ramos[1]

Received: 1 March 2024 / Revised: 24 July 2024 / Accepted: 5 August 2024

Abstract
Post-Traumatic Stress Disorder (PTSD) is a mental health condition that arises from exposure to traumatic events, affecting various aspects of human well-being. The complexity and variability of symptoms pose challenges for accurate diagnosis and monitoring, exacerbated by accessibility barriers. In response, alternative methodologies leveraging biometric data have emerged, such as facial movements, speech, or voice-to-text transcriptions, and analyzing them using Explainable Artificial Intelligence (XAI) techniques. Numerous studies have explored the presence or absence of PTSD, yet few have concentrated on either explicable indicators or distinctions among these indicators, such as the patient's sex. This research used an XAI algorithm to identify patterns related to PTSD in three biometric data sets: facial movements, speech, and voice-to-text transcriptions. Such biometric data are grouped by sex. Utilizing the DAIC-WOZ database, this study involves feature selection and characterization. Experiment configurations addressed participant segmentation, feature reduction, and standardization. Training phases employed machine-learning classification algorithms with their corresponding performance evaluation. The interpretability stage explored the relationship between input features and class output. The findings reveal that among the three biometric data sets evaluated in this work (facial movements, speech, and voice-to-text transcriptions), speech characterization is the most effective in identifying PTSD indicators, suggesting a uniform speech pattern associated with breathy and tense voice and weak phonation in PTSD patients. Sex-specific analysis enhances prediction performance, revealing distinctions in the associated speech features. The Women's model prioritizes tense voice and vocal volume variations, reduced glottal closure, and interrupted phonation. Conversely, the Speech-Men model reflects reduced resonance, making the voice thinner and weaker, indicating altered vocal quality. As for facial movements, sex-specific characteristics are not evident, but some features focused on lips are associated with PTSD. Similarly, PTSD is related to alertness, determination, and anxiety in both women and men. In conclusion, using an XAI algorithm to differentiate sex-based patterns in biometric data contributes to a better understanding of PTSD indicators, offering potential advancements in personalized diagnostic strategies.

Keywords PTSD · Sex-based analysis · Explainable artificial intelligence · SHAP · Biometric data

1 Introduction

Post-traumatic stress disorder (PTSD) is a condition that can develop after experiencing a profound traumatic event (Kim et al. 2020), such as when someone's life is in danger or they experience significant harm. While the initial reports of PTSD were related to war veterans, a variety of experiences, including sexual violence, abuse, natural disasters, exposure to civil wars, accidents, and childbirth, among others, can also cause PTSD (Coventry et al. 2020). PTSD can also develop from witnessing a traumatic event.

Every year, approximately 5 out of 100 adults experience PTSD. For instance, in 2020, around 13 million individuals in the U.S. faced PTSD (US Department of Veterans Affairs 2018). Particularly, according to Paula (2007), about 8% of women and 4% of men may experience PTSD at some stage in their lifetime.

Diagnosing PTSD can be challenging due to individual differences and symptom variability. Mental health specialists use various assessment tools, including medical

✉ Sulema Torres-Ramos
sulema.torres@academicos.udg.mx

1 División de Tecnologías para la Integración Ciber-Humana, CUCEI-Universidad de Guadalajara, Blvd. Marcelino García Barragán No. 1451, 44430 Guadalajara, Jalisco, Mexico

examinations, health assessments, the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V) (American Psychiatric Association 2022), and symptom questionnaires. According to DSM-V, four symptom criteria for PTSD apply to adults, adolescents, and children older than six. These symptoms must persist for at least one month and involve re-experiencing symptoms, avoidance, arousal and reactivity symptoms, and cognition and mood symptoms.

Existing approaches for diagnosing PTSD could be affected by several access impediments such as financial factors, cultural prejudice, unfavorable attitudes toward therapy, resource limitations, and the episodic character of psychiatric conditions (Low et al. 2020) such as PTSD. Additionally, diagnostic accuracy is affected by the subjectivity of evaluation processes, which depend on the patient's acceptability, authenticity, and the treating medical staff's expertise (Muhorakeye and Biracyaza 2021). As a result, alternative methodologies for PTSD diagnosis have emerged utilizing data from biological markers, biosignals, neuroimaging, and biometric data such as facial expressions, speech, and voice-to-text transcriptions, among others. These emerging methodologies aim to reduce subjectivity in the diagnosis process.

Within this framework, the analysis of biometric data from facial movements, speech, and voice-to-text transcriptions, combined with machine learning tools, offers alternatives for early detection and understanding of PTSD. Among the state-of-the-art in the field, the review conducted by Othmani et al. (2023) details available databases and techniques, based on video sensors and EEG signals, for the assessment and detection of PTSD. The authors emphasize the importance of proposing new resolutions in telemedicine to improve the medical diagnosis of mental health conditions such as PTSD, through quantitative proposals based on automatic and less biased systems.

Regarding the above, most studies have utilized automatic feature extraction from several biometric data such as facial movements, speech, and voice-to-text transcriptions, to categorize the presence or absence of PTSD (Gupta et al. 2022; Sawalha et al. 2022; Scherer et al. 2013). However, among previous works, only a few have focused on identifying patterns leading to PTSD detection without falling into subjectivity and diagnostic biases.

For instance, the authors of Balbin et al. (2017) analyzed parameters like heart rate, skin conductance, and facial movements to assist healthcare professionals in evaluating patients at risk of developing PTSD. To do so, the authors used the Facial Action Units coding system (also known as AUs) to assess the potential development of the PTSD condition. However, the authors do not present an evaluation using PTSD-diagnosed patients but only patients considered at risk. Moreover, they mention that the limited amount of subjects used for the experiment may affect the risk estimation.

Other studies have focused on speech alterations in patients diagnosed with PTSD. For example, the authors of Marmar et al. (2019) identified traits contributing to outstanding performance in PTSD detection, noting slower speech and less intonation variability in individuals diagnosed with PTSD. However, such research did not evaluate the role of sex as an important variable in the detection.

Similarly, there is a study conducted in Scherer et al. (2015), where a decrease in the space between vowels was observed, indicating that participants with PTSD and depression exhibit more tense vocal characteristics. Despite its encouraging results, this work does not consider the variability of speech characteristics related solely to PTSD but to another condition, depression.

In the analysis of the voice-to-text transcription, specifically in the classification of PTSD versus Control patients. On the one hand, the authors of Sawalha et al. (2022) conducted sentiment analysis obtaining scores related to emotional polarity and intensity. However, their machine learning classification method, Gradient Boosting, aims only to define the presence or absence of PTSD without explaining the classification criteria. On the other hand, in He et al. (2017), the authors worked with trauma survivor autobiographies and applied bag-of-words, achieving a sensitivity of 0.75. However, its analysis was limited by the lack of confirmation of PTSD diagnosis in the study subjects and variability in the processes of acquiring autobiographies.

According to the research performed in Schultebraucks et al. (2022), aiming to extract features from facial expressions, movement parameters, prosody, and natural language to classify participants between PTSD and Major Depressive Disorder (MDD) using deep learning algorithms, they achieved an Area Under Curve (AUC–ROC) of 0.9 detecting both conditions, giving more weight to natural language features concerning the PTSD class. While the results are encouraging, the study does not perform an analysis using sex as a possible discriminant variable.

