A Survey Of Machine Learning Techniques On Speech Based Emotion Recognition And Post Traumatic Stress Disorder Detection

Chappidi Suneetha¹, Dr. Raju Anitha²

Abstract:
This paper reviews the literature on a wide array of methodologies utilized for emotion recognition from stressed speech. These techniques include models for the detection of post-traumatic stress disorder (PTSD), neural networks, and long and short-term memory networks (LSTM). The relevance of selecting alternative classification models has also been emphasized. The crucial issues to consider for future emotion recognition research in general, specifically in the Indian context, have been highlighted where applicable. Finally, possible future trends in stressed emotion recognition are discussed, including the use of region-based convolutional neural network (RCNN), you look only once (YOLO), recurrent neural networks (RNN), and LSTM techniques to solve new challenges in this domain.

Keywords: stressed emotion recognition, CNN, YOLO, RNN, LSTM

DOI Number: 10.4704/nq.2022.20.14.NQ88010

1. Introduction
Emotional stress detection computers strive to make human-computer interfaces more effective and natural. Stress, receives little attention, in spite of its certainty that one suffers stress at work or in everyday settings. These scenarios, for example, while driving during rush hour [2], might be perilous. In these cases, an intelligent car management system could be extremely useful, relaxing the driver by playing new music or alerting him to his present emotional state. This example shows how valuable an automatic stress recognizer may be [3][4]. However, such applications are currently challenging to deploy because data is difficult to come by, and the available datasets are largely military recordings [5].

Speech is created as a sequence of complex synchronized articulator movements, respiratory system airflow, and vocal system in physiological time [6]. While variations in articulator posture produce speech, not all utterances produced by a speaker will be identical in every way. This reason is that, in many circumstances, the person is under emotional stress, which influences the utterance and causes articulator movements to deviate [7][8]. Listeners can handle or process tiny variations in human communications far better compared to automatic human-machine interaction [9]. Stress and its impacts on human speech production, perception, and automated speech systems are currently poorly understood [10]. The complexity of the human speech signal is that comprises data about the speaker, his or her state, the acoustic surroundings, the person's aim, their language history, accent and dialectical features, and other paralinguistic elements [11]. One of the major goals of this survey is to discuss stressed emotion recognition methods in addition to voice and speaker recognition. In voice

¹Corresponding Author: Chappidi Suneetha
Address: ¹Research Scholar, Department of Computer Science and Engineering, Koneru Lakshmaia Education Foundation, Vaddeswaram, India., E-mail: 183030075@kluniversity.in
²Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaia Education Foundation, Vaddeswaram, India., E-mail: anitharaju@kluniversity.in

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Received: Accepted:
recognition systems, the inequality at intervals in training and test utterance has a wide reflection on performance. Observing that reaction of stress on the speech production system of humans will undoubtedly enhance the performance of speech recognition technologies and aid in synthesizing of speech to replicate emotions [12][13]. It is necessary to have the knowledge on the qualities of “stress” before moving further towards analysis and development of such aforementioned systems, as the stress qualities will define the need for a speech based system [14]. Stress is a challenging concept to define because it exists on a spectrum and is not always a binary choice [15]. A single definition cannot, in general, include all instances. For practical purposes, most definitions may be regarded as a little hazy. Despite this, we will use principles of linguistics to illustrate the importance of syllable in our elucidation. As a consequence, we use the term "speech under stress" to describe a scenario in which a person is speaking under duress and the production process of speech is altered [16]. "Speech under neutral conditions" refers to speaking that takes place in stress-free, pressure-free conditions. Usually as outcome, stress is a psychological state that arises in acknowledgement to expanding threat or task demand and is often associated with specific emotions. Even for cases in which a human does not desire it, these changes also might have effect on human speech. Any deviation from the neutral manner of speech, including speaking style, word selection, and body language, is classified as communication under stress. These changes may cast an effect on the person’s speech, even if it’s undesired. Speaking during stress is defined as any departure from the neutral mode of speech, encompassing speaking style, word selection, word usage, including sentence length [17]. As a result, communication under stressful settings refers to speech produced outside of a natural, conversational framework due to some environmental cause or emotional state [18]. The physical and emotional position of a speaker may differ across different stressful settings. Some examples include emergency responders, essentially police, fire, ambulance, and military soldiers engaged in peacekeeping or military activities [19].

