Emotion-Based Speech Analysis for Disaster Response and Crisis Management

Rafa Siddiqua ID:23166023
Rabea Akhter ID: 23366029
Samiu Mostafa Ishan ID: 23273011

Abstract:

Emotion Based Discourse Examination is crucial in emergency situations for the board, working with the understanding of close to home states during catastrophes. The most recent Speech Emotion Recognition (SER) techniques for crisis scenarios are discussed in this abstract. Approaches, for example, coordinating BLSTM and LSTM organizations and using IMEMD-CRNN framework exhibit promising progressions in feeling acknowledgment. Different models enveloping consideration based BiLSTM, CNN-LSTM, Transformers, and Progressive Consideration Organizations exhibit adequacy in perceiving feelings during emergencies. This study focuses on the significance of different process for enhanced speech emotion recognizition and provides a idea for the use of development in order to improved crisis management models and architectures in the future.

Key word: Speech Emotion Recognition, IMEMD-CRNN,CNN-LSTM, BLSTM, LSTM

Introduction:

In an environment of rapid technological advancement, understanding and analyzing human emotions in emergency situations has become an important pursuit in many fields. This review is designed to provide a comprehensive research and report on various aspects of emotional intelligence in stress, including emotional intelligence (SER), social emotional analysis energy during natural disasters [5] [1], and the concept of signal communication in crises. detection [4]] and sentiment analysis on platforms such as Twitter during critical events [1]. Speech Emotion Recognition (SER) and emergencies. In human-computer interaction (HCI), speech recognition plays an important role in this breakthrough. Thoughts expressed in words are especially important in emergency situations [7]. Its applications include call centers, healthcare and digital marketing. However, improving the accuracy of natural language emotion recognition is still a challenge. Research is constantly working towards the creation of systems that can overcome language barriers, recognize different views of the speaker, and work well in noisy environments [2]. The role of social media in disaster response. The emergence of social media platforms has revolutionized the dissemination of messages and information. Expression of emotions, especially during natural disasters [8]. This chapter highlights the importance of public opinion surveys in assessing public opinion, influencing the way public opinion is influenced, and influencing emergency decision-making, providing a good insight into disaster management [6]. Voice Analysis Research in CrisisIt is important to understand the emotions sent by voice signals in crisis situations. Traditional methods use multitasking to identify emotions in the mind to help solve communication problems when sending voice data across the network in

critical situations [4]. Sentiment analysis and discovery on Twitter. Twitter's immediacy makes it useful during important events, providing a quick understanding of public opinion and sentiment[1]. This chapter explores the integration of event detection and sentiment analysis using Twitter profiles, using the Las Vegas shooting study as an example [1]. This comprehensive review integrates and presents a wide range of research theories in crisis situations and suggests that competition among leaders and misunderstandings of the response process are important for solving future problems in crisis management.

Aim and Objective:

This research has the target to compare different emotion recognition by using speech and to figure out which is the best one by provide a brief comparison in between different implementation, methodology and datasets. Provide a better view for the readers to make them understan what they can use or what will be the best way to use.
Aim:

1. Compare in between provided models and their implementation.
2. Compare in between the result and figure out what will have the best to use according to their accuracy.
3. Put out a final result for them..

Literature Review: In [1], the authors Sung-Woo Byun and Seok-Pil Lee constructed a Korean emotional speech database for speech emotion analysis. In [2], the authors proposed an approach based on emotional perception, which designs an implicit emotional attribute classification. The authors of [3] used two unidirectional LSTM layers for text recognition and fully connected layers are used for acoustic emotion recognition, which are then merged to produce the predicted emotion categories. The authors of [4] used a technique in their project, where it classifies emotions into five different categories. Different approaches for developing speech recognition, which are language and speaker independent, are briefly discussed by the author of [5].

Proposed Methodology:

Emotion-based speech analysis for disaster response and crisis management typically involves a combination of algorithms and techniques from natural language processing (NLP), machine learning, and signal processing.
Author Tris, Sirai and Massto have used different features in order to recognise the emotions from different speech where they have used different features like extraction from speech, acoustic feature extraction and speech emotion recognition models where the whole speech and voice segments using bidirectional LSTM networks, with or without attention models [6]. Here the author has utilized discourse Feeling Acknowledgment where acoustic elements are separated from discourse fragments after quietness expulsion.

