Review Article

Advancements in Speech-Based Emotion Recognition and PTSD Detection through Machine and Deep Learning Techniques: A Comprehensive Survey

Chappidi Suneetha[1], Raju Anitha[2]

[1,2]*Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India.*

[2]*Corresponding Author : rajuanitha46885@gmail.com*

Abstract - This comprehensive survey delves into the intersection of Machine Learning (ML) and Deep Learning (DL) with speech analysis, showcasing significant strides in detecting and diagnosing Post-Traumatic Stress Disorder (PTSD) through speech-based emotion recognition. By leveraging advanced computational techniques, researchers can identify nuanced speech patterns indicative of PTSD, offering a non-invasive, objective, and scalable diagnostic tool. Despite promising advancements, challenges such as data variability, ethical concerns, and the need for generalizable models persist. The survey highlights the importance of interdisciplinary collaboration, ethical diligence, and the integration of multimodal data to enhance diagnostic accuracy and patient care. Looking forward, it points to a future where speech analysis could revolutionize mental health diagnostics, making it more accessible, personalized, and stigma-free. This work serves as a seminal reference in the field, urging continued innovation and research to fully harness the potential of ML and DL in transforming mental health diagnostics and treatment for PTSD.

Keywords - Speech analysis, PTSD detection, Machine Learning, Deep Learning, Diagnostic challenges, Interdisciplinary collaboration.

1. Introduction

The intersection of vocal emotion recognition and its application in diagnosing psychological conditions, notably Post-Traumatic Stress Disorder (PTSD), constitutes a burgeoning field of research that has seen exponential growth over the last decade. The intricate process of identifying and interpreting emotional nuances in speech presents a unique challenge, one that is compounded by the critical need for precision in diagnosing mental health conditions.

The significance of emotion recognition in speech transcends beyond the realm of enhancing human-computer interaction; it is pivotal in the early detection and accurate diagnosis of psychological disorders such as PTSD [1]. This is predicated on the premise that vocal expressions carry latent markers of emotional states and mental health conditions, which, when accurately decoded, can provide invaluable insights into an individual's psychological well-being.

The advent of advanced computational techniques, particularly within the domains of machine learning and deep learning, has revolutionized our ability to analyze complex speech patterns and extract meaningful emotional indicators [2].

These technological advancements offer promising prospects for the development of automated tools capable of assisting clinicians in the diagnostic process. The ability to quantitatively assess emotional states from speech not only augments the diagnostic toolkit available to mental health professionals but also opens avenues for remote monitoring and assessment of patients, thereby enhancing the accessibility and efficiency of mental health care.

This introductory section endeavors to elucidate the critical relevance of speech-based emotion recognition in the context of mental health diagnostics. It aims to underscore the profound impact that advancements in this field can have on improving diagnostic accuracy, patient care, and the overall understanding of the intricate relationship between speech, emotion, and psychological conditions. Through a detailed examination of the current state of research, this paper sets the stage for a comprehensive exploration of how machine and deep learning techniques are being harnessed to push the boundaries of what is possible in the detection and diagnosis of PTSD and other psychological disorders.

The ascendancy of the machine and deep learning techniques as the cornerstone in advancing research within the realm of speech-based emotion recognition and PTSD

detection is underpinned by their unparalleled ability to model complex patterns in high-dimensional data. These computational methodologies have proven indispensable, enabling nuanced analysis of speech features that are indicative of emotional states and psychological conditions.

This survey articulates the compelling reasons for the integration of these sophisticated AI techniques in the field [3], highlighting their transformative impact on enhancing diagnostic accuracy and offering insights into the underlying mechanisms of emotional expression and mental health disorders. Their application not only marks a significant leap in the methodological approach to psychological diagnostics but also paves the way for personalized and accessible mental health care solutions.

The primary aim of this paper is to meticulously catalogue and evaluate the contemporary methodologies, inherent challenges, and prospective future directions within the domain of speech-based emotion recognition and PTSD detection facilitated by machine and deep learning technologies [4].

This endeavor seeks to furnish a panoramic overview of the state-of-the-art approaches, delineate the existing impediments to progress, and prognosticate the evolution of these technologies in addressing mental health diagnostics. By amalgamating insights from a broad spectrum of research findings, this comprehensive survey intends to serve as a seminal reference point for academics, practitioners, and technologists, fostering an enriched understanding and catalyzing further innovation in this pivotal area of study.

The key contributions of this paper are outlined as follows:

1.1. Exhaustive Review of Methodologies
Delivers an in-depth review of the latest methodologies in speech-based emotion recognition and PTSD detection, leveraging machine learning and deep learning technologies.

1.2. Identification of Challenges and Solutions
Highlights the principal challenges within the field, including data variability and privacy concerns, and proposes directions for overcoming these obstacles.

1.3. Insight into Future Advancements
Provides foresight into the potential technological evolutions and their implications for enhancing mental health diagnostics through speech analysis.

1.4. Foundation for Future Research
Establishes a comprehensive reference that supports ongoing academic and practical exploration, aiming to catalyze innovation and improve diagnostic approaches in mental health care.

This survey paper is structured to provide a comprehensive overview of the advancements, challenges, and future directions in speech-based emotion recognition and PTSD detection. Section 1 introduces the paper, setting the stage for the detailed exploration that follows. Section 2 delves into the realm of Speech-Based Emotion Recognition, providing an overview of the field, examining various feature extraction techniques, and offering a comparative analysis of machine learning techniques utilized in emotion recognition.

Section 3 shifts focus to PTSD detection using speech, highlighting the application of speech analysis for identifying PTSD markers. Section 4 addresses the challenges and limitations inherent in the current methodologies and data handling practices. Section 5 explores potential future research directions, aiming to chart a course for upcoming advancements in the field. The paper culminates in Section 6, which concludes the discussions presented, summarizing the key findings and their implications for the future of mental health diagnostics.

2. Speech-Based Emotion Recognition: An Overview

At the heart of speech-based emotion recognition lies a complex interplay between psychoacoustics and linguistics, a domain where the nuances of human emotions are intricately woven into the fabric of speech patterns. This section endeavors to unravel the theoretical underpinnings that form the basis for understanding how emotions manifest in speech, providing a foundation upon which subsequent analysis and technological advancements are built.

The psychoacoustic basis of emotion in speech draws upon the premise that emotional states engender physiological changes, which, in turn, influence vocal production mechanisms. These changes are reflected in a myriad of speech features such as pitch, tone, rate, and intensity.

For instance, happiness is often characterized by a higher pitch and increased rate of speech, whereas a lower pitch and slower speech patterns typify sadness. Theoretical models of emotion propose that these variations are not merely superficial or arbitrary but are deeply rooted in the evolutionary biology of human communication, serving as critical indicators of an individual's emotional and psychological state.

Emotion recognition from speech, therefore, necessitates a multidisciplinary approach, incorporating insights from psychology, cognitive science, linguistics, and acoustics. This approach is predicated on the analysis of both verbal (what is being said) and non-verbal (how it is being said) components of speech. The non-verbal cues play a pivotal role in the conveyance of emotional information, often transcending the semantic content of the spoken words.