High cognitive load, sleep deprivation, irritation over misleading evidence, emotion such as anxiety, illness, emotional exhaustion, and other modern-day multitasking situations all contribute to stress. The following areas have a higher concentration of emotions:

1. Diagnostics—involve the actions, conversation analysis including potential threats [20].
2. Protection as well as security — undersea explorers, spacecraft astronauts, power distribution technicians, [21], army forces confronting a panel of assessors, but also criminal justice trainees [22][23].
3. Psychology - patients’ emotional states [24]. Many investigations have been fetched on workload, task training stress, in noisy environments [25].

Stress from cognitive or physical strain, the Lombard effect from a loud environment, accent shift, and linguistic variability are all elements that influence speech output and the speaker, as seen in figure 1. These numerous variables or settings, impair the enactment of automatic speech organisms & human speech perception. In noisy, stressful circumstances requiring voice search along with speech recognition, speaker verification, and dialogue systems, we discovered that decreasing noise alone is insufficient to counteract performance losses. Even if loud stressful events could be avoided entirely, stress-induced production variability, particularly the Lombard phenomenon, also has a considerably greater impact upon speech system stability.

Figure 1 Stress and other factors have an impact on the speech system

Understanding the multitude of ways that stress besides emotion affects speaking output in real-world settings is more important than speech technology progression. There are three key areas of speech important from the standpoint of communication:
b) For voice and speaker recognition, speech systems, and technology use variations from the acoustic signal, and
c) Hearing and human perception, especially resolve whether a person is stressed or not stressed.

2. Literature Review

There are various approaches are present for Stress emotion recognition. Some of the important literature that covers the more important emotion recognition is discussed below:

Lalitha S, Gupta D, et al. [26] describe a novel speech emotion recognition (SER) system that is multilingual, that takes into account, five distinct languages and dialect variation. Standard speech emotion databases are usually generated by mixed-lingual corpora. A compact feature set is also created, which includes discrete combinations of increased perceptive traits and modified H-coefficients along with a diverse set of speech features. It also provides evidence for detecting emotions in SER technologies that are linguistic or mixed-language using samples from a variety of languages and dialects.

Passardi S, Peyk P, et al. [27] Post traumatic stress disorder (PTSD) detection based on facial attributes is yet to be researched. People with PTSD reported higher levels of thought or emotion stopping feelings and alexithymia without all these, especially emotional deregulation indicated a poorer level of recognition of unpleasant facial expressions. Although face recognition and reactions to perceived facial emotion are unaffected by PTSD, alexithymia is more significant than depressive symptoms for recognizing unpleasant emotions in the face.

A. Bérubé, C. Blais, et al. [28] Mothers were shown photographs of their children’s faces with changing boundaries of the six fundamental emotional expressions. The sensitive behaviors of partnerships between children and their mothers were then coded. Each mental instability questionnaire assessed the mothers’ records of childhood maltreatment. The goal of this research is to identify the abusive history and maltreatment during childhood which impacts the bondage with the mother.

Haider F, Pollak S, et, al. [29] Reducing the number of qualities utilized in inductive inference is one technique to attain this goal. We compare three approaches in this study, we apply state-of-the-art features extraction approaches such as infinite latent feature selection (ILFS), relief and fisher (generalizable fisher scores) to something like a new deep learning strategy called active feature selection (APS). Three emotion recognition data sets are evaluated by combining two pairs of conventional audible phonetic elements Berlin Database of Emotional Speech (EmoDB), Surrey Audio-Visual Expressed Emotion (SAVEE) and Italian Emotional Speech Database (EMOVO).

Egger M, Ley M, et, al. [30] This study should aid practitioners, researchers, and engineers in determining the exact elucidation for a prearranged situation. The use of a video camera to examine facial features is a non-contact method. Whether utilizing computers, cell phones, even tablets with built-in cameras, this would be very beneficial. Smart wearable’s make coming into contact with a person, performing stochastic activities, subsequent clinical standards with electro-dermal activities as well as heart-related signals maybe entered in a discreet manner. Heterogeneous sentimental supercomputers outperform unimodal solutions in terms of categorization accuracy.