Under the consideration of sex as a crucial factor in PTSD analysis, previous studies confirm this observation and indicate variations in sex-based assessments of psychiatric disorders. For instance, in Stratou et al. (2015) the authors proposed using a sex-differentiated approach, mentioning a significant improvement in the detection of non-verbal signs linked to PTSD. In García-Valdez et al. (2023), differences were identified in the performance of decision tree models, highlighting sex-affected speech features. However, these studies do not consider other biometric data and only use decision trees as classification algorithms.

Lately, there has been an increasing interest in understanding the decision process of several black-box machine-learning algorithms. So, several Explainable Artificial Intelligence techniques (XAI) have been developed allowing us to understand how and why a particular decision is reached, which is crucial in applications where transparency and interpretability are essential.

In the context of PTSD research, where precision and comprehension are paramount, XAI ensures that the identified indicators are not only accurate but also comprehensible to researchers and clinicians. This transparency aids in building trust in the AI-based diagnostic tools and facilitates a deeper understanding of how specific features contribute to PTSD detection.

Studies like those conducted in Schultebraucks et al. (2022) and Kathan et al. (2023) have employed XAI algorithms. The authors of Schultebraucks et al. (2022) found that features derived from voice-to-text transcriptions are most correlated with PTSD patients. In the case of Kathan et al. (2023), a relation was identified with speech features, achieving an AUC of 82% using SVM. However, these studies did not account for the participants' sex and the exploration of multiple classification algorithms to identify the one with the highest performance.

In this regard, multiple studies have focused on diagnosing the presence or absence of PTSD using machine learning tools. However, few publications focus on finding indicators related to PTSD based on the analysis of facial movements, speech, and voice-to-text transcriptions, considering sex as a grouping variable and using an XAI algorithm for explicability purposes. This leads us to the following research question: Are there sex-affected patterns in biometric data: facial movements, speech, or voice-to-text transcriptions related to PTSD when using explainable artificial intelligence? Therefore, the present work aims to automatically identify sex-related patterns associated with PTSD indicators through the analysis of biometric data: facial movements, speech, or voice-to-text transcriptions.

2 Materials and methods

Figure 1 illustrates the proposed methodology consisting of five steps: PTSD dataset, feature selection and characterization, experimental dataset construction, ML modeling, and model explainability. A detailed explanation of each step is provided.

2.1 PTSD database

To conduct our experiment, we selected the DAIC-WOZ database. The database, titled "The Distress Analysis Interview Corpus - Wizard of Oz (DAIC-WOZ)", was created in 2014 by the University of California Gratch et al. (2014). Its objective is to assist in the diagnosis of mental health conditions, including PTSD.

The DAIC-WOZ database contains 189 participants aged between 18 and 70, all native English speakers. The DAIC-WOZ protocol involved participants interacting with an avatar named Ellie during interviews. The individuals controlling Ellie made decisions regarding appropriate questions and responses, aiming to establish a connection with the participants and prompt emotional and neutral responses. Throughout this interaction, participants were recorded on video and audio. Additionally, the dialogue between Ellie and the participants was transcribed from the audio, resulting in text data.

Furthermore, the database incorporates scores from a standardized self-assessment questionnaire known as PCL-C, based on the DSM-V criteria for PTSD diagnosis. Questionnaire scores range from 17 to 85, reflecting the severity of PTSD. The first range (17–29) indicates little to no severity, the second range (30–44) suggests moderate to moderately high severity, and the last range (45–85) indicates high severity and a positive PTSD diagnosis.

Upon exploring the data, we observed that it does not include interview videos but rather feature vectors derived from facial movements. However, both the audio recordings of the interviews and the computation of speech features have been included. Additionally, transcripts of the interviews are available in TXT format.

2.2 Feature selection and characterization

2.2.1 Feature selection

The speech and facial movement data were provided in the format of feature vectors. As described by Gratch et al. (2014), the DAIC-WOZ database contains data that were calculated using two processing systems. Speech data were processed using COVAREP (Degottex et al. 2014), while facial movements were characterized using the OpenFace software (Jacob and Stenger 2021).

Initially, after visual and statistical examination, feature vectors with repeated zero values across all instances were eliminated, assuming that such behavior reflects errors during data acquisition.

Subsequently, in the feature vectors provided by the database, interviewer characterization, as well as non-speech segments, were removed to focus on analyzing the voices of the participants. In contrast, the characterization of facial movement features comes from the interview video, which only shows the participant. It is important to note that this video was not provided by the authors of the database.

Regarding the facial movement features, the analysis incorporates the following action units: AU1, AU2, AU4, AU5, AU6, AU7, AU9, AU10, AU12, AU14, AU15, AU17,

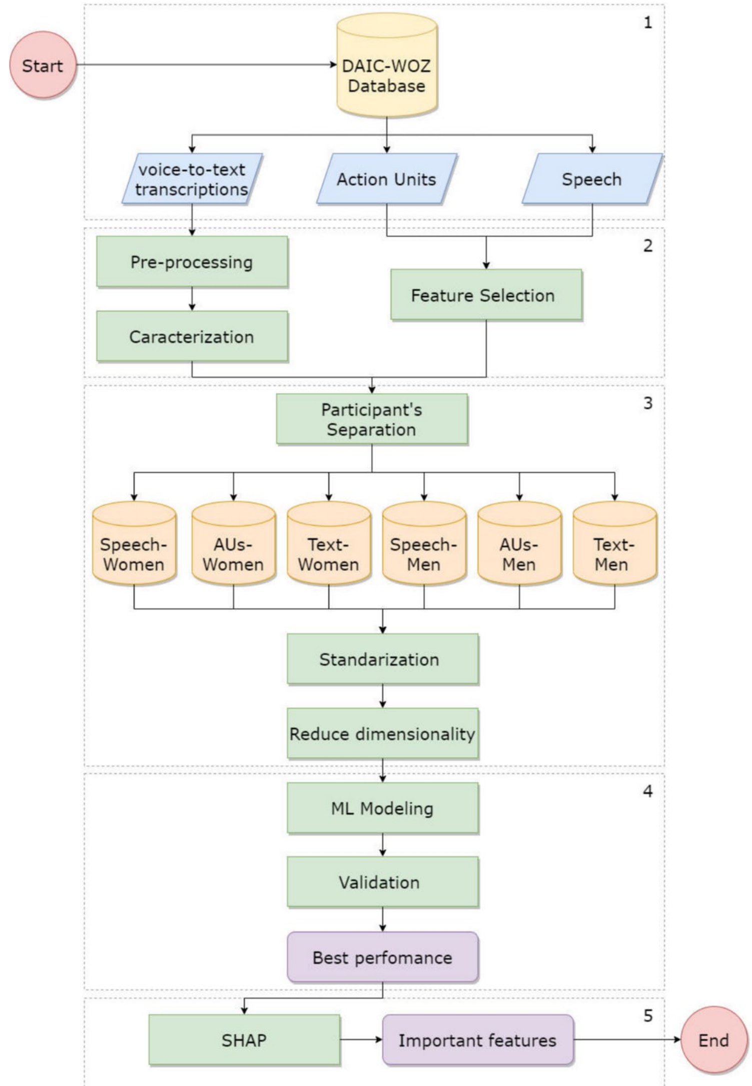


Fig. 1 Five-step based methodology followed for the study: database, feature selection and characterization, experimental dataset construction, ML modeling and model explainability

AU20, AU23, AU25, AU26, AU28, and AU45. Each action unit represents a specific facial movement as outlined in the Facial Action Coding System (FACS) (Jacob and Stenger 2021). This analytical approach involves the consideration of eighteen distinct features.