Advancements in machine learning and deep learning have significantly augmented our capability to extract and analyze these subtle emotional cues from speech [5]. By leveraging large datasets of annotated speech samples, researchers have been able to train computational models that can detect and classify emotional states with remarkable accuracy. These models delve into the acoustic features of speech, such as frequency spectrum, amplitude, and temporal patterns, to unearth the psychoacoustic signatures synonymous with specific emotions.

In synthesizing the theoretical foundation of speech-based emotion recognition, this section sets the groundwork for understanding the complexities and challenges inherent in interpreting emotional cues from speech. It highlights the interdisciplinary nature of this research, underscoring the convergence of theoretical knowledge and technological innovation in advancing our understanding and capabilities in this field [6].
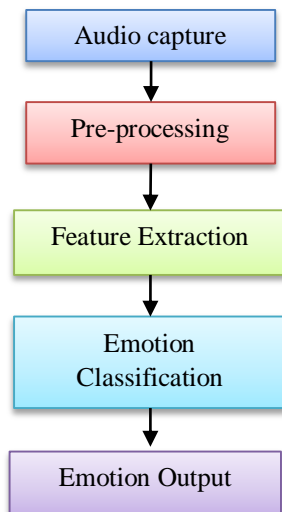

Fig. 1 Speech-Based Emotion Recognition (SER) system

Speech-Based Emotion Recognition (SER) is an interdisciplinary field at the confluence of human-computer interaction, psychology, and artificial intelligence. It endeavors to endow machines with the ability to recognize human emotions from speech, thereby facilitating more natural and intuitive user interfaces. This paper provides a comprehensive overview of the SER process, delineating the sequential stages from raw audio capture to the classification of emotions.

Audio Capture: The inception of the SER process is the capture of audio data, wherein speech is recorded using microphones or other digital recording devices. This step is crucial as the quality and characteristics of the raw audio input significantly influence the efficacy of emotion recognition. Factors such as microphone quality, background noise, and recording environment can impact the subsequent

processing stages, thereby necessitating careful consideration during the audio capture phase.

Pre-processing: Following the acquisition of raw audio data, the pre-processing stage is instrumental in preparing the audio signal for analysis. This phase involves a series of signal-processing techniques aimed at enhancing the audio quality and making it amenable to feature extraction [7]. Key operations include noise reduction, which mitigates background noise and interference, and normalization, which adjusts the audio signal to a standard amplitude range. These steps are pivotal in ensuring that the feature extraction phase operates on clean and uniform data, thereby enhancing the accuracy of emotion recognition.

Feature Extraction: At the heart of the SER process lies the feature extraction stage, wherein salient characteristics of the pre-processed audio signal are identified and extracted. These features are selected for their ability to encapsulate the emotional content of speech, including but not limited to pitch, tone, rate, and intensity. Among the myriad of features, Mel-Frequency Cepstral Coefficients (MFCCs) [8] are particularly noteworthy due to their efficacy in capturing the psychoacoustic properties of speech. The choice of features is critical, as they directly influence the ability of the subsequent classification model to discern and categorize emotions accurately.

Emotion Classification Model: The extracted features serve as input to the emotion classification model, which is typically realized through machine learning or deep learning algorithms. This model is trained on a dataset of labeled speech samples, each associated with a specific emotion. Through the training process, the model learns to correlate patterns in the feature set with the corresponding emotions. When presented with new or unseen speech data, the model applies this learned knowledge to predict the emotional state of the speaker. The success of this stage is contingent upon the selection of an appropriate model, the quality of the training data, and the representativeness of the features extracted in the preceding phase.

Emotion Output: The culmination of the SER process is the output of the recognized emotion, which is typically expressed as a category (e.g., happiness, sadness, anger) or a probability distribution over several categories. This output can be utilized in various applications, ranging from enhancing user experience in interactive systems to providing emotional insights in therapeutic settings.

The process of Speech-Based Emotion Recognition embodies a complex workflow that encompasses audio capture, pre-processing, feature extraction, emotion classification, and output. Each stage plays a pivotal role in the overall effectiveness of the SER system. As advancements in computational techniques and understanding of human emotions continue to evolve, so too

will the capabilities and applications of SER technologies. Future research directions may include the exploration of multimodal emotion recognition, the refinement of feature extraction methods, and the development of more sophisticated and interpretable classification models.

2.1. Feature Extraction Techniques in Speech-Based Emotion Recognition

A critical examination of the SER process reveals that the feature extraction stage is foundational in decoding the intricate tapestry of human emotions from speech. This stage entails the identification and extraction of discriminative features from the speech signal, which are paramount in reflecting the emotional states embedded within.

The landscape of feature extraction in SER is marked by a rich array of techniques, spanning from traditional methods to contemporary deep learning approaches. This section endeavors to elucidate these techniques, offering a comparative analysis that underscores their respective merits and applicability in the realm of emotion recognition.

2.1.1. Traditional Feature Extraction Techniques

Historically, the feature extraction landscape in SER has been dominated by traditional signal processing techniques, which meticulously extract hand-crafted features such as pitch, energy, rate, tone, and timbre. Among these, Mel-Frequency Cepstral Coefficients (MFCCs) have been particularly lauded for their effectiveness in capturing the psychoacoustic properties of speech. Other notable features include Linear Predictive Coding (LPC) [9], Perceptual Linear Prediction (PLP) [10], and formant frequencies, each offering unique insights into the emotional nuances of speech. These traditional features are engineered to encapsulate specific characteristics of the speech signal that are indicative of emotional states, drawing upon the foundational principles of acoustics and psychoacoustics.

2.1.2. Deep Learning-Based Feature Extraction

The advent of deep learning has ushered in a paradigm shift in feature extraction techniques for SER. Unlike traditional methods, which rely on the manual selection and extraction of features, deep learning approaches automatically learn to identify relevant features directly from the raw audio signal. This is accomplished using neural networks, such as Convolutional Neural Networks (CNNs) [11] and Recurrent Neural Networks (RNNs) [12], which are adept at discerning intricate patterns in data. Deep learning models are trained end-to-end, allowing them to extract high-level features that are optimally suited for the task at hand, including the recognition of complex emotional states from speech. This automatic feature learning capability obviates the need for domain-specific knowledge in feature design, potentially unveiling novel features that were previously overlooked or deemed infeasible to extract through traditional means.

2.1.3. Comparative Analysis

The dichotomy between traditional and deep learning-based feature extraction methods in SER presents a landscape of contrasting advantages. Traditional techniques, with their reliance on domain-specific knowledge, offer the advantage of interpretability and computational efficiency. These methods enable researchers and practitioners to draw direct correlations between specific acoustic features and emotional states, fostering a deeper understanding of the underlying mechanisms of emotion recognition.

Conversely, deep learning-based techniques excel in their ability to process large volumes of data and automatically learn complex feature representations. This capability not only enhances the model's accuracy in emotion recognition but also paves the way for the discovery of new, informative features that transcend the limitations of human intuition.