Yin Z, Liu L, et, al. [31] Locally robust electroencephalogram feature selection (LRFS) is a novel technique for finding generalizable characteristics of electroencephalography (EEG) in a variety of patient subgroups. These obtained EEG features were initially characterized through probability density functions in the LRFS paradigm. The inter-individual stability of EEG parameters are all for the resemblance from all probability density functions in between two subjects. The resulting stability is being used to identify locally robust EEG properties, using emotional margin loss indicating the relevance of each feature.

Pandey P, Seeja KR, et, al. [32] Emotion recognition electroencephalography waves were used to demonstrate sentiment identification process. The intrinsic mode functions are used to extract two features for emotion categorization: the power spectral density’s (PSD’s) optimal values and also the first discrepancy.
Although ERA logical signals (DEAP) raises) scores by recording act of personal ment paradigm. In Effects of early S on social and emotional d as a potential When General Divide 80 healthy h hair cortisol levels, and these theemotional feelings of patients were altered characteristics on moderation was looked into. Furthermore, the imp emotion effects of these two types of dermatologists’ career success relationship. It is anticipated that it will encourage further organizations to discover the relationship between employment segregation and ERA using different approaches and from different perspectives to get a fuller picture of the ERA-career success relationship. Visser LN, Tollenaar MS, et, al. [35] Overall effects of these two types of dermatologists' emotion-oriented conversation upon patients' medical training recollection, including the feasibility of mediated by such a lowering throughout stressful situations, were explored. Furthermore, the impact of personal characteristics on moderation was looked into. An oncologist’s response in words to theemotional feelings of patients were altered during a recorded, prearranged bad-news session. The three situations that were developed were emotion-oriented speech, continuous conversation and silence based on emotions. A total of 217 people were assigned to all three circumstances at random. Hartling C, Fan Y, et, al. [36] Effects of early childhood distress in emotion recognition on the face is moderated by a profile of genetic variability linked to variation in the hypothalamic-pituitary-adrenal (HPA) axis activity. If the findings are replicated in a different cohort, they could point to a biological foundation for vulnerability to the negative effects of ELS on social and emotional functioning. The entire discovery adds to either a growing body of evidence emphasizing the importance of this process in determining ELS impacts or to a growing body of evidence emphasizing the importance of this mechanism in determining ELS impacts. Zeier P, Sandner M, et, al. [37] Divide 80 healthy people into two groups, one working on a stress 'two groups either working on a stress' (TSST) target and remaining control task. Both groups then pay attention on updated emotion management paradigm. In this exercise, participants are asked to generate distinct appraisals of diverse emotional scripts as possible to reduce unpleasant emotions. We can analyze individual fluency (number of reappraisals) and flexibility (diversity of reappraisals) scores by recording the participants' reappraisals. Wirkner J, Ventura-Bort C, et, al. [38] To measure sympathetic and neutrality scenario interpretation and afterward recognition memory in female volunteers, high cortisol quantities have been used as a potential biomarker for psychological stress. Event-related potentials demonstrated that emotional material elicited higher late positive potentials (LPPs) than neutral content during picture viewing. Total LPP volumes are associated positively with hair cortisol levels, and these brain potentials changed in response to long-term hair cortisol levels. Smith KE, Leitzke BT, et, al. [39] The researchers investigated how child psychological evaluations are influenced by the interaction of stress exposure. A total of 89 young people ages (11–15 years) were disclosed to one-of-a social stressor via video. The children took an emotion...
Chappidi Suneetha, Raju Anitha, /A Survey Of Machine Learning Techniques On Speech Based Emotion Recognition And Post Traumatic Stress Disorder detection

...examine before and participants witnessed combined representations including facial movements as well as circumstances suggesting consistent versus irreconcilable facial expressions following the stressor. Observation of the eyes was used to track alterations affecting attentiveness towards stimuli. To label the emotions for children, expose them to acute stress and acquire face relevant data is most recommended.