The speech features provided by the DAIC-WOZ dataset are as follows: fundamental frequency (F0), maxima dispersion quotient (MDQ), the first two harmonics of the differentiated glottal source spectrum (H1, H2), voicing (VUV), normalized amplitude quotient (NAQ), quasi-open quotient (QOQ), Mel Frequency Cepstral Coefficients (MCEP 0–12), Mel cepstral coefficients (MCEP 13–24), parabolic spectral parameter (PSP), spectral slope of wavelet responses (peak/slope), shape parameter of the Liljencrants–Fant model of the glottal pulse dynamic (Rd), Rd conf, harmonic model and phase distortion mean values (HMPDM 0–24), and deviations (HMPDD 0–12), for a total of seventy-three features.

2.2.2 Characterization

A different methodology was applied to the voice-to-text transcriptions data. Initially, the transcription data was cleaned by removing text segments of Ellie's participation. This was done to focus on the participants' contributions. Additionally, text segments shorter than four characters were excluded, along with those containing punctuation, accents, and special characters. Finally, word standardization was applied to ensure that all words were in lowercase.

In the characterization, features from three groups: sentiment analysis, emotion classification, and context classification are obtained. Sentiment analysis reveals the direction of emotion expressed in a text, with values ranging from -1 for negative text to +1 for positive text. Values close to 0 describe a neutral text. These scores were obtained using the VADER package (Valence Aware Dictionary and sEntiment Reasoner) (Isnan et al. 2023), a Python tool for sentiment analysis based on the emotional polarity of a text. It provides a score indicating how positive, negative, or neutral a text is, resulting in three features.

Subsequently, emotion classification was performed, where input text was evaluated for emotions related to fear, anger, anticipation, trust, surprise, sadness, disgust, or joy. This classification was achieved using the NRCLexicon package (Zuhanda et al. 2023), which analyzes words in a text related to emotions and returns a score indicating the intensity of each emotion for every text examined, resulting in eight features.

Finally, context classification is calculated using LIWC (Linguistic Inquiry and Word Count) (Koutsoumpis et al. 2022). The calculation is performed through the LIWC package. The output includes six features related to the context in which a text can be categorized: social, affective, cognitive, perceptual, biological, and relative.

2.3 Experimental dataset construction

The construction of the experimental dataset involved organizing participants based on their PCL-C scores. The control group comprised 97 participants with no severity, while the PTSD group consisted of 55 participants with high severity, totaling 152 participants. The dataset included 60 women and 92 men. The study excluded participants with a moderate PCL score (37 participants) as interpreting such cases is complicated and may be linked to the presence of other conditions. This analysis focuses on sex differentiation; thus, six experiments were designed according to the biometric data. In all experiments, the evaluation encompassed two groups: participants diagnosed with PTSD and a control group. The first two experiments examined speech: 'Speech-Women' analyzed features of women, while 'Speech-Men' evaluated men. The next two experiments focused on Action Units (AUs): 'AUs-Women' assessed women, and 'AUs-Men' studied men. The final two experiments concentrated on features extracted from voice-to-text transcriptions: 'Text-Women' evaluated text features among women, and 'Text-Men' did the same for men.

2.3.1 Dimensionality reduction

Subsequently, a dimensionality reduction approach was employed to eliminate features displaying similar patterns and reduce computation time. This approach was applied in each of the previously described experiments. Spearman's correlation (Dancey and Reidy 2007), a non-parametric method robust against outliers, was calculated to assess the correlation between two sets of measurements taken from the same individuals. Following (Dancey and Reidy 2007), a correlation threshold of 0.7 was considered high; therefore, all features with this value were discarded.

2.3.2 Standardization

Furthermore, for the sake of comparability and consistency, normalization was applied in the experiments. In this study, the statistical technique used is z-score standardization (Smithson et al. 2024), which transforms the data to have a mean of zero and a standard deviation of one. This technique is particularly useful when dealing with data from populations with diverse distributions, ensuring that all features are in the same standardized format. Additionally, this approach is recommended for its robustness in handling outliers in the evaluated variables.

2.4 Machine learning modeling

Initially, a search for machine learning algorithms was conducted, specifically focusing on supervised classification methods. Selected methods included Decision Trees (Han et al. 2011), Support Vector Machine (Tanveer et al. 2022), Random Forests (Manzali and Elfar 2023), Artificial Neural Networks (Abdolrasol et al. 2021), and Wickramasinghe and Kalutarage (2021). The choice of these algorithms was based on their advantages.

Decision trees are widely used in machine learning due to their intuitive structure, which facilitates the interpretation of the decision-making process. Unlike other more complex algorithms, such as neural networks, they are inherently interpretable, making them ideal when transparency and explainability are prioritized. Furthermore, they are robust and relatively simple to implement, making them a preferred choice for a variety of classification and prediction problems.

Support Vector Machines are favored in machine learning for their capability to handle complex datasets with high-dimensional feature spaces efficiently. They excel at finding an optimal hyperplane to separate different classes and can handle linear and non-linear problems effectively using kernel functions.

Random Forests stand out for their robustness in combining multiple decision trees trained on random data samples. This technique reduces overfitting and provides accurate estimates even on large and complex datasets. Their ability to handle missing data without prior imputation makes them a reliable choice in various machine-learning applications. Each decision tree in the forest randomly picks a subset of features at each split, so the importance of different features can vary in each model (Borup et al. 2023). This randomness helps the model generalize better to new data by avoiding overfitting to specific features.

Artificial Neural Networks excel in the classification process due to their ability to learn hierarchical representations of data and identify complex patterns. Their flexibility allows them to adapt optimally to different datasets and specific problems, making them effective in analyzing intricate and nonlinear data, such as pattern recognition and time series prediction.

Naïve-Bayes is valued in machine learning for its simplicity and efficiency. By assuming independence between features, it efficiently calculates prior and conditional probabilities based on Bayes' theorem, making it particularly effective for handling large datasets. Its fast training speed and low computational requirements make it suitable for real-time applications or scenarios with limited computational resources. Additionally, Naïve-Bayes performs well even with limited training data, making it a robust option for various real-world applications.

The training phase of machine learning methods is conducted with hyper-parameter configurations, except for the Naïve-Bayes algorithm, where parameters were selected to optimize the learning process. Data processing was carried out using the scikit-learn machine learning library.

For Decision Trees, a binary tree approach was adopted, meaning instances are split into two distinct branches at each step of the tree-building process. A minimum of two instances in the tree leaves was set. Additionally, fewer than five instances are not allowed for splitting, and the maximum tree depth is limited to five. A learning rate threshold of 95% was applied to prevent the model from becoming overly specific. Parameters for the SVM algorithm were adjusted with a cost of 1, an RBF kernel function, a numerical tolerance of 0.0010, and a maximum of 100 iterations. Neural network parameter tuning involved 100 neurons in one hidden layer, a ReLu activation function, and optimal weights obtained through Adam optimization with a maximum of 200 iterations. Random Forests were configured with a maximum of 10 trees, and the subset split threshold was set to 5.