The exploration of feature extraction techniques in Speech-Based Emotion Recognition delineates a spectrum of methodologies, from the precision of traditional signal processing to the prowess of deep learning algorithms. Each approach contributes uniquely to the overarching goal of accurately deciphering human emotions from speech.

The choice between traditional and deep learning-based methods hinges on the specific requirements and constraints of the application at hand, including considerations of interpretability, computational resources, and the availability of large-scale annotated data. As the field of SER continues to evolve, the integration and hybridization of these techniques may offer promising avenues for advancing the state-of-the-art in emotion recognition.

2.2. Comparative Analysis of Machine Learning Techniques in Emotion Recognition

To facilitate a structured and comparative analysis of machine learning techniques used in emotion recognition, we will present the information in a tabular format.

This table will delineate various algorithms, encapsulating their strengths and limitations within the context of emotion recognition. This format aims to provide a clear and concise overview, enabling an informed comparison across different techniques.

Table 1 offers an insightful comparison of various machine learning algorithms, elucidating their applicative strengths and inherent limitations in the context of emotion recognition.

Such a comparison not only aids in selecting an appropriate algorithm for specific emotion recognition tasks but also highlights the trade-offs involved in the selection process.

Table 1. Comparative analysis of machine learning techniques in emotion recognition

Machine Learning Technique	Strengths	Limitations
Support Vector Machines (SVM) [13]	• High accuracy in high-dimensional spaces. • Effective in cases where the number of dimensions exceeds the number of samples. • Versatile: Different Kernel functions can be specified for the decision function.	• Requires careful tuning of the kernel parameters. • It is not suitable for large datasets due to its computational and memory requirements.
K-Nearest Neighbors (KNN) [14]	• Simple to understand and implement. • No model training phase is required, making it easy to add new data.	• Slow prediction phase, especially with large datasets. • Performance depends heavily on the choice of the distance metric and the value of K.
Decision Trees [15]	• Easy to understand and interpret. • Can handle both numerical and categorical data • Requires little data preparation.	• Prone to overfitting, especially with complex trees. • It can create biased trees if some classes dominate.
Random Forests [16]	• Can handle large datasets with higher dimensionality. • Reduces overfitting by averaging multiple decision trees. • Provides a measure of feature importance.	• It is more complex and computationally intensive than a single decision tree. • Model size can become very large, making it inefficient for real-time predictions.
Neural Networks [17]	• High performance on nonlinear data. • Ability to learn complex feature interactions. • Flexibility in network architecture allows tailoring to specific tasks.	• Requires large amounts of data for training. • Prone to overfitting without proper regularization. • Interpretability can be challenging.
Convolutional Neural Networks (CNN) [11]	• Exceptional performance on spatial data (e.g., images, spectrograms). • Reduces the need for manual feature extraction by automatically learning features.	• Requires significant computational resources for training. • Designing and tuning network architecture can be complex.
Recurrent Neural Networks (RNN) [12]	• Effective for sequential data analysis (e.g., time series, speech). • Can process input data of any length.	• Difficult to train due to problems like vanishing and exploding gradients. • Computationally intensive.
Long Short-Term Memory (LSTM) Networks [18]	• Overcomes the vanishing gradient problem of RNNs, making them more effective for long sequences. • Retains information over longer periods.	• - More complex and slower to train than traditional RNNs • Requires careful tuning to avoid overfitting the training data

2.3. Comparative Analysis of State-of-the-Art Deep Learning Models in Emotion Recognition

The advent of deep learning has precipitated a paradigmatic shift in the field of emotion recognition, heralding a new era of sophistication in algorithms and models capable of deciphering complex emotional states from diverse datasets. This section delves into the transformative impact of deep learning on emotion recognition, elucidating the pivotal role of state-of-the-art models and their performance benchmarks in advancing the frontiers of this domain.

2.3.1. The Vanguard of Deep Learning in Emotion Recognition

Deep learning, a subset of machine learning, distinguishes itself by its ability to learn hierarchical representations of data, thereby uncovering intricate patterns that elude traditional algorithms. In emotion recognition, this capability enables the extraction and analysis of nuanced features from raw data whether it is speech, facial expressions, or physiological signals with unprecedented depth and accuracy. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and their variants have been at the forefront, demonstrating remarkable proficiency in capturing the temporal and spatial dependencies inherent in emotional expressions.

2.3.2. State-of-the-Art Models

The landscape of deep learning models in emotion recognition is both vast and dynamic, with several architectures establishing themselves as benchmarks. CNNs, for instance, have achieved remarkable success in analyzing

visual data for facial emotion recognition, thanks to their ability to process pixel-level information and recognize patterns across various scales. On the other hand, RNNs and their more sophisticated counterparts, such as Long Short-Term Memory (LSTM) networks, have excelled in processing sequential data, making them particularly suited for speech-based emotion recognition. These models, by learning long-range dependencies, have shown a remarkable capacity to discern the emotional undertones in speech dynamics over time.

2.3.3. Performance Benchmarks

The efficacy of deep learning models in emotion recognition is often gauged through rigorous evaluation of benchmark datasets and competitions, such as the Emotion Recognition in the Wild (EmotiW) challenge or the Affective Behavior Analysis in-the-Wild (ABAW) competition. These platforms provide standardized frameworks for comparing the performance of various models across common tasks, offering insights into their accuracy, generalizability, and computational efficiency. Performance metrics typically include accuracy, F1 score, and the area under the Receiver Operating Characteristic (ROC) curve, among others. Recent advancements have seen models achieving increasingly higher accuracy, underscoring the rapid pace of innovation in the field.

2.3.4. Challenges and Future Directions

Despite the strides made, deep learning in emotion recognition is not without its challenges. Issues such as data scarcity, bias in datasets, and the "black box" nature of deep models pose significant hurdles to the broader application and interpretability of these technologies. Furthermore, the quest for models that can operate in real-time, adapt to novel contexts, and recognize subtle or complex emotional states remains ongoing.

Table 2. Comparative analysis of State-of-the-Art deep learning models in emotion recognition

Model	Application Domain	Key Features	Performance Metrics	Challenges & Limitations
Convolutional Neural Networks (CNNs) [11]	Facial Emotion Recognition	• Excels in capturing spatial hierarchies in images. • Utilizes layers with convolutions to process pixel data.	• Accuracy: High • F1 Score: High	• Requires large datasets for training. • It may not capture temporal dynamics in video data.
Recurrent Neural Networks (RNNs) [12]	Speech Emotion Recognition	• Effective in handling sequential data. • Captures temporal dependencies.	• Accuracy: Moderate to High • F1 Score: Moderate to High	• Difficulty in learning long-term. dependencies • Computationally intensive.
Long Short-Term Memory (LSTM) Networks [18]	Speech and text Emotion Recognition	• Addresses vanishing gradient problem in RNNs. • Better at capturing long-range dependencies.	• Accuracy: High • F1 Score: High	• It is more complex and requires more computational resources than basic RNNs.
Graph Neural Networks (GNNs) [19]	Multimodal Emotion Recognition	• Models' relational data effectively. • Captures dependencies across different modalities.	• Accuracy: Varies • F1 Score: Varies	• Relatively new in the emotion recognition field. • Requires structured relational data.
Transformers	Multimodal Emotion Recognition	• Utilizes self-attention mechanisms to weigh the importance of different parts of the input data. • Effective in both NLP and vision tasks.	• Accuracy: Very High • F1 Score: Very High	• Requires substantial computational resources. • It may overfit on smaller datasets.