Martin D, Croft J, et, al, [40] conducted a comprehensive assessment of literature upon this correlation among probability feature for illusion or psychosis and emotion recognition processing deficiencies. Since there was enough data, meta-analyses were performed; alternatively, descriptive descriptions have been used. In both overall and specific facial emotion recognition deficiencies, meta-analyses were conducted.

### Table 1. Comparative study of stress emotion recognition techniques.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Methodology</th>
<th>Advantages</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lakshmi S, Surya</td>
<td>2018</td>
<td>Multilingual and monolingual speech</td>
<td>High accuracy</td>
<td>High accuracy</td>
</tr>
<tr>
<td>Premkumar R, Anand</td>
<td>2018</td>
<td>Post-traumatic stress disorder</td>
<td>Detects even the slightest facial muscle activation</td>
<td>High accuracy</td>
</tr>
<tr>
<td>Breiden J, Ernst C</td>
<td>2018</td>
<td>Facial recognition and nonverbal expression</td>
<td>Provides consistent translation for all emotions</td>
<td>High accuracy</td>
</tr>
<tr>
<td>Haseen P, Rehman S</td>
<td>2018</td>
<td>Evaluation of automated feature extraction methods</td>
<td>Increased prediction</td>
<td>High accuracy</td>
</tr>
<tr>
<td>Raja M, Raj K</td>
<td>2018</td>
<td>Breath recognition from physiological signals</td>
<td>Observers</td>
<td>High accuracy</td>
</tr>
<tr>
<td>Fu L, Li L</td>
<td>2018</td>
<td>The ROI feature extraction of facial robust</td>
<td>Deriving neural con. eliciting and safety-related perceptions</td>
<td>Average accuracy</td>
</tr>
<tr>
<td>Pathy A, Bajaj A</td>
<td>2018</td>
<td>VED and Deep learning</td>
<td>Unmatched at accuracy</td>
<td>High accuracy</td>
</tr>
<tr>
<td>Shrivastava P, Pranav E</td>
<td>2018</td>
<td>E-Learning environment</td>
<td>It benefits from both consistency and the semantic transformers</td>
<td>High accuracy</td>
</tr>
<tr>
<td>Kaushik I, Bhaskar S</td>
<td>2018</td>
<td>Offline emotion detection</td>
<td>Higher accuracy</td>
<td>Higher accuracy</td>
</tr>
<tr>
<td>Nader UA, Taklauer MC</td>
<td>2018</td>
<td>Cognitive emotion oriented communication</td>
<td>Suitable design's logical validity is inevitably biased</td>
<td>High accuracy</td>
</tr>
<tr>
<td>Burling C, Fox F</td>
<td>2018</td>
<td>Interview of RPA/HR: process</td>
<td>Evaluating higher-order functions</td>
<td>High accuracy</td>
</tr>
<tr>
<td>Cidre P, Soberon W</td>
<td>2018</td>
<td>Potential emotion models</td>
<td>High accuracy</td>
<td>High accuracy</td>
</tr>
<tr>
<td>Villasenor E, Petersen P,</td>
<td>2018</td>
<td>Cognitive arousal and motivation</td>
<td>High accuracy</td>
<td>High accuracy</td>
</tr>
<tr>
<td>Smith J, Sutherland W</td>
<td>2018</td>
<td>Response to stress and anxiety-based laboratory data</td>
<td>Significant evidence in both human responses and their impact evaluating the process</td>
<td>High accuracy</td>
</tr>
<tr>
<td>Martin D, Croft J</td>
<td>2018</td>
<td>Intra-person and inter-personal recognition</td>
<td>Suitable recognition</td>
<td>High accuracy</td>
</tr>
</tbody>
</table>

The above comparison table represents stress emotion recognition techniques and indicates that the case of emotion recognition by e-learning achieved the highest accuracy. Compared to the above techniques Approaches for recognition of emotions, with such a focus on e-learning settings achieves high accuracy as well as it takes the Stability requirement as well as exponential modification each have their merits.

### 3. Studies Related To Stress Emotion Recognition Based Rcnn&Yolo

Neural networks (NN) are made up of layers of processing units called neurons, each one has its own set of connections. Each neuron multiplies an initial value by weight, adds the results to additional values flowing into the same neuron, modifies the result by the bias of the neuron, and finally normalizes the output with an activation function.