For the evaluation of the models, a 10-fold cross-validation approach is implemented. This method involves dividing the dataset into ten subsets and performing the evaluation process ten times. In each iteration, a different fold is used as the test set, while the remaining nine folds serve as the training set. Performance measures, including F1-score, precision, recall, specificity, and accuracy, are calculated during this evaluation process to assess the quality of the model. To comprehend these metrics, it is essential to examine the numbered equations from Eqs. (1), (2), (3), (4), and (5).

$$F1\text{-}Score = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{5}$$

2.5 Model explainability

The interpretable machine learning technique applied is known as SHAP, which stands for SHapley Additive exPlanations (Broeck et al. 2022). It proposes a method for improving the interpretability of machine learning algorithms by offering individual-level explanations based on Shapley values derived from cooperative game theory. The SHAP methodology involves applying Shapley values to assign contributions or weights to each feature based on a specific prediction outcome. This aids in identifying the most important features in the classification process. Shapley values are computed by averaging the marginal contributions of each feature, considering all possible combinations, to ensure an equitable attribution of the prediction value to each feature (Sun et al. 2023). The advantages of SHAP include the interpretability of decisions in machine learning models, uncovering potential biases, visually interpretable graphics, and result consistency. The SHAP library in Python was employed to implement this tool, obtaining SHAP values for the target class, namely, the PTSD class.

The Beeswarm plots, computed from SHAP values, are interpreted by analyzing their three dimensions: the vertical axis indicates the assessed features sorted by their SHAP value. The feature with the greater average SHAP value is at the top (more influential) and the one with the lower average SHAP value is at the bottom (less influential). The horizontal axis represents the SHAP values, where positive values influence predicting the class of interest more. In contrast, negative values suggest an influence on the prediction of the opposing class. Note that the SHAP value indicates the degree of changes in logarithmic probabilities. Finally, each dot on the graph represents an instance of the dataset with the color indicating the feature value of such an instance. The greater the feature value the redder the instance. On the other hand, the lower the feature value the bluer the dot.

3 Results

Performance metrics (AUC, Accuracy, F1-score, Precision, Recall, and Specificity) were applied using 10-fold cross-validation for six experiments: speech-women, speech-men, AUs-women, AUs-men, text-women, and text-men. Results for each experiment are observed in Tables 1, 2, 3, 4, 5, and 6, respectively. The scores represent only the classification

Table 1 Classification performance of the Speech-women models

Algorithm	Evaluation metrics					
	AUC	*Accuracy*	*F1-score*	*Precision*	*Recall*	*Specificity*
Decision Tree	0.858	0.792	0.721	0.751	0.694	0.854
Support Vector Machine	0.721	0.681	0.581	0.593	0.570	0.752
Random Forest	**0.997**	0.975	**0.967**	0.987	0.949	0.992
Neuronal Network	0.996	0.975	**0.967**	0.972	0.963	0.982
Naïve Bayes	0.761	0.689	0.632	0.585	0.686	0.691

Table 2 Classification performance of the Speech-men models

Algorithm	Evaluation metrics					
	AUC	*Accuracy*	*F1-score*	*Precision*	*Recall*	*Specificity*
Decision Tree	0.800	0.742	0.633	0.749	0.548	0.875
Support Vector Machine	0.730	0.681	0.571	0.629	0.524	0.788
Random Forest	**0.997**	0.974	**0.967**	0.989	0.947	0.992
Neuronal Network	0.995	0.973	0.966	0.970	0.962	0.980
Naïve Bayes	0.721	0.655	0.611	0.563	0.669	0.645

Table 3 Classification performance of the AUs-women models

Algorithm	Evaluation metrics					
	AUC	*Accuracy*	*F1-score*	*Precision*	*Recall*	*Specificity*
Decision Tree	0.623	0.597	0.490	0.579	0.425	0.741
Support Vector Machine	0.464	0.470	0.470	0.432	0.515	0.432
Random Forest	**0.865**	0.785	**0.762**	0.771	0.752	0.812
Neuronal Network	0.561	0.569	0.329	0.569	0.232	0.852
Naïve Bayes	0.609	0.587	0.444	0.577	0.361	0.778

Table 4 Classification performance of the AUs-men models

Algorithm	Evaluation metrics					
	AUC	*Accuracy*	*F1-score*	*Precision*	*Recall*	*Specificity*
Decision Tree	0.600	0.647	0.121	0.574	0.067	0.971
Support Vector Machine	0.474	0.479	0.478	0.372	0.665	0.374
Random Forest	**0.845**	0.784	**0.656**	0.764	0.575	0.900
Neuronal Network	0.564	0.641	0.008	0.485	0.004	0.997
Naïve Bayes	0.584	0.631	0.178	0.443	0.111	0.921

Table 5 Classification performance of the Text-women models

Algorithm	Evaluation metrics					
	AUC	*Accuracy*	*F1-score*	*Precision*	*Recall*	*Specificity*
Decision Tree	0.552	0.552	0.602	0.221	0.500	0.142
Support Vector Machine	0.495	0.495	0.539	0.329	0.391	0.283
Random Forest	0.742	0.539	0.571	0.382	0.447	0.333
Neuronal Network	0.742	0.549	0.596	0.291	0.483	0.208
Naïve Bayes	0.742	0.561	0.571	0.426	0.455	0.400

Table 6 Classification performance of the Text-men models.

Algorithm	Evaluation metrics					
	AUC	*Accuracy*	*F1-score*	*Precision*	*Recall*	*Specificity*
Decision Tree	0.525	0.650	0.057	0.346	0.031	0.969
Support Vector Machine	0.489	0.576	0.272	0.327	0.233	0.704
Random Forest	0.518	0.610	0.233	0.353	0.174	0.872
Red Neuronal	0.532	0.655	0.036	0.365	0.019	0.994
Naïve Bayes	0.536	0.643	0.111	0.366	0.065	0.982

of the PTSD class; the best performances are highlighted in bold.

In Tables 1 and 2, it can be seen that the models evaluating speech features using Random Forest and Neural Network show the highest metric performance. Particularly, the F1-Score reaches a value of 0.967 using Random Forest in both speech-women and speech-men. Furthermore, the AUC in both experiments is equal to 0.997.

Similarly, concerning AUs features, Tables 3 and 4 indicate that when employing the Random Forest algorithm, the evaluation metrics show improvement. Specifically, in the AUs-Women experiment, an AUC of 0.865 and an F1-score of 0.762 are achieved. In contrast, in the AUs-Men experiment, an AUC of 0.845 and an F1-score of 0.656 are obtained.

In the case of the voice-to-text transcription datasets, Tables 5 and 6 show that the evaluated metrics did not reach values greater than 60%, so their models were not considered accurate. Based on these performances, it was decided to analyze only those models generated with random forests because they exhibit higher performance in all experiments. Additionally, experiments related to voice-to-text transcription were not explored in detail due to their low performance.

Therefore, the next step focuses on applying the SHAP algorithm to Random Forests due to its superior performance among all Machine Learning algorithms. The Beeswarm plots help visualize the most important features in predicting a particular class. In this case, the assessment focuses on the PTSD class, aiming to identify patterns within the PTSD group.

Given the inherent randomness in the feature selection process of Random Forests, a specialized methodology was developed to mitigate this variability. The SHAP algorithm was employed, and the experiment was conducted iteratively over ten runs. Each run produced a matrix storing SHAP values for the features. The study aimed to identify features that consistently ranked highly across different runs by averaging these SHAP values across all iterations. This iterative approach allows for the identification of features that repeatedly emerge as significant contributors to predicting the PTSD class. By focusing on these frequently occurring features, the analysis seeks to uncover robust patterns that enhance the model's predictive capabilities and interpretability.