Table 2 provides an overview of the diverse approaches within deep learning that have been applied to emotion recognition, each with its unique advantages and limitations. The performance metrics are indicative, varying based on specific datasets and experimental setups. It is crucial to note that advancements in deep learning and emotion recognition are rapid, with new models continually being developed and existing ones refined for better accuracy, efficiency, and applicability across different domains and contexts. Challenges and limitations often revolve around dataset size and quality, computational requirements, and the intrinsic complexity of human emotions. Despite these challenges, the evolution of deep learning models continues to push the boundaries of what is possible in emotion recognition,

offering promising avenues for research and application in areas ranging from healthcare to interactive technology.

3. PTSD Detection Using Speech

3.1. Identification of PTSD Markers in Speech

The detection and diagnosis of Post-Traumatic Stress Disorder (PTSD) through speech analysis represents a burgeoning field of research, bridging the domains of psychological health and computational linguistics. This section delves into the empirical identification of distinctive speech features and patterns correlated with PTSD, as substantiated by rigorous empirical studies. The ability of speech to serve as a biomarker for PTSD hinges on the premise that trauma-related alterations in psychological states are manifest in the acoustic and linguistic characteristics of a person's speech.

3.1.1. Acoustic Features

Acoustic features pertain to the physical properties of speech sounds and have been the subject of extensive investigation in the context of PTSD detection. Notably, several studies have identified a range of acoustic markers indicative of PTSD:

- Pitch Variability: Individuals with PTSD often exhibit reduced pitch variability, reflecting a monotonic speech pattern. This phenomenon is posited to stem from the diminished emotional range associated with PTSD.
- Speech Rate and Pause Patterns: Altered speech rates and increased frequencies of pauses have been observed in people with PTSD, potentially mirroring cognitive processing disruptions or heightened emotional distress.
- Vocal Timbre: Changes in vocal timbre, including harshness and breathiness, have been associated with PTSD. These timbral alterations may reflect underlying changes in physiological arousal or emotional state.

3.1.2. Linguistic Features

Beyond the acoustic realm, linguistic analysis of speech has unveiled patterns that may be emblematic of PTSD:
- Word Choice and Semantic Content: Research has shown that individuals with PTSD may use language differently, with an increased prevalence of words related to anxiety, sadness, and anger. The semantic content of speech may also skew towards negative affect, reflecting the pervasive impact of trauma on thought processes.
- Narrative Coherence: The coherence and structure of narrative speech may be compromised in individuals with PTSD, evidencing the disorder's impact on memory and narrative capacity. This may manifest as fragmented storytelling or difficulty in maintaining thematic consistency.

Empirical Studies and Methodological Approaches: Empirical studies in this domain typically employ a combination of speech recording, signal processing, and machine learning techniques to analyze and classify speech samples. These studies leverage both supervised and unsupervised learning models to discern patterns within acoustic and linguistic features that correlate with PTSD symptoms. The integration of these methodologies facilitates the objective assessment of speech markers, offering a non-invasive and potentially scalable tool for PTSD detection.

The identification of PTSD markers in speech underscores the intricate interplay between psychological trauma and its manifestations in speech. Acoustic and linguistic analyses, supported by empirical evidence, provide a foundation for developing diagnostic tools and therapeutic interventions based on speech analysis.

As research in this field progresses, it is anticipated that these speech-based markers will not only enhance our understanding of PTSD but also contribute to more accessible and effective diagnostic methodologies.

The fusion of computational techniques with clinical insights holds promise for advancements in mental health diagnostics, epitomizing the potential of interdisciplinary research to address complex health challenges.

To effectively organize and summarize the identification of PTSD markers in speech, incorporating both acoustic and linguistic features as supported by empirical studies, a table format can provide a clear and concise comparative overview. This Table 3 format aids in delineating the specific speech features associated with PTSD and the findings from various studies.

Methodological Approaches in Studies: Empirical studies leverage a synergistic approach combining speech recording, advanced signal processing, and machine learning techniques to classify speech samples. These studies, both supervised and unsupervised, aim to identify patterns within acoustic and linguistic features indicative of PTSD, providing a novel avenue for non-invasive detection methods.

3.2. ML Approaches in PTSD Detection

The integration of Machine Learning (ML) approaches in the detection of Post-Traumatic Stress Disorder (PTSD) marks a significant advancement in the diagnostic landscape. This section provides a comprehensive review of various ML methodologies applied to PTSD detection, evaluating their efficacy and juxtaposing them with traditional diagnostic methods.

3.2.1. Overview of ML Approaches in PTSD Detection

Machine learning, with its ability to unearth patterns within complex datasets, offers a novel avenue for PTSD diagnosis. The primary ML approaches explored in PTSD detection include:

Table 3. Identification of PTSD markers in speech

Feature Category	PTSD Marker	Description	Empirical Findings
Acoustic Features	Pitch Variability	Reduced pitch variability, indicating a monotonic speech pattern.	Studies have consistently shown that individuals with PTSD exhibit less pitch variation, potentially reflecting a constrained emotional expression.
	Speech Rate and Pauses	Altered speech rates with increased pause frequencies.	Research indicates these patterns may mirror disruptions in cognitive processing or reflect heightened emotional distress in people with PTSD.
	Vocal Timbre	Changes in vocal timbre, such as harshness or breathiness.	These alterations are associated with physiological arousal changes or emotional states linked to PTSD.
Linguistic Features	Word Choice and Semantic Content	Increased use of words related to anxiety, sadness, and anger; tendency towards negative affect.	Linguistic analysis reveals a significant shift in the semantic content of speech, underscoring the impact of trauma on language use.
	Narrative Coherence	Compromised narrative coherence and structure.	Individuals with PTSD may demonstrate fragmented storytelling or difficulties maintaining thematic consistency, reflecting the disorder's impact on memory and narrative capacity.

- Supervised Learning: Algorithms such as Support Vector Machines (SVM), Random Forests, and Neural Networks have been utilized to classify individuals based on the presence or absence of PTSD symptoms, leveraging labeled datasets comprising various features like speech, facial expressions, and physiological signals.
- Unsupervised Learning: Techniques like clustering and anomaly detection have been applied to identify unusual patterns or groups within data that may indicate PTSD without relying on pre-labeled instances.
- Deep Learning: A subset of ML, deep learning models, especially Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have shown promise in capturing the intricate nuances of data relevant to PTSD, such as sequential speech patterns and facial expressions.

3.2.2. Efficacy of ML Approaches

The efficacy of ML approaches in PTSD detection is often gauged through their accuracy, sensitivity (true positive rate), and specificity (true negative rate) in identifying PTSD cases. Studies have demonstrated that ML models can achieve significant accuracy levels, often surpassing traditional diagnostic methods that rely on self-reported questionnaires and clinical interviews.