A neural network (NN) is a system of neurons, or simply an artificial neural networks (NN) made up of artificial neurons or nodes, in today’s terms [41]. A neural network can therefore be physiological or synthetic, and it could be applied to solve artificial intelligence challenges. Weight values are used to depict the connectivity among neurons in the biological analogy. Relationship of excitability have a positive weight, while inhibitory links are negative bonds. The inputs are given weight before being totalled. The output's amplitude is finally controlled by an activation function. For illustration, a suitable frequency range is often 0 to 1, but it could alternatively be -1 to 1.

![Figure 2. Structure of Neural Network](image-url)

**A. Neural Networks Of Various Types In Deep Learning**

There are three major kinds of neural networks which provide the foundation to most pre-trained deep learning models, however current analysis specifically looks into CNN and RNN:

**B. Convolution Neural Network (Cnn)**

Convolution neural network (CNN) and its robust architectural and functional variations is currently the vogue inside the community of deep learning. Those neural networks are employed in several activities as well as sectors,
although image and video analytics is where they’re most prevalent.

C. Advantages Of Convolution Neural Network (Acnn)

Convolutional neural network (CNN) identifies the limitations but it does not broadcast anything. The extraction about the suitable knowledge from the incoming data is aided by these filters. It grasps the spatial nature of an appropriate feature input.

In a target speech under stress, geographical features depend on the organization of levels of stress and also the emotions which emerge among them. These aid in the accurate identification of sentiments [42].

Wang S, Li J, et, al. [43] The Laban Movement Analysis (LMA) method is a dance emotion expression method that comprises the characteristic parameters. Authentic dance movement data is converted onto three unique interpretation factors to produce dance emotion data, using body structure, spatial orientation, as well as pressure effect. The dancing emotion collection is tested and trained using CNN and LSTM hybrid neural network models. Eventually, decision tree algorithms, random forests, CNNs, and LSTMs were established and checked for capability to illustrate the CNN-LSTM model’s usefulness.

Kwon S, et, al. [44] demonstrated a one-dimensional adaptation to help neural network-based end-to-end genuine SER model. Then use a multi-learning approach in which the system acquired long-term situational associations whilst identifying geographically relevant psychological effects using voice signals. In an attempt to discover a connection, researchers employed the residual blocks with such a skip connection component, emotional cues, as well as the sequence learning subsystem to acquire the long-term contextual relationships throughout the input data.

Yao Z, Wang Z, et, al. [45] developed a confidence-based fusion technique in which several multi-task learning-based sub-classifiers were used for normative identification for emotional states: DNN plus expression high-level statistical functions (HSFs), convolution neural network (CNN) with multiple segment-level mel-spectrograms (MS), as well as RNN with multiple frame-level low-level descriptors (LLDs). To combine various CNN or RNN outcomes, researchers used a balanced pooling technique based on transfer learning. The brain network was willing to concentrate upon psychologically important regions, arguably as a result of the the activation function.

Cimtay Y, Ekmekcioglu E, et, al. [46] employed state-of-the-art CNN topologies which were pre-trained to considerably improve specific topic recognition accuracy. In contrast to earlier studies that generate spectrum band power statistics using EEG readings, they employed unprocessed EEG data without demodulating, pre-adjustments, as well as acceptability. Through, eliminating minutiae extraction from the training technique, the chance of omitting unique information inside the original reducing power, as well as the deep neural network’s ability to identify unknown features is increased.

Khare SK, Bajaj V, et, al. [47] multiple CNNs are proposed for automated feature extraction and categorization. To begin, the suggested approach using only a time-frequency notation, turns the sanitized EEG data sets into some kind of picture. To integrate frequency EEG signals onto pictures, a normalized pseudo-Wigner–Ville allocation is used. Pre-trained ‘alex net’, ‘ResNet50’, ‘visual geometry group’ (VGG16), and also a configurable CNN, were given those images. Overall accuracy, precision, Mathew’s coefficient of correlation, F1-score, as well as the false-positive rate of four CNNs are all evaluated.