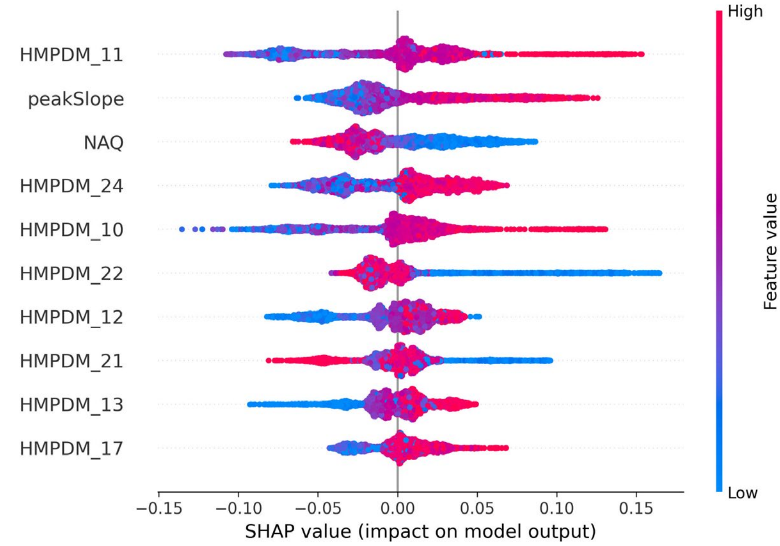
Firstly, interpretability analysis was conducted for Speech-Women. Figure 2a shows the Beeswarm plot with the ten most important features obtained from the average SHAP values (Beeswarm plots for each iteration are available in Supplementary Data File 1). The HMPDM (11, 24, 10, 22, 12, 21, 13, 17), peakSlope, and NAQ or normalized amplitude quotient features stand out. Features supporting a PTSD class prediction include high values (red) patterns, such as peakSlope, and values of HMPDM (24, 10, 13, 17). However, the NAQ feature and values of HMPDM (22, 21), with low values (blue), positively contribute to the PTSD classification process.

Subsequently, the ten most important features of the Speech-Men experiment are shown in Fig. 2b (Beeswarm plots for each iteration are available in Supplementary Data File 2). The obtained features are HMPDM (11, 20, 15, 9, 10, 22, 17, 14), PSP and H1H2. Features supporting a PTSD class prediction include high values (red) patterns, such as values HMPDM (11) and H1H2. However, the values of HMPDM (15, 9, 22) and PSP, with low values (blue), positively contribute to the PTSD classification process.

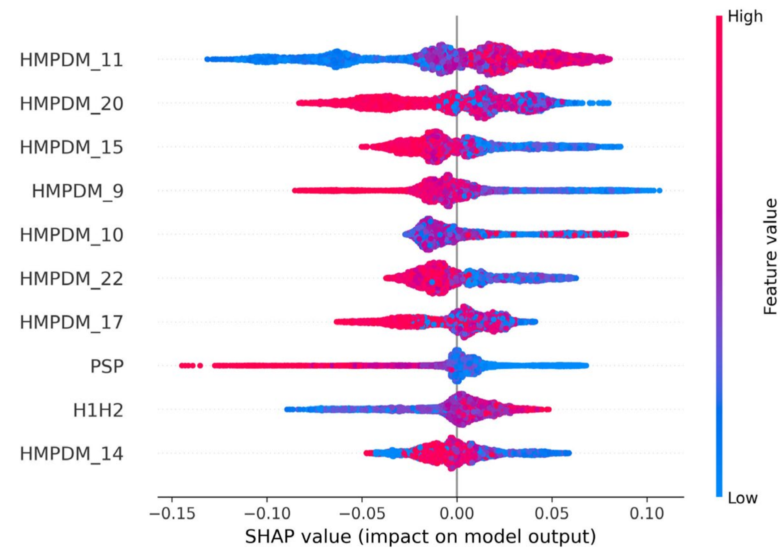
Figure 3a displays the ten most important features obtained from the average SHAP values for the AUs-Women experiment (Beeswarm plots for each iteration are available in Supplementary Data File 3). These features are displayed on the left vertical axis: Lip-Corner-Puller-c, Lip-Corner-Depressor-c, Upper-Lid-Raiser-r, Chin-Raiser-r, Lip-Tightener-c, Blink-c, Outer-Brow-Raiser-r, Lip-Corner-Depressor-r, Brow-Lowerer-r, and Lip-Stretcher-r. Features supporting a PTSD class prediction include high values (red) patterns, such as Upper-Lid-Raiser-r, Outer-Brow-Raiser-r, Brow-Lowerer-r, and Lip-Stretcher-r. However, the Chin-Raiser, with low values (blue), positively contributes to the PTSD classification process.

Regarding the AUs-Men experiment, Fig. 3b displays the ten most important features obtained: Lip-Corner-Puller-c, Blink-c, Chin-Raiser-r, Lips-Parts-r, Outer-Brow-Raiser-r, Lip-Suck-c, Brow-Lowerer-c, Brow-Lowerer-r, Lip-Corner-Depressor-r, and Lip-Corner-Depressor-c. Beeswarm plots for each iteration of the AUs-Men experiment are available in Supplementary Data File 4.

Fig. 2 Beeswarm plots for Speech experiments



(a) The average SHAP values plot applied to the Post-Traumatic Stress Disorder class, assessing the importance offeatures in the Speech-Women experiment



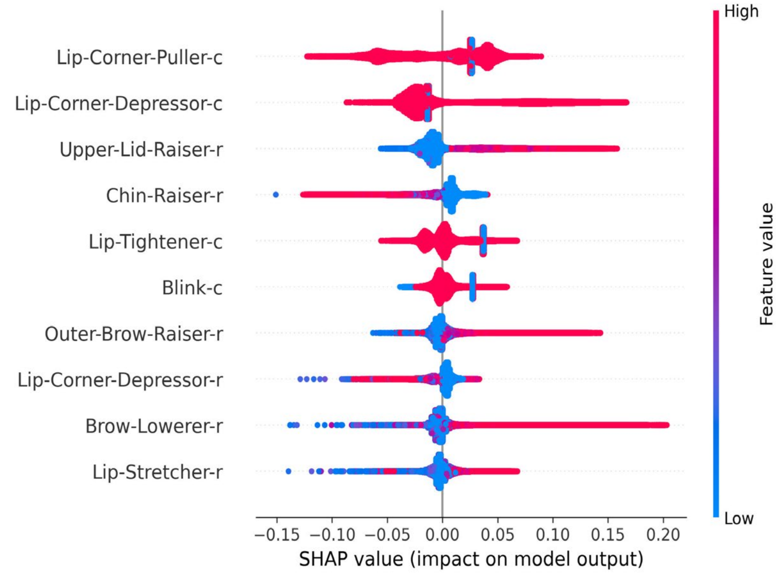
(b) The average SHAP values plot applied to the Post-Traumatic Stress Disorder class, assessing the importance offeatures in the Speech-Men experiment

Table 7 presents the most relevant features related to speech and action units (AUs) for both women and men. For speech, the most significant features belong to the HMPDM. However, some features differ between men and women. For instance, features peakSlope and NAQ are related to the Speech-Women model, whereas PSP and H1H2 are associated with the Speech-Men model. In contrast, facial movements show shared features like Lip-Corner-Puller, Lip-Corner-Depressor, Chin-Raiser, Blink, Outer-Brow-Raiser, and Brow-Lowerer among both sexes.