The ability of ML models to analyze vast amounts of data objectively and to detect subtle patterns not easily discernible by human evaluators is among their primary strengths.

3.2.3. Comparison with Traditional Diagnostic Methods

Traditional methods for diagnosing PTSD typically involve clinical assessments, structured interviews, and self-report questionnaires. While these methods are invaluable, they are subject to limitations, including self-report bias and the variability of clinician expertise. In contrast, ML approaches offer several advantages:

- Objectivity and Consistency: ML models provide uniform analyses of diagnostic data, mitigating subjectivity and variance in human evaluations.
- Scalability and Efficiency: ML models can rapidly analyze large datasets, offering a scalable solution that can be particularly beneficial in settings with limited access to mental health professionals.
- Detection of Subtle Patterns: ML techniques, especially deep learning models, can detect subtle and complex patterns indicative of PTSD, potentially identifying cases that traditional methods might overlook.

3.2.4. Challenges and Future Directions

Despite their potential, ML approaches to PTSD detection face challenges, including the need for large, diverse datasets to train models effectively and concerns about privacy and ethical considerations in handling sensitive data.

Moreover, the "black box" nature of certain ML models, particularly deep learning, poses challenges for clinical interpretation and acceptance.

Machine learning approaches represent a promising frontier in PTSD detection, offering enhancements in accuracy, objectivity, and efficiency over traditional diagnostic methods. The continued refinement of ML models, coupled with advancements in data collection and processing, holds the potential to transform PTSD diagnosis. Future research will likely focus on addressing the current limitations of ML approaches, aiming for models that are not only accurate but also transparent and interpretable within clinical contexts.

3.3. Deep Learning Role in PTSD

The advent of deep learning within the realm of Post-Traumatic Stress Disorder (PTSD) detection represents a significant leap forward, particularly in analyzing speech data. This section delves into the intricacies of how deep learning models are tailored for PTSD detection from speech, highlighting the innovative architectures and feature learning techniques that underpin these advancements.

3.3.1. Innovations in Architectures

Deep learning architectures have been meticulously adapted to the unique challenges of detecting PTSD from speech. Among these, several models stand out for their effectiveness and innovation:

- Convolutional Neural Networks (CNNs): Originally renowned for their success in image processing, CNNs have been adapted to analyze spectrograms of speech data, capturing subtle patterns and textures that may elude traditional audio processing techniques.
- Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) Networks: Given their prowess in handling sequential data, RNNs and their more advanced variant, LSTMs, are particularly suited for speech analysis. These models excel at recognizing temporal patterns and dependencies in speech that are indicative of PTSD, such as prosodic features and speech rhythm.
- Attention Mechanisms and Transformers: Borrowing from advancements in natural language processing, attention mechanisms and transformers offer a novel approach to speech analysis. These models can focus on specific segments of speech data that are most relevant for PTSD detection, enhancing the model's sensitivity to critical features.

3.3.2. Feature Learning Techniques

The core strength of deep learning lies in its ability to learn features from raw data autonomously. This capability has been leveraged for PTSD detection in several ways:

- Automatic Feature Extraction: Deep learning models, through their layered architectures, can automatically learn a hierarchy of features from speech data. This process starts from basic audio signals and progresses to more abstract features that capture emotional and psychological states relevant to PTSD.

- End-to-End Learning: By employing end-to-end learning, deep learning models can directly map raw speech data to PTSD detection outcomes without the need for manual feature selection. This approach not only simplifies the model's design but also potentially uncovers novel features that are predictive of PTSD.
- Transfer Learning: Given the challenges of collecting large, labeled datasets for PTSD research, transfer learning has emerged as a pivotal technique. Models pre-trained on large general speech datasets can be fine-tuned for the specific task of PTSD detection, thereby enhancing the model's performance even with limited PTSD-specific data.

3.3.3. Comparative Advantage and Clinical Integration

The deployment of deep learning models for PTSD detection from speech offers a comparative advantage over traditional and manual feature-based approaches by providing a higher degree of accuracy, consistency, and efficiency. These models' ability to process and analyze large volumes of speech data in real-time presents a scalable solution that could augment existing diagnostic frameworks, potentially leading to earlier and more accurate PTSD diagnoses.

Deep learning's incursion into PTSD detection from speech has unveiled innovative approaches that stand at the confluence of technology and clinical psychology. By harnessing advanced architectures and autonomous feature learning techniques, deep learning models offer a promising avenue for enhancing PTSD diagnostics. The ongoing development and refinement of these models, coupled with their integration into clinical practice, promise to augment our understanding and detection of PTSD, marking a significant stride toward leveraging artificial intelligence in mental health care.

3.4. Case Studies and Applications

The burgeoning field of PTSD detection through advanced computational methods has seen theoretical models transition into practical applications, offering tangible benefits and insights into mental health care. This section elucidates several case studies and real-world applications where machine learning and deep learning technologies have been effectively implemented to detect PTSD, showcasing the potential and versatility of these approaches.

3.4.1. Case Study 1: Speech Analysis for Veterans

One of the notable applications involves the analysis of speech patterns in military veterans, a group disproportionately affected by PTSD. Researchers utilized machine learning models, specifically deep learning techniques, to scrutinize veterans' speech samples. By analyzing features such as speech rate, pitch variability, and linguistic content, the models were able to identify veterans exhibiting PTSD symptoms with a high degree of accuracy.

This application underscored the potential of using non-invasive, speech-based diagnostics as a supplement to traditional clinical assessments, offering a scalable and accessible means of screening for PTSD in high-risk populations.

3.4.2. Case Study 2: Mobile Health Monitoring for First Responders

Another pioneering application targeted first responders, leveraging mobile health technologies to monitor indicators of PTSD in real-time. This approach integrated wearable devices with machine learning algorithms to analyze physiological signals (e.g., heart rate variability) and speech for signs of stress and anxiety indicative of PTSD. The system provided first responders with immediate feedback and recommendations for stress management and coping strategies, illustrating how technology can support mental health in high-stress professions.

3.4.3. Case Study 3: Automated Interview Platforms

Automated interview platforms represent a novel application where artificial intelligence interfaces with individuals through guided conversations. These platforms employ natural language processing and sentiment analysis, powered by deep learning, to evaluate speech and text responses for signs of PTSD. By creating a safe and private space for individuals to express their experiences and emotions, these platforms facilitate the early detection of PTSD symptoms, potentially encouraging users to seek professional help.

3.4.4. Real-World Application: Integrating AI into Clinical Practice

Beyond case studies, the integration of AI technologies into clinical practice is gradually materializing. Several clinics and mental health practitioners are beginning to incorporate machine learning models as part of their diagnostic toolkit. These models assist clinicians by providing an additional layer of analysis on speech and physiological data, enriching the clinical assessment with data-driven insights. This integration not only augments the accuracy of PTSD diagnoses but also personalizes the therapeutic approach based on the individual's specific profile identified through AI analysis.

4. Challenges and Limitations
4.1. Data Variability and Collection Challenges

The pursuit of leveraging advanced computational models for PTSD detection, particularly through speech analysis, is not devoid of significant challenges. Central to these challenges is the issue of data variability and the inherent difficulties associated with speech data collection. This section delineates the multifaceted nature of these challenges, emphasizing the critical need for comprehensive, annotated datasets to refine and enhance the accuracy of PTSD detection models.