Sarath S, et, al. [48] focused on evaluating sentiments hidden throughout the face, such as stress, worry, or excitement, that can cause a pleasant diffusion around the features. Those infrared graphics were estimated by making use of the thermal value of the pixel than the intensity of every pixel. Image processing methods were used for pulling out the region of interest which also included segregating the object from its backdrop. Given the wealth of publicly available methodologies for recognizing emotional states, economy, resilience, as well as efficiency, could all be enhanced by combining several modalities, which are focused on sensory facial features, thermal face imaging, as well as structural data.

Laroca R, Severo E, et, al. [49] describe a reliable and effective automatic License Plate
Recognition (ALPR) system that is based on the cutting-edge YOLO detection system. CNNs are formed, then fine-tuned at every ALPR phase to guarantee that they have been adaptable in a variety of circumstances. Researchers design a two-stage strategy for graphic identification and classification utilizing modest data augmentation including such upside-down license plates and reversed scripts. 

Chen RC, et al. [50] offer a robust automatic recognition system suitable for swiftly detecting license plate numbers whilst also ensuring accurate results. The classic YOLO was altered to make a compact YOLO singular detector containing 36 models. To prevent issues regarding microscopic element categorization, the detector uses a sliding window with all categories from each frame. The value loss was 0.40, while average detection, as well as recognition latency, was recorded as 825.81 ms, according to the research. The approach essentially affects a specific development stage identification and tracking technique, which doesn't involve feature extraction or echo cancellation.

Table 2. Comparative Study of Stress emotion recognition using RCNN & YOLO technique

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Methodology</th>
<th>Approach</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang LJ</td>
<td>2020</td>
<td>CNN and LSTM</td>
<td>Extracts enemy features but also maintains the temporal relationship of the data, also perfect for extracting lower motion elements.</td>
<td>Throat control 95%</td>
</tr>
<tr>
<td>Xiong S</td>
<td>2020</td>
<td>CNN, followed by Weak Learning 1986 approach</td>
<td>YOLO-DSP and EMDB datasets were used, as before</td>
<td>75% accuracy</td>
</tr>
<tr>
<td>Yau T, Wang Z</td>
<td>2020</td>
<td>YOLO-DSP, HHD, CO, and LSTM</td>
<td>CNN and LSTM</td>
<td>Weighted accuracy of 93.1% and average accuracy of 91.9%</td>
</tr>
<tr>
<td>Cutie D, Zhan M</td>
<td>2020</td>
<td>CCG, Decision Recognition</td>
<td>Adaptable categorization has the advantage of preserving the original distribution data.</td>
<td>65% accuracy</td>
</tr>
<tr>
<td>Kumar SK, Hajj S</td>
<td>2020</td>
<td>CNN</td>
<td>Stability as well as complexity</td>
<td>High accuracy</td>
</tr>
<tr>
<td>Saraf S</td>
<td>2020</td>
<td>EEG-based using Yolo Algorithms</td>
<td>YOLO can be enhanced</td>
<td></td>
</tr>
<tr>
<td>Larson R, Ferraro E</td>
<td>2019</td>
<td>A Robust Real-Time Automatic License Plate Recognition</td>
<td>They are far larger private datasets, which is a considerable advantage especially for deep learning approaches.</td>
<td>Improvement in accuracy</td>
</tr>
<tr>
<td>Chen RC</td>
<td>2019</td>
<td>Storing/Predicting Datasets Using Yolo R</td>
<td>It employs both CNN for feature learning and automated localization to get around the character representation technique.</td>
<td>High accuracy</td>
</tr>
</tbody>
</table>

Stress emotion recognition utilizing RCNN algorithms can be seen in comparison table 2. Dance emotion recognition depending on laban activity recognition, employing CNN as well as LSTM, achieves a significant precision of 92 basis points. It also has an outstanding experience, besides feature extraction of dance emotions compared to the other methodologies [16]. Yolo states in contrast to the processes discussed above, are used to recognize stress feelings. Darknet-Yolosliding-window automatic license recognition is combined with CNN for learning features and automated localization. Deep learning achieves great accuracy while avoiding the character segmentation step.