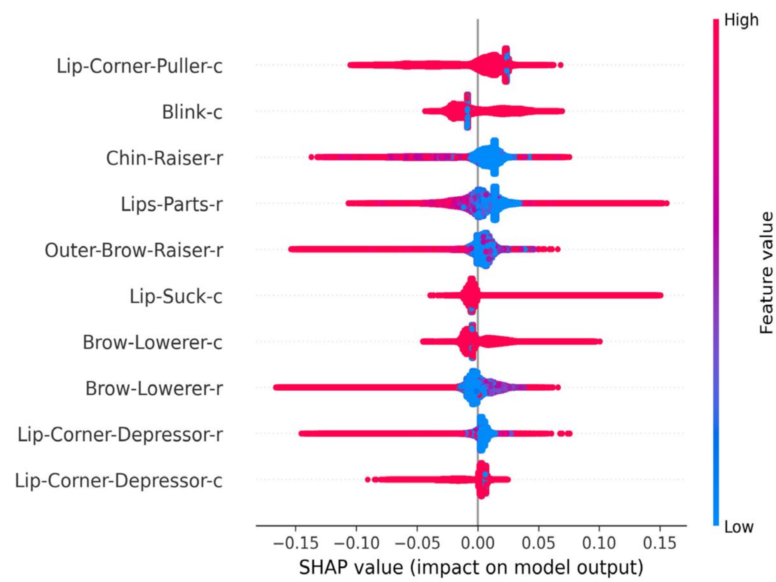
4 Discussion

In the analysis of the three biometric datasets, it is notable that the performance of models based on voice-to-text transcriptions was lower compared to those based on speech and action units (AUs) features. This finding contrasts with previous studies by Schultebraucks et al. (2022) and Sawalha et al. (2022), where metrics above 90% AUC were reported. The discrepancy in voice-to-text transcription

Fig. 3 Beeswarm plots for Action Units experiments



(a) The average SHAP values plot applied to the Post-Traumatic Stress Disorder class, assessing the importance offeatures in the AUs-Women experiment



(b) The average SHAP values plot applied to the Post-Traumatic Stress Disorder class, assessing the importance offeatures in the AUs-Men experiment

performance raises questions about the applied text preprocessing methodology. The use of tokenization and cleaning requires a detailed review in future research to determine its impact on results and facilitate more accurate comparisons with existing literature.

On the other hand, models based on speech features showed the highest performance in this study, particularly those trained using Random Forests. This suggests that analyzing the voice signals of PTSD patients may be crucial in identifying indicators for this mental health condition.

Examining the features identified by the SHAP algorithm concerning speech models, spectral features emerged as the predominant feature group of significance. Regarding spectral features, the values of the Harmonic Mean Phase Distortion Model (HMPDM) demonstrated a significant impact on the model's ability to discern patterns related to PTSD. Therefore, a detailed analysis is required to fully understand

Table 7 Key features of PTSD

Speech		AUs	
Women	Men	Women	Men
HMPDM-11	HMPDM-11	Lip-Corner-Puller-c	Lip-Corner-Puller-c
peakSlope	HMPDM-20	Lip-Corner-Depressor-c	Blink-c
NAQ	HMPDM-15	Upper-Lid-Raiser-r	Chin-Raiser-r
HMPDM-24	HMPDM-9	Chin-Raiser-r	Lips-Parts-r
HMPDM-10	HMPDM-10	Lip-Tightener-c	Outer-Brow-Raiser-r
HMPDM-22	HMPDM-22	Blink-c	Lip-Suck-c
HMPDM-12	HMPDM-17	Outer-Brow-Raiser-r	Brow-Lowerer-c
HMPDM-21	PSP	Lip-Corner-Depressor-r	Brow-Lowerer-r
HMPDM-13	H1H2	Brow-Lowerer-r	Lip-Corner-Depressor-r
HMPDM-17	HMPDM-14	Lip-Stretcher-r	Lip-Corner-Depressor-c

the information provided by these features. This set of values was created as a proposal to evaluate the quality of human voice (Degottex et al. 2014) applied in voice recognition algorithms. According to the description of this set of features, they are linked to a speech signal that is uniformly consistent, in other words, without abrupt changes.

The authors of Marmar et al. (2019) noted that the speech of PTSD patients tends to be more monotonous and less activated, thus confirming our findings and allowing us to define that the speech of PTSD patients tends to exhibit this uniformity. However, something that Marmar et al. (2019) and other authors such as Schultebraucks et al. (2022); Gupta et al. (2022) did not consider is the possibility of finding differences when analyzing the speech of men and women separately, which is reflected in our results.

Based on the model that considered only women highlighted values from the HMPDM features. Additionally, the metrics NAQ and peakSlope were exclusively associated with models focusing on women (see Table 7), allowing us to analyze the spectral tilt of the glottal source, derived from wave analysis. According to previous studies, the peakSlope is associated with the presence of tense voice in speech analysis, as well as vocal volume variation (Kuang and Liberman 2018; Szklanny and Tylki-Szymańska 2018). Similarly, a feature known as NAQ was identified. This measure quantifies the fraction of time during phonation when the glottis is considered closed, and several studies recognize it as a relevant indicator to assess the quality of human voice (Kania et al. 2004). The detailed analysis of the Beeswarm provided by the SHAP algorithm (see Fig. 2a) reveals that low NAQ values positively support the prediction of the PTSD class. These low values suggest a reduction in glottal closure, indicating interrupted or weak phonation due to inadequate closure of the vocal cords (Alku et al. 2002).

In contrast, in the Speech-Men experiment, the model identifies features that are linked to the HMPDM values. In particular, two features are exclusively associated with the men's experiment (see Table 7). H1H2 is a feature that indicates the first two harmonics of the distinct glottal source spectrum (Barche 2024). This feature is used to analyze changes in phonation and is related to the glottal opening phase (Kreiman et al. 2007). In addition, another feature related to the Speech-Men model is the Parabolic Spectral Parameter (PSP). This feature is a voice analysis parameter that quantifies the waveform of the glottal volume flow (Kiran Reddy et al. 2021), providing information about the symmetry and slope of the waveform to characterize the glottal source in more detail. In Fig. 2b, it is observed that low PSP values are associated with the PTSD class. A low PSP in the voice suggests reduced resonance, making the voice sound thinner and weaker. This indicates difficulties in the production and control of phonation, reflecting potential muscle (Barche 2024). Consequently, vocal quality is altered, affecting speech intelligibility and communication effectiveness.

According to the FACS system (Jacob and Stenger 2021), there are established relationships between action units and emotions. For instance, the Lip-Corner-Puller elevates the corners of the lips, commonly associated with expressions of happiness and joy. In contrast, the Lip-Corner-Depressor lowers the corners of the lips, typically indicating expressions of sadness and disgust. The Upper-Lid-Raiser lifts the upper eyelids, which can signal surprise or fear in response to something unexpected (Kiran Reddy et al. 2021). Meanwhile, the Chin-Raiser lifts the chin, typically interpreted as a sign of disgust or disdain towards something or someone. The Lip-Tightener tightens the lips, reflecting effort or intense concentration. Lastly, Blink, the act of blinking, can indicate fatigue, nervousness, or simply an involuntary bodily response (Bologna et al. 2024). In summary, action units that focus on facial movements, particularly those of the lips, are often associated with the PTSD class in both men and women models (see Table 7).