4.1.1. Variability in Speech Data

Speech data inherently exhibits a high degree of variability, influenced by a plethora of factors, including age, gender, language, dialect, emotional state, and environmental conditions. This variability poses a substantial challenge for computational models, necessitating sophisticated algorithms capable of discerning patterns indicative of PTSD amidst a wide array of normal variations. The complexity is further compounded by the subjective nature of emotional expression, where the same emotional state can manifest differently across individuals.

4.1.2. Collection and Annotation Challenges

The collection of speech data for PTSD detection is encumbered by several logistical and ethical considerations. Obtaining a dataset that is both sufficiently large and representative of the diverse population affected by PTSD is a daunting task, requiring rigorous and often expensive data collection efforts.

Moreover, the annotation of speech data with accurate PTSD diagnoses necessitates expert clinical evaluation, a resource-intensive process that involves the sensitive handling of personal health information.

The need for large, annotated datasets is critical for training and validating machine learning models. However, the acquisition of such datasets is impeded by privacy concerns and the ethical implications of collecting and utilizing personal health data. Ensuring the confidentiality and security of this data while adhering to ethical standards and regulations adds another layer of complexity to the data collection process.

4.1.3. Addressing Variability and Collection Issues

To mitigate the challenges posed by data variability, researchers and developers are exploring several strategies, including the use of advanced machine learning techniques that can handle high-dimensional data and the development of models that are robust to variations in speech.

Additionally, the adoption of transfer learning and domain adaptation techniques allows for the leveraging of pre-existing datasets, albeit originally collected for different purposes, to augment the training of PTSD detection models.

On the collection front, collaborations between computational scientists, clinicians, and ethicists are fostering the development of standardized protocols for speech data collection and annotation. These efforts aim to balance the need for comprehensive data with respect for participant privacy and ethical considerations.

Furthermore, initiatives to crowdsource data collection and engage the broader community are emerging as innovative solutions to augment dataset sizes while ensuring diversity and representativeness.

4.2. Ethical and Privacy Concerns

The utilization of speech data for emotion and PTSD detection navigate a complex landscape of ethical considerations and privacy concerns. This intricate domain demands a careful examination of consent protocols, privacy safeguards, and the potential for misuse, ensuring that the pursuit of technological advancements does not compromise individual rights and ethical standards.

4.2.1. Informed Consent

Informed consent stands as a cornerstone in the ethical collection and use of speech data for PTSD detection. Participants must be fully aware of how their data will be used, the purpose of the research, and any potential risks or benefits involved. This transparency is crucial in respecting the autonomy of individuals and fostering trust between researchers and participants. The dynamic nature of consent in digital domains, where data may be reused for future studies, calls for ongoing consent processes that allow individuals to reassess their participation as the scope of research evolves.

4.2.2. Privacy and Data Security

The privacy of individuals contributing speech data is paramount, given the sensitive nature of the information and its potential to reveal intimate details about one's mental health status. Adhering to strict data anonymization protocols and implementing robust cybersecurity measures are essential to safeguarding participant privacy. Furthermore, the de-identification of speech data, while preserving its utility for PTSD detection, presents a significant technical challenge that requires innovative solutions to balance privacy with research needs.

4.2.3. Potential Misuse and Discrimination

The potential for misuse of speech analysis technologies raises ethical alarms, particularly in contexts where such tools could be employed for surveillance or discriminatory purposes. The capability to detect emotions or mental health conditions from speech could, if misapplied, lead to stigmatization or unwarranted scrutiny. Guidelines and regulations must be established to prevent the use of these technologies in ways that could harm individuals or groups, ensuring that applications are aligned with principles of fairness and equity.

4.2.4. Addressing Ethical and Privacy Concerns

Addressing these ethical and privacy concerns necessitates a collaborative approach, engaging ethicists, technologists, policymakers, and stakeholders in the development of guidelines and standards for the ethical use of speech data in PTSD detection. Key to this endeavor is the establishment of multidisciplinary ethics committees to oversee research practices and the integration of ethical considerations into the design and deployment of technologies. Moreover, public engagement and dialogue can play a pivotal role in shaping societal norms and expectations around privacy and the ethical use of speech analysis for mental health purposes.

4.3. Accuracy and Generalizability Issues

In the realm of utilizing speech data for PTSD detection, achieving high accuracy and ensuring the generalizability of models across diverse populations and conditions emerge as formidable challenges. These issues are pivotal, as they directly impact the efficacy and applicability of diagnostic technologies in varied real-world scenarios. This section delves into the intricacies of these challenges, shedding light on the hurdles faced by researchers in crafting models that are both precise and universally applicable.

4.3.1. Variability across Populations

One of the primary obstacles to achieving generalizability lies in the inherent variability of speech patterns across different populations. Factors such as cultural background, age, gender, and linguistic diversity significantly influence speech characteristics, potentially affecting the model's ability to detect PTSD indicators accurately. Models trained on datasets that lack representation from a broad spectrum of the population may exhibit biases, resulting in diminished accuracy when applied to underrepresented groups.

4.3.2. Adapting to Diverse Conditions

The context in which speech data is collected can vastly impact its analysis. Environmental noise, recording quality, and the speaker's emotional state at the time of recording are just a few examples of conditions that can alter speech patterns. These variations pose significant challenges for models, which must be robust enough to account for such fluctuations and still accurately identify markers of PTSD.

4.3.3. Achieving High Accuracy

Attaining high accuracy in PTSD detection through speech analysis is contingent upon the model's ability to distinguish between nuanced emotional states and the complex manifestations of PTSD symptoms in speech. The subtlety of these indicators, coupled with the variability, makes it challenging to develop models that consistently perform well across different datasets and conditions. Moreover, the rarity of PTSD instances in the general population further complicates the task, as models require sufficient exposure to positive examples to learn effectively.

4.4. Strategies for Enhancing Generalizability and Accuracy

To address these challenges, researchers are exploring a variety of strategies aimed at enhancing the generalizability and accuracy of PTSD detection models:

- Diverse and Representative Datasets: Building datasets that are inclusive of a wide range of demographics,

languages, and conditions is critical for training models that perform well across diverse populations.

- Advanced Modeling Techniques: Utilizing advanced machine learning and deep learning techniques, such as transfer learning and domain adaptation, can help models adjust to new populations and conditions more effectively.
- Continuous Learning and Adaptation: Implementing models that can continuously learn and adapt from new data can help maintain their accuracy over time and across varying contexts.

4.5. Exploring Datasets for PTSD Classification and Detection

The foundation of any successful machine learning or deep learning model lies in the quality and diversity of the datasets on which it is trained and validated. In the context of PTSD classification and detection, the availability of comprehensive and representative datasets is crucial for developing accurate and generalizable models. This section provides an overview of key datasets that have been pivotal in advancing PTSD detection efforts, highlighting their characteristics and the roles they play in fueling research and innovation in this area.