Highlights incorporate time and ghostly area highlights, MFCCs (Mel-recurrence cepstral coefficients), and chromas. Alongside that, different profound learning designs (LSTM, consideration models) are assessed for discourse-based feeling acknowledgment.

The author additionally utilized the word Implanting feeling acknowledgment where they separated printed information from records tokenized and changed over into word embeddings alongside various profound learning structures (CNN, LSTM, LSTM with consideration) are investigated for text-based feeling acknowledgment.

In Consolidating Discourse and Text Highlights creator has proposed a methodology that includes joining the acoustic elements from discourse with word embeddings from text information with various models, including CNN, LSTM, and mixes of organizations, which are assessed for joined feeling acknowledgment [6].

As per creator Sun, Li and Mama they have zeroed in on a methodology called IMEMD-CRNN (Further developed Concealing Exact Mode Decay - Convolutional Repetitive Brain Organization) for foreseeing feelings in discourse signals. The strategy comprises of three fundamental modules: IMEMD-based profound discourse signal deterioration, extraction of time-recurrence highlights from IMFs (Inherent Mode Capabilities), and discourse feeling acknowledgment in light of CRNN (Convolutional Repetitive Brain Organization) [7].Here creator have utilized different IMEMD-based Close to home Discourse Signal Decay like EMD (Observational Mode Disintegration), Covering Signal-based EMD (MSEMD) and Further developed Concealing EMD (IMEMD). In EMD they have utilized motional discourse signal deterioration Non-fixed signals are separated by this into IMFs (Natural Mode Capabilities) and a buildup likewise various addresses mode mixing issues in EMD by using a sinusoidal veiling sign to disconnect different repeat parts alongside that creator additionally proposes a unique procedure to fabricate disguising signs that relieve mode mixing. It adds a disguising sign to the principal sign, breaks down it using EMD, and dispenses with the veiling sign to get high-repeat parts [7]. Extraction of Highlights In light of IMEMD Tone Elements: uses IMEMD decay to separate Hilbert range dispersion and shape highlights, among other ghostly elements. Mel-repeat cepstral coefficients (SMFCC) from the recreated signal procured through IMEMD, close by the first and second auxiliaries of SMFCC for discovering passing information. Convolutional Discontinuous Cerebrum Association (CRNN) which contains four 2D CNN blocks, followed by bidirectional GRUs and totally related layers, utilizing softmax inception at the outcome layer, alongside that they have likewise utilizes the Adam enhancer and cross-entropy misfortune to prepare the organization over different ages at a foreordained learning rate and little bunch size [7].The overall objective of the proposed method is to use a CRNN architecture and IMEMD-based signal decomposition, followed by feature extraction, to boost the robustness and accuracy of speech emotion recognition systems. The IMEMD strategy means to relieve mode blending in EMD, while the

CRNN model is utilized to gain and order feelings from the removed elements. This approach is intended to improve feeling forecast in discourse and announces consolidating signal handling procedures with profound learning techniques.

According to Vydana, P. Vikash, T. Vamsi, K. P. Kumar and A. K. Vuppala using Identifying emotionally Important Areas where one needs to calculate is utilized to recognize sincerely critical sections inside an expression [8]. These fragments address the durationally short emotive motions made by the speaker.Along with that highlight extraction of ghastly vectors when the genuinely huge districts are recognized, phantom vectors are processed from the discourse information inside these distinguished portions. Using model turn of events where gaussian Blend Displaying (GMM) which strategy is utilized to make models for feeling acknowledgment utilizing the unearthly vectors extricated from the genuinely critical portions [8].