Regarding the identified action units in AUs-Women, the most relevant features include Lip-Corner-Puller, Lip-Corner-Depressor, Upper-Lid-Raiser, Chin-Raiser, Lip-Tightener, Blink, Outer-Brow-Raiser, Brow-Lowerer, and Lip-Stretcher. The combination of Lip-Corner-Puller and Lip-Corner-Depressor suggests a forced smile (Kiran Reddy et al. 2021), possibly masking genuine emotions. The presence of Upper-Lid-Raiser, Chin-Raiser, and Lip-Tightener reflects alertness and determination, common responses to traumatic memories. Additionally, Blink indicates anxiety, while Outer-Brow-Raiser and Brow-Lowerer express surprise, anger, or sadness, particularly when combined with Lip-Corner Depressor (Takemoto et al. 2023). Lip-Stretcher alongside repeated Lip-Corner-Puller emphasizes emotional tension.

In contrast, the identified action units in AUs-Men highlight Lip-Corner-Puller, Blink, Chin-Raiser, Lips-Parts, Outer-Brow-Raiser, Lip-Suck, Brow-Lowerer, and Lip-Corner-Depressor. Similar to women, Lip-Corner-Puller and Lip-Corner-Depressor indicate a forced smile (Ghamen and Caplier 2011). Blink, associated with anxiety, along with Chin-Raiser and Lips-Parts, reflects determination and surprise. Additionally, Outer-Brow-Raiser suggests surprise or interest, while Lip-Suck and Brow-Lowerer convey tension, insecurity, and anger (Gavrilescu and Vizireanu 2019).

5 Conclusion

This investigation reveals the association between features extracted from facial movements, speech, and voice-to-text transcriptions with indicators of PTSD, presenting significant findings for early detection and a deeper understanding of this mental health condition.

The results indicate that, among the three biometric data, speech characterization stands out as the most effective in identifying PTSD indicators. Specifically, the findings suggest that spectral features carry more weight in predicting the PTSD class, with emphasis on the group of HMPDM values. This leads to the conclusion that the speech of PTSD patients tends to be uniform when treated as a signal and is associated with breathy or weak phonation.

The obtained information reflects a subtle improvement in performance when making a sex-based separation among the evaluated patients, as well as differences in associated speech features. Among these differences, the Speech-Women model is related to the presence of tense voice in speech analysis and variations in vocal volume. They also signify reduced glottal closure and interrupted or weak phonation due to inadequate closure of the vocal cords. In contrast, the Speech-Men model reflects changes in phonation and the opening phase of the glottis. Additionally, it is associated with reduced resonance, which makes the voice sound thinner and weaker, indicating altered vocal quality.

Therefore, it is recommended to consider sex when analyzing the speech of patients diagnosed with PTSD in future studies. This consideration forms the groundwork for devising personalized strategies in the detection of this condition.

Regarding the AUs features, the results suggest that action units focusing on lip movements show an association with the PTSD class. Additionally, the AUs-Women model shows associations with a forced smile, which may potentially mask genuine emotions. Other actions suggest alertness and determination, while blinking signifies anxiety. Expressions of surprise or sadness are evident, particularly in combination with other facial gestures. Similarly, AUs-Men model displays forced smiles and expressions of anxiety. Their facial movements convey determination, surprise, interest, as well as tension and anger.

The study's limitations are acknowledged, such as the lack of validation with other databases to test stability, the need for analysis with a larger number of participants, and the absence of collaboration with medical professionals. These limitations create opportunities for additional research that can enhance and expand upon these findings.

In future work, the addition of new sources of features could provide a more comprehensive insight. Additionally, it is suggested to review and adjust the processing strategies applied, considering alternative approaches to achieve better performance. The insights gleaned from this study offer a robust foundation for potential applications in the analysis of various mental health disorders beyond the scope of the current investigation. This extends the reach of our findings, opening avenues for further exploration and research in the realm of mental health analysis.

Finally, this study not only advances the identification of PTSD indicators but also aims to contribute to the development of personalized and quantitative diagnostic tools in the medical field. By supporting a broader analysis of PTSD and striving for faster and more measurable detection methods, this paper enhances the understanding and management of PTSD. As a result, society could benefit from improved diagnostic and treatment approaches, ultimately enhancing the quality of life for those affected by PTSD.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13721-024-00485-y.

Acknowledgements Not applicable.

Author contributions AAGV, STR, IRG, RASR conceptualized the study, analyzed and interpreted data; AAGV drafted the work; STR, IRG, RARS substantively revised the manuscript. All authors read and approved the final manuscript.

Funding The author, Andrea Alejandra García-Valdez, thanks CONAHCyT for the scholarship granted to pursue her postgraduate studies (CVU 1184323).

Data availability The data that support the findings of this study are openly available at the following URL: https://dcapswoz.ict.usc.edu/ (Gratch et al. 2014).

Declarations

Conflict of interest The authors declare no Conflict of interest.

References

Abdolrasol MG, Hussain SS, Ustun TS, Sarker MR, Hannan MA, Mohamed R, Ali JA, Mekhilef S, Milad A (2021) Artificial neural networks based optimization techniques: a review. Electronics 10(21):2689

Alku P, BäckströMT Vilkman E (2002) Normalized amplitude quotient for parametrization of the glottal flow. J Acoust Soc Am 112(2):701–710

American Psychiatric Association (2022) Diagnostic and statistical manual of mental disorders: DSM-5-TR. American Psychiatric Association Publishing, Washington

Balbin JR, Pinugu JNJ, Basco AJS, Cabanada MB, Gonzales PMV, Marasigan JCC, Sejera MM (2017) Development of scientific system for assessment of post-traumatic stress disorder patients using physiological sensors and feature extraction for emotional state analysis. In: 2017IEEE 9th International conference on humanoid, nanotechnology, information technology, communication and control, environment and management (HNICEM). IEEE, pp 1–6

Barche P (2024) Acoustic analysis of voice disorders from clinical perspective. PhD thesis, International Institute of Information Technology Hyderabad

Bologna M, Paparella G, Valls-Solé J, Hallett M, Berardelli A (2024) Neural control of blinking. Clin Neurophysiol. https://doi.org/10.1016/j.clinph.2024.02.023

Borup D, Christensen BJ, Mühlbach NS, Nielsen MS (2023) Targeting predictors in random forest regression. Int J Forecast 39(2):841–868. https://doi.org/10.1016/j.ijforecast.2022.02.010

Broeck G, Lykov A, Schleich M, Suciu D (2022) On the tractability of SHAP explanations. J Artif Intell Res 74:851–886

Coventry PA, Meader N, Melton H, Temple M, Dale H, Wright K, Cloitre M, Karatzias T, Bisson J, Roberts NP et al (2020) Psychological and pharmacological interventions for posttraumatic stress disorder and comorbid mental health problems following complex traumatic events: systematic review and component network meta-analysis. PLoS Med 17(8):1003262

Dancey CP, Reidy J (2007) Statistics without maths for psychology. Pearson education, Harlow