4.5.1. Key Datasets in PTSD Research
The PTSD-Repository

The PTSD-Repository is a curated collection of data sourced from clinical studies and research on PTSD. It encompasses a wide range of data types, including speech recordings, physiological signals, and self-report questionnaire responses. This repository serves as a vital resource for researchers seeking to explore the multifaceted nature of PTSD through various modalities.

Veterans Affairs (VA) Speech Dataset

Sponsored by the U.S. Department of Veterans Affairs, this dataset consists of speech samples collected from military veterans, a group with a high prevalence of PTSD. The dataset is notable for its inclusion of detailed clinical assessments accompanying each speech sample, providing researchers with valuable ground truth data for model training and validation.

The International Affective Picture System (IAPS) for PTSD

While not exclusively focused on speech data, the IAPS dataset includes emotional stimuli that have been used in conjunction with speech analysis to investigate emotional reactivity in individuals with PTSD. This approach allows for the exploration of how PTSD affects emotional processing, providing indirect insights into the disorder's impact on speech and communication.

Public Speech Datasets with Emotional Annotations

Several publicly available speech datasets, initially designed for emotion recognition research, have been repurposed for PTSD detection studies. These datasets, such as the Emo-DB and RAVDESS, include emotional annotations that can be leveraged to identify speech patterns associated with PTSD symptoms despite not being specifically collected for PTSD research.

Characteristics and Challenges

The datasets mentioned above share the common goal of facilitating the detection and classification of PTSD through various data modalities. However, they also embody the challenges inherent in PTSD research, including the need for large, annotated datasets that accurately reflect the diversity of the affected population. Furthermore, ethical considerations regarding privacy and consent are paramount in the collection and use of these datasets, especially when dealing with sensitive information related to mental health.

Datasets for PTSD Classification and Detection: Overview and Properties

To provide a structured overview of the datasets pivotal for PTSD classification and detection, the following table summarizes key datasets, highlighting their sources, characteristics, and the unique contributions they offer to the field of PTSD research.

5. Future Directions in PTSD Detection through Speech Analysis

The burgeoning field of PTSD detection through speech analysis stands at the precipice of significant advancements, driven by rapid technological evolution and an expanding understanding of both PTSD and speech dynamics. As researchers and practitioners navigate the complexities of this interdisciplinary domain, several key areas emerge as pivotal for future exploration.

These directions not only promise to refine the accuracy and applicability of PTSD detection methods but also to broaden the impact of these technologies on mental health diagnostics and treatment. The following segments delineate these prospective avenues, highlighting the path forward for research and development in PTSD detection through speech analysis.

5.1. Enhancing Model Generalizability

A primary focus for future research lies in enhancing the generalizability of machine learning models across diverse populations and linguistic contexts. Achieving broad applicability necessitates the development of models that are robust to variations in speech patterns due to demographic, cultural, and linguistic differences.

Future efforts could include the creation of more diverse and representative speech datasets and the exploration of transfer learning and domain adaptation techniques to tailor models to specific populations.

Table 4. Overview of key datasets for PTSD classification and detection

Dataset Name	Source	Data Type (s)	Key Characteristics	Purpose/Use
PTSD-Repository [20]	Clinical Studies	Speech, Physiological Signals, Questionnaire Responses.	Comprehensive collection from diverse studies, providing a multi-modal approach to PTSD research.	Facilitates a holistic view of PTSD, supporting models that integrate multiple data types for detection.
Veterans Affairs (VA) Speech Dataset [21]	U.S. Department of Veterans Affairs	Speech	Includes detailed clinical assessments focused on military veterans.	Enables the study of PTSD in a population with a high prevalence of the disorder, emphasizing speech patterns related to trauma.
International Affective Picture System (IAPS) for PTSD [22]	Psychological Research	Emotional Stimuli (Images) and Accompanying Speech/Reaction Data.	Used to study emotional reactivity; not exclusively speech data but relevant for indirect speech analysis.	Assists in understanding the emotional processing aspects of PTSD and its manifestation in speech.
Public Speech Datasets with Emotional Annotations (e.g., Emo-DB [23], RAVDESS [24], SAVEE [25], IEMOCAP [26])	General Emotion Recognition Research	Speech with Emotional Annotations	Initially not intended for PTSD research but includes valuable emotional annotations.	Repurposed for PTSD detection to identify speech patterns and emotional states associated with the disorder.

5.2. Multimodal Data Integration

The integration of speech analysis with other diagnostic modalities presents a promising avenue for improving PTSD detection accuracy. Future research could explore the synergistic potential of combining speech data with facial expressions, physiological signals, and behavioral data to construct a more holistic view of an individual's mental health status. Developing algorithms capable of processing and analyzing these multimodal data sources could unveil nuanced indicators of PTSD, offering a more comprehensive diagnostic tool.

5.3. Advancements in Deep Learning Architectures

Continued advancements in deep learning architectures and algorithms hold the potential to impact PTSD detection through speech analysis significantly. Future directions may include the exploration of novel neural network models that offer enhanced interpretability and the ability to capture complex, subtle patterns in speech indicative of PTSD.

Additionally, the application of attention mechanisms and transformer models could provide new insights into the temporal and contextual aspects of speech that are relevant for PTSD diagnosis.

5.4. Ethical and Privacy Considerations

As technologies for PTSD detection through speech analysis advance, so too must the frameworks for addressing ethical and privacy concerns. Future research should prioritize the development of protocols and technologies that safeguard participant data, ensure informed consent, and maintain transparency in model development and application. The ethical implications of automated PTSD detection, including potential biases and the impact on individuals diagnosed, will require ongoing scrutiny and ethical oversight.

5.5. Interdisciplinary Collaboration

The complex nature of PTSD detection through speech analysis necessitates interdisciplinary collaboration among psychologists, computer scientists, linguists, ethicists, and healthcare professionals. Future research directions could benefit from these collaborative efforts, leveraging diverse expertise to refine diagnostic models, address ethical challenges, and ensure that technologies are developed and deployed in a manner that benefits individuals with PTSD.

5.6. Clinical Integration and Real-World Applications

Translating research advancements into clinical practice and real-world applications remains a crucial objective. Future studies could focus on the implementation and evaluation of speech analysis technologies in clinical settings, assessing their impact on diagnostic processes, treatment outcomes, and patient experiences. The development of user-friendly tools and platforms that integrate speech analysis for PTSD detection into healthcare workflows will be essential for widespread adoption.