According to Liu, Y. Mou, Y. Ma, C. Liu and Z. Dai the review proposes a methodology for perceiving feelings in discourse through a sliding window-based technique combined with counterfeit brain organization (ANN) displaying.Where one need to uses non stop sliding windows of a particular length (Mδ) and slide distance (Δ) to perceive feelings window by window inside a discourse test [9]. Identifying emotionally significant inorder to critical locales inside every expression utilizing sliding windows, creating a succession addressing close to home vectors.Also weight conveyance capability and lattice development which put a weight circulation capability to portray the commitment of feeling from span level to window-level feelings. Develops a framework (G) to plan span level close to home vectors to window-level profound vectors.This models close to home dispersions inside every window utilizing Gaussian likelihood thickness capabilities, approximating profound commitment inside every window. Also use fusion and extraction of features from where one can determines span level close to home vectors (e*) in view of the weight dispersion and straightforwardly perceives window-level profound vectors (E). Both e* and E are considered as highlights and melded. End with testing the EMO-DB Dataset: Assesses the proposed model utilizing the Berlin Feeling Data set (Emotional DB), choosing explicit feelings for examination. Conducts five-overlay cross-approval for preparing and testing sets, guaranteeing speaker autonomy [9].

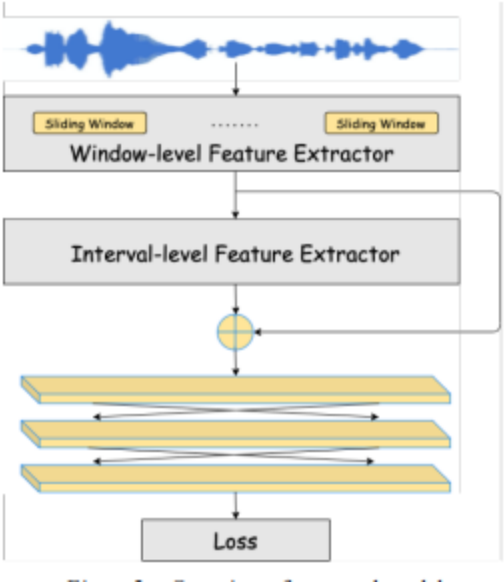

Figure: Overview of proposed model

According to Sharma, Dutta and Pradhan, coordinate discourse information with physiological signs like pulse, skin conductance, or looks to make a multimodal dataset. Using combination strategies like late combination (consolidating highlights at a later stage) or early combination (incorporating highlights at the info level) to thoroughly catch profound prompts more [12] . Consolidating logical data from the discourse content, environmental factors, or progressing occasions to all the more likely figure out the close to home state. Using context oriented embeddings or consideration systems to gauge the significance of various logical components in feeling acknowledgment [13]. Utilizing pre-prepared models on huge close to home discourse datasets and calibrating them on unambiguous fiasco related profound discourse datasets [14]. Using move figuring out how to conquer information shortage in calamity explicit situations. Growing continuous feeling discovery frameworks that can dissect progressing discourse information to give quick criticism or help in emergency the board procedures. Using lightweight models upgraded for speed and exactness. Stretching out feeling based discourse examination to different dialects pervasive in misfortune impacted areas. Involving cross-lingual models or multilingual methodologies for feeling acknowledgment in discourse [15].

Data Analysis:

Author Tris, Sirai and Massto have discussed about IEMOCAP dataset in their paper Speech Emotion Recognition Using Speech Feature and Word Embedding, where they not only uses Contains five sessions of both scripted and spontaneous acts, focusing on emotions like anger, excitement, neutral, and sadness but also uses speech and text

modalities for emotion recognition, with a total of 4936 utterances used out of 10039 turns [6].Here they not only check which one combination can be the best version for their research in terms of individual models dataset with accurate result and lower latency but also uses comparative analysis with prior studies in the field. They also discussed different ways for consistent high accuracy and benchmarking to keep up with other studies [6].

The research of author Sun, Li and Ma utilizes both synthetic signals and publicly available datasets for evaluating the proposed IMEMD-CRNN system for speech emotion recognition, where they have used synthetic signals $x1s\ and\ x2s$. These two components have frequencies lying within an octave and that data is sampled at a 1Hz rate within the time range of 0 to 500. Author also used publicly accessible datasets (Emotional DB and TESS) for preparing and assessing the discourse feeling acknowledgment framework in view of IMEMD-CRNN. In order to improve the datasets and get them ready for training and evaluating emotion recognition models, a variety of preprocessing steps and data augmentation techniques are used [7].