Degottex G, Kane J, Drugman T, Raitio T, Scherer S (2014) Covarep—a collaborative voice analysis repository for speech technologies. In: 2014 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 960–964

García-Valdez AA, Román-Godínez I, Salido-Ruiz RA, Torres-Ramos S (2023) Sex-based speech pattern recognition for post-traumatic stress disorder. In: Congreso Nacional de Ingeniería Biomédica. Springer, Berlin, pp 192–200

Gavrilescu M, Vizireanu N (2019) Predicting depression, anxiety, and stress levels from videos using the facial action coding system. Sensors 19(17):3693

Ghamen K, Caplier A (2011) Positive and negative expressions classification using the belief theory. Int J Tomogr Stat 17(S11):72–87

Gratch J, Artstein R, Lucas GM, Stratou G, Scherer S, Nazarian A, Wood R, Boberg J, DeVault D, Marsella S et al (2014) The distress analysis interview corpus of human and computer interviews. In: LREC. Reykjavik, pp 3123–3128

Gupta S, Goel L, Singh A, Agarwal AK, Singh RK (2022) TOXGB: teamwork optimization based XGBoost model for early identification of post-traumatic stress disorder. Cogn Neurodyn 16(4):833–846

Han J, Kamber M, Pei J (2011) Data mining: concepts and techniques, 3rd edn. Morgan Kaufmann Publishers Inc., San Francisco

He Q, Veldkamp BP, Glas CA, Vries T (2017) Automated assessment of patients' self-narratives for posttraumatic stress disorder screening using natural language processing and text mining. Assessment 24(2):157–172

Isnan M, Elwirehardja GN, Pardamean B (2023) Sentiment analysis for TikTok review using VADER sentiment and SVM model. Procedia Comput Sci 227:168–175

Jacob GM, Stenger B (2021) Facial action unit detection with transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 7680–7689

Kania RE, Hans S, Hartl DM, Clement P, Crevier-Buchman L, Brasnu DF (2004) Variability of electroglottographic glottal closed quotients: necessity of standardization to obtain normative values. Arch Otolaryngol Head Neck Surg 130(3):349–352

Kathan A, Triantafyllopoulos A, Amiriparian S, Milkus S, Gebhard A, Hohmann J, Muderlak P, Schottdorf J, Schuller BW, Musil R (2023) The effect of clinical intervention on the speech of individuals with PTSD: features and recognition performances. In: Proc. INTERSPEECH 2023, pp 4139–4143. https://doi.org/10.21437/Interspeech.2023-1668

Kim TD, Lee S, Yoon S (2020) Inflammation in post-traumatic stress disorder (PTSD): a review of potential correlates of PTSD with a neurological perspective. Antioxidants 9(2):107

Kiran Reddy M, Helkkula P, Madhu Keerthana Y, Kaitue K, Minkkinen M, Tolppanen H, Nieminen T, Alku P (2021) The automatic detection of heart failure using speech signals. Comput Speech Lang 69:101205. https://doi.org/10.1016/j.csl.2021.101205

Koutsoumpis A, Oostrom JK, Holtrop D, Van Breda W, Ghassemi S, Vries RE (2022) The kernel of truth in text-based personality assessment: a meta-analysis of the relations between the big five and the linguistic inquiry and word count (liwc). Psychol Bull 148(11–12):843

Kreiman J, Gerratt B, Antoñanzas-Barroso N (2007) Measures of the glottal source spectrum. J Speech Lang Hear Res 50:595–610. https://doi.org/10.1044/1092-4388(2007/042)

Kuang J, Liberman M (2018) Integrating voice quality cues in the pitch perception of speech and non-speech utterances. Front Psychol 9:2147

Low DM, Bentley KH, Ghosh SS (2020) Automated assessment of psychiatric disorders using speech: a systematic review. Laryngosc Investig Otolaryngol 5(1):96–116

Manzali Y, Elfar M (2023) Random forest pruning techniques: a recent review. Oper Res Forum 4:43

Marmar CR, Brown AD, Qian M, Laska E, Siegel C, Li M, Abu-Amara D, Tsiartas A, Richey C, Smith J et al (2019) Speech-based markers for posttraumatic stress disorder in us veterans. Depress Anxiety 36(7):607–616

Muhorakeye O, Biracyaza E (2021) Exploring barriers to mental health services utilization at Kabutare district hospital of Rwanda: perspectives from patients. Front Psychol 12:638377

Othmani A, Brahem B, Haddou Y et al (2023) Machine learning-based approaches for post-traumatic stress disorder diagnosis using video and EEG sensors: a review. IEEE Sens J. https://doi.org/10.1109/JSEN.2023.3312172

Paula S (2007) Epidemiology and impact of PTSD. https://www.ptsd.va.gov/professional/treat/essentials/epidemiology.asp. Online; Accessed 10 Dec 2023

Sawalha J, Yousefnezhad M, Shah Z, Brown MR, Greenshaw AJ, Greiner R (2022) Detecting presence of PTSD using sentiment analysis from text data. Front Psychiatry 12:811392

Scherer S, Stratou G, Gratch J, Morency L-P (2013) Investigating voice quality as a speaker-independent indicator of depression and PTSD. In: Interspeech, pp 847–851

Scherer S, Lucas GM, Gratch J, Rizzo AS, Morency L-P (2015) Self-reported symptoms of depression and PTSD are associated with reduced vowel space in screening interviews. IEEE Trans Affect Comput 7(1):59–73

Schultebraucks K, Yadav V, Shalev AY, Bonanno GA, Galatzer-Levy IR (2022) Deep learning-based classification of posttraumatic stress disorder and depression following trauma utilizing visual and auditory markers of arousal and mood. Psychol Med 52(5):957–967

Smithson CJ, Chow JK, Chang T-Y, Gauthier I (2024) Measuring object recognition ability: reliability, validity, and the aggregate z-score approach. Behav Res Methods. https://doi.org/10.3758/s13428-024-02372-w

Stratou G, Scherer S, Gratch J, Morency L-P (2015) Automatic nonverbal behavior indicators of depression and PTSD: the effect of gender. J Multimodal User Interfaces 9:17–29

Sun H, Lui S, Huang X, Sweeney J, Gong Q (2023) Effects of randomness in the development of machine learning models in neuroimaging studies of schizophrenia. Schizophr Res 252:253–261

Szklanny K, Tylki-Szymańska A (2018) Follow-up analysis of voice quality in patients with late-onset Pompe disease. Orphanet J Rare Dis 13(1):1–7

Takemoto A, Aispuriete I, Niedra L, Dreimane LF (2023) Differentiating depression using facial expressions in a virtual avatar communication system. Front Digit Health 5:1080023

Tanveer M, Rajani T, Rastogi R, Shao Y-H, Ganaie M (2022) Comprehensive review on twin support vector machines. Ann Oper Res 339:1223–1268

US Department of Veterans Affairs (2018) VA: how common is PTSD in adults? https://www.ptsd.va.gov/understand/common/common_adults.asp. Online; Accessed 10 Dec 2023

Wickramasinghe I, Kalutarage H (2021) Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. Soft Comput 25(3):2277–2293

Zuhanda MK, Syofra AHS, Mathelinea D, Gio PU, Anisa YA, Novita N (2023) Analysis of twitter user sentiment on the monkeypox virus issue using the NRC lexicon. Jurnal Mantik 6(4):3854–3860