6. Conclusion

In summary, this comprehensive survey underscores the significant strides and inherent challenges in the realm of PTSD detection through speech analysis, bridging the domains of artificial intelligence, computational linguistics, and mental health diagnostics. The integration of advanced machine learning and deep learning techniques with speech analysis has emerged as a promising avenue for enhancing PTSD diagnostics, offering a non-invasive, objective, and scalable tool that holds the potential to transcend traditional diagnostic methodologies. However, the research journey unveils complex ethical and privacy concerns necessitating rigorous attention to ensure the respectful and secure handling of sensitive speech data. Furthermore, the pursuit of generalizable and interpretable models highlights the ongoing need for technological advancements that can adapt across diverse populations and linguistic variances, thereby reinforcing the integrity and applicability of PTSD detection tools. Looking forward, the anticipation of integrating speech analysis with other diagnostic modalities opens a new frontier for creating more comprehensive and accurate mental health diagnostics. This survey not only reiterates the importance of the research area but also reflects optimistically on the transformative impact anticipated from future advancements in technology and methodology. The promise of these innovations to revolutionize mental health diagnostics and treatment is profound, offering a beacon of hope for those affected by PTSD and signaling a pivotal shift towards more accessible, personalized, and stigma-free approaches to mental health care. As the field advances, these technological developments must be navigated with an unwavering commitment to ethical standards, interdisciplinary collaboration, and the enhancement of patient care, ensuring that the potential benefits are realized in a manner that is both responsible and impactful.

References

[1] Thole H. Hoppen, Ahlke Kip, and Nexhmedin Morina, "Are Psychological Interventions for Adult PTSD more Efficacious and Acceptable when Treatment is delivered in Higher Frequency? A Meta-Analysis of Randomized Controlled Trials," *Journal of Anxiety Disorders*, vol. 95, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Nicole M. Christ et al., "Using Machine Learning to Predict Sudden Gains in Intensive Treatment for PTSD," *Journal of Anxiety Disorders*, vol. 100, pp. 1-10, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Narayanan Srinivasan, and S.R. Balasundaram, "Minimizing Errors in Air Traffic Speech Using Rule-Based Algorithms," *International Journal of Computer Engineering in Research Trends*, vol. 10, no. 12, pp. 42-48, 2023. [CrossRef] [Publisher Link]

[4] Alice Othmani et al., "Machine-Learning-Based Approaches for Post-Traumatic Stress Disorder Diagnosis Using Video and EEG Sensors: A Review," *IEEE Sensors Journal*, vol. 23, no. 20, pp. 24135-24151, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Mishaim Malik et al., "Automatic Speech Recognition: A Survey," *Multimedia Tools and Applications*, vol. 80, no. 6, pp. 9411-9457, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Shrikanth Narayanan, and Panayiotis G. Georgiou, "Behavioral Signal Processing: Deriving Human Behavioral Informatics from Speech and Language," *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1203-1233, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[7] Julna Nazer, and K. Sajeer, "Automatic Speech Recognition-A Survey," *International Journal of Computer Engineering in Research Trends*, vol. 3, no. 4, pp. 190-193, 2016. [Google Scholar] [Publisher Link]

[8] Maria Labied et al., "An Overview of Automatic Speech Recognition Preprocessing Techniques," *2022 International Conference on Decision Aid Sciences and Applications*, Chiangrai, Thailand, pp. 804-809, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Uppu Jithendra, Usha Mittal, and Priyanka Chawla, "Audio Detection Using Mel-Frequency Cepstral Coefficients," *2021 9th International Conference on Reliability*, *Infocom Technologies and Optimization (Trends and Future Directions)*, Noida, India, pp. 1-5, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Yang Liu et al., "LPCSE: Neural Speech Enhancement through Linear Predictive Coding," *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, Rio de Janeiro, Brazil, pp. 5335-5341, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Sara Sandabad, Achraf Benba, and Hasna Nhaila, "Parkinson's Syndrome Diagnosis Applying Perceptual Linear Prediction Cepstral Analysis on Several Speech Recordings," *International Journal of Engineering Trends and Technology*, vol. 70, no. 9, pp. 214-221, 2022. [CrossRef] [Publisher Link]

[12] Fanling Huang, and Yangdong Deng, "TCGAN: Convolutional Generative Adversarial Network for Time Series Classification and Clustering," *Neural Networks*, vol. 165, pp. 868-883, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Vishal Passricha, and Rajesh Kumar Aggarwal, "Convolutional Support Vector Machines for Speech Recognition," *International Journal of Speech Technology*, vol. 22, no. 3, pp. 601-609, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[14] K.A. Senthil Devi, and B. Srinivasan, "Spoken Keyword Spotting System Design Using Various Wavelet Transformation Techniques with BPNN Classifier," *International Journal of Computer Engineering in Research Trends*, vol. 4, no. 3, pp. 111-118, 2017. [Publisher Link]

[15] Kai Yu et al., "Context Adaptive Training with Factorized Decision Trees for HMM-Based Statistical Parametric Speech Synthesis," *Speech Communication*, vol. 53, no. 6, pp. 914-923, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[16] Fatemeh Noroozi et al., "Vocal-Based Emotion Recognition Using Random Forests and Decision Tree," *International Journal of Speech Technology*, vol. 20, no. 2, pp. 239-246, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[17] Miao Liu et al., "Non-Intrusive Speech Quality Assessment Based on Deep Neural Networks for Speech Communication," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Hyeong II Koh, Sungdae Na, and Myoung Nam Kim, "Speech Perception Improvement Algorithm Based on a Dual-Path Long Short-Term Memory Network," *Bioengineering*, vol. 10, no. 11, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] K. Islam, and Z. ElSayed, "Speech-Based Emotion Recognition and PTSD Detection through Machine and Deep Learning," *International Journal of Computer Engineering in Research Trends*, vol. 11, no. 3, pp. 46-53, 2024. [CrossRef] [Publisher Link]

[20] Maya O'Neil et al., "Development of a Publicly Available Database of Randomized Controlled Trials for Posttraumatic Stress Disorder: The PTSD-Repository," *Archives of Physical Medicine and Rehabilitation*, vol. 101, no. 11, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[21] Parimal A. Itankar, and Hirendra R. Hajare, "Indoor Environment Navigation for Blind with Voice Feedback," *International Journal of Computer Engineering in Research Trends*, vol. 3, no. 12, pp. 609-612, 2016. [Google Scholar] [Publisher Link]

[22] M. Muzammil Parvez, H. Salam, and Y. Hoffmann, "Next-Generation Speech Analysis for Emotion Recognition and PTSD Detection with Advanced Machine and Deep Learning Models," *Synthesis: A Multidisciplinary Research Journal*, vol. 1, no. 1, pp. 11-21, 2023. [Publisher Link]

[23] Hussain Basha Pathan, Shyam Preeth, and M. Bhavsingh, "Revolutionizing PTSD Detection and Emotion Recognition through Novel Speech-Based Machine and Deep Learning Algorithms," *Frontiers in Collaborative Research*, vol. 1, no. 1, pp. 35-44, 2023. [Publisher Link]

[24] Vandana Singh, and Swati Prasad, "Speech Emotion Recognition Using Fully Convolutional Network and Augmented RAVDESS Dataset," *2023 International Conference on Advanced Computing Technologies and Applications*, Mumbai, India, pp. 1-7, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[25] Zhen-Tao Liu et al., "Speech Emotion Recognition Based on an Improved Brain Emotion Learning Model," *Neurocomputing*, vol. 309, pp. 145-156, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[26] Samarth Tripathi, Sarthak Tripathi, and Homayoon Beigi, "Multi-Modal Emotion Recognition on IEMOCAP Dataset Using Deep Learning," *arXiv*, pp. 1-5, 2018. [CrossRef] [Google Scholar] [Publisher Link]