4. Studies Related To Speech Based Post Traumatic Stress Disorder Detection

Passardi S, Peyk P, et al. [51] High mental capability such as emotion control are impaired in patients with PTSD. Facial mimicry (automated facial adaptations towards observed facial emotion) is not analyzed till now in the literature on PTSD. Similarly, although abnormalities in recognizing facial emotions have also been documented, the underlying causes are unknown.

Ragsdale KA, Jones KR, et al. [52] adopted a 2-week course of treatment which involved long term exposure, cognitive schema therapy, with psychological capacity, independent of the pathophysiology, particularly resolving PTSD plus intellectual difficulties brought on to by mild Traumatic Brain Injury (mTBI). Another of the service’s main objectives should be to establish fully prepared expectancies to minimize the incidence of capability of sensing from treating suspected mTBI symptoms. The PTSD+mTBI partial hospitalization was finalized by thirty patients with filled or idiopathic intracranial hypertension PTSD and historiography of self-reported TBI. By exploring the required neurobiological rationale for ‘eye movement desensitization and reprocessing’ (EMDR), researchers propose a potential that a recently introduced neurophysiological focused nine construct model for experienced selfhood might assist address this difficulty.

Patterson J, Martin CJ, et al. [55] Researchers have developed a subjective procedure called interpretative phenomenological analysis, which particularly incorporates a specific framework, to understand better, the lived experience of patient advocate engagement. To learn more about how women suffering from post-traumatic stress disorder (PTSD) engage with physicians throughout care delivery, individualized face-to-face tests were held with
six women who met the entire treatment guidelines for Post-Traumatic Stress Disorder Post Childbirth (PTSD-PC) including six midwives who offered postpartum care. Particularly, where individuals sense throughout their relationships and what that implies to them. 

Gilmore AK, Lopez C, et al. [56] present additional predictive analytics with a current randomized controlled testing, estimating the worth of prolonged exposure (PE) provided in person. Especially, PE administered through telehealth among female veterans having military sexual trauma (MST)-related PTSD is presented. Military sexual trauma is pervasive amongst military personnel, and that can lead to depression like post-traumatic stress disorder and anxiety. Spite about substantiality there are a handful of scientific proof treatments for MST-related PTSD, include extended exposure (PE) psychotherapy, it is uncertain if there are any features that correlate towards treatment termination too early.

Table 4. Comparative study of detecting speech-based post-traumatic stress disorder technique

<table>
<thead>
<tr>
<th>Author Name</th>
<th>Year</th>
<th>Methodology</th>
<th>Advantage</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passari: G, Payi P</td>
<td>2020</td>
<td>Post-traumatic stress disorder</td>
<td>Detecting even the smallest facial muscle activations but does not show whether the muscle activation was generated by focal mimicry or was an effective response to the scene.</td>
<td>Higher accuracy</td>
</tr>
<tr>
<td>Fraghshehia EA, James KR</td>
<td>2020</td>
<td>Treatment of Gossip Treated written by Brain and Post-traumatic stress disorder</td>
<td>___</td>
<td>Medium accuracy</td>
</tr>
<tr>
<td>Vlasovic AK, Lehenbey A</td>
<td>2019</td>
<td>Post-Traumatic Stress also Alcohol Use Disorders</td>
<td>___</td>
<td>___</td>
</tr>
<tr>
<td>Fraghshehia AA, Fuglsang AA</td>
<td>2019</td>
<td>Three-dimensional model of the expiriencelessness</td>
<td>___</td>
<td>___</td>
</tr>
<tr>
<td>Patterson J, Neuro C</td>
<td>2019</td>
<td>Post-Traumatic Stress Disorder (PTSD) post childbirth</td>
<td>___</td>
<td>___</td>
</tr>
<tr>
<td>Gilmore AK, Lopez C</td>
<td>2020</td>
<td>Military Sexual Trauma-Related Posttraumatic Stress Disorder</td>
<td>___</td>
<td>___</td>
</tr>
</tbody>
</table>

An investigation involving stress emotion recognition as well as post-traumatic stress disorder approaches is mentioned in Table 3. Comparative analysis towards the components discussed earlier in this process, cranial imitation, cranial emotion, as well as symptomology in post-traumatic stress disorder possess a high level of faultlessness and also added benefit efficacious to determine for the gentle specific facial action potentials. Nevertheless, it doesn't provide any documentation upon whether muscle activation has been triggered by facial mimicry or if it was a right answer towards the stimuli.

4. Studies Related to PTSD Using RNN and LSTM

Zhang T, Wu J, et al. [57] in their study state, that in order to represent the speech signal more abstractly, traditional prosodic acoustic components are introduced and contrasted with new i-vector features. For mapping the qualities to emotion labels, a RNN technique is presented.

Ma J, Tang H, et al. [58] in their work state that the multimodal residual (MMRes) LSTM connectivity is just a bidirectional residual LSTM network. To understand the connection between the EEG along with other physiological particulars, the MMRes-LSTM is given high input values at every LSTM layer across the modality. It incorporates the spatiotemporal shortening pathways of the residual network as well as the LSTM, enabling quick learning emotion-related high-level information. DEAP, a publicly accessible dataset with EEG-based feeling of love, hate, guilt etc., recognition, has been used to figure out the proposed structure.

Nakisa B, Rastgoo MN, et al. [59] proposed a new framework for leveraging differential evolution (DE) to automatically improve LSTM hyperparameters. Here the initial study focused on hyperparameter optimization within the framework for emotional identification in such a systematic manner. Utilizing a novel algorithm derived through embedded devices, they evaluate and compare the conceptual methodology versus various state-of-the-art hyperparameter optimization methodologies. These results indicate how tweaking LSTM hyperparameters improves four-quadrant dimensional emotional expressions by 14% whilst also increasing precision by 14%.

Nath D, Singh M, et al. [60] in their study, employed the DEAP dataset, which is made up of pre-processed EEG and physiological data and is freely available. Their research is focused on analysing EEG signals to construct a headgear model that can track emotions in real-time. This article evaluates the recognition
accuracy of several predictors for such valence versus arousal subdomains utilizing band power, a frequency-domain feature extracted from EEG data.

Table 5. A comparison of RNN and LSTM techniques for recognizing post-traumatic stress disorder

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Methodology</th>
<th>Advantage</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang T, Wu J</td>
<td>2015</td>
<td>I-Vector Feature and RNN Model</td>
<td>Gaining a better grasp of the correlation by exchanging values among modalities is beneficial.</td>
<td>-</td>
</tr>
<tr>
<td>Ma J, Tang H</td>
<td>2019</td>
<td>Multimodal Residual LSTM Network</td>
<td>Swifly converge by producing a large number of diverse answers.</td>
<td>92.87% and 92.30%</td>
</tr>
<tr>
<td>Nakiss B, Rastg oo MN</td>
<td>2018</td>
<td>Long Short-Term Memory Hyperparameter Optimization</td>
<td>Averag e accuracy 90.5%</td>
<td>-</td>
</tr>
<tr>
<td>Nath D, Singh M</td>
<td>2020</td>
<td>LSTM Network</td>
<td>High accuracy 98.3%</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5 shows findings of a study employing RNN and LSTM approaches, identifying stress emotions in PTSD. While differentiating the above-noticed methodologies, we deduce that emotion recognition depends on LSTM and EEG. The LSTM network is quite effective.

5. Conclusion
The drawing of speech into emotions, aids in the naturalness of current speech systems’ performance. In the recent past, an exhaustive amount of work has been done in this area. This review study conducted on fairly recent work in stress emotion identification may pique the interest of the research community in filling some critical research gaps. This paper discussed most of the recent research undertaken in stressed-speech emotion recognition, with significance of stress under speech emotion frameworks, speech features, and models of classification. The paper also addresses many research questions in the investigation of voice-based emotion identification. Following a thorough examination of previous work, seminal articles, beneficial in this topic work were picked for investigation.

References


Cimtay Y, Ekmelcioglu E. Investigating the use of pretrained convolutional neural network on cross-

www.neuroquanotology.com
Chappidi Suneetha, Raju Anitha, / A Survey Of Machine Learning Techniques On Speech Based Emotion Recognition And Post Traumatic Stress Disorder Detection