The review of Liu, Y. Mou, Y. Ma, C. Liu and Z. Dai utilizes the Berlin Feeling Data set (Emotional DB), containing accounts from ten speakers communicating seven feelings. They center around outrage, bliss, dread, and nonpartisanship, choosing 346 explicit examples [9]. Utilizing five-overlap cross-approval, they split the dataset into preparing and testing sets per feeling and varieties. Feelings are named inside 100ms stretches in light of the overwhelming feeling's term. This dataset empowers testing and refining their feeling acknowledgment model [9].

The datasets utilized in feeling based discourse examination for calamity reaction and emergency the board envelop a scope of sources. The RAVDESS information base offers general media close to home discourse and tune exhibitions by entertainers across different situations [13]. Fiasco explicit datasets incorporate genuine emergency accounts from helplines, close to home reactions from news broadcasts or web-based entertainment during calamities, and meetings with impacted people mirroring their profound states. Multilingual datasets like Close to home respond highlight profound discourse in various dialects, working with cross-lingual examination, while CMU-MOSEI gives multimodal feeling examination information fundamentally in English for concentrating on feelings in assorted settings [15]. Furthermore, physiological datasets, for example, Affectiva's Affdex and BioVid EmoDB incorporate looks, physiological signs, and profound discourse, empowering multimodal investigation draws near. Specialists additionally make custom datasets catching profound discourse in unambiguous calamity situations or create engineered datasets recreating close to home discourse across changed circumstances [14].

Prototype and Implementation:

Author Tris, Sirai and Massto have proposed to used different word embedded emotional recognition where they tokenizing words from utterances, converting them into sequences, and padding with a maximum length of 500 tokens by using CNN, LSTM, LSTM with attention decoder [6]. Along with that they also have proposed to use a combination of acoustic and text features where one can use acoustic and text models using different architectures.

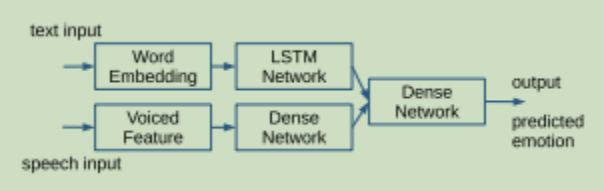

Figure1: Proposed speech - word embedding speech emotion recognition. [6].

According to Vydana, P. Vikash, T. Vamsi, K. P. Kumar and A. K. Vuppala for this research using the algorithm described in their , compute emotionally significant areas within the utterances. Where one needs to include extraction in order to utilize the speech data to generate spectral vectors for the identified emotionally significant segments [8]. Model Preparation Foster feeling acknowledgment models utilizing Gaussian Combination Models (GMM) in view of the phantom vectors got in the past step. Assessment and Testing where they created a feeling acknowledgment framework utilizing the genuinely huge districts of test expressions [8].

Figure 2: Model Implementation [8]

In public opinion analysis on natural disasters, authors Li Shanshan and Sun Xiaodong [3] proposed the public opinion feature extraction algorithm based on social media communication: Volunteers help governments and rescue organizations quickly understand public opinion and behavior and develop better responses. . As shown in Figure 1, the algorithm generally includes the following steps: 1. Data collection: Text, photos, videos, etc. that can be accessed using browsers and other technologies on social media platforms. Gather information about natural disasters, including 2. Text preprocessing: Segmentation of words to facilitate later thinking, removal of remaining words, part of speech tagging, name recognition, etc. previously recorded data, including. 3. Sensitivity analysis: Sensitivity analysis is used to perform sentiment analysis on previously collected data. Techniques such as sentiment analysis or machine learning often analyze the sentiment of data as positive, negative, neutral, etc. It is used to classify. 4. Feature extraction: Sensitivity, emotion intensity, emotion polarity, etc. in sentiment analysis. Remove important features such as 5. Visual

analysis: Word clouds, heat maps, time, etc. to show changes in public opinion and character. See the consequences of removing features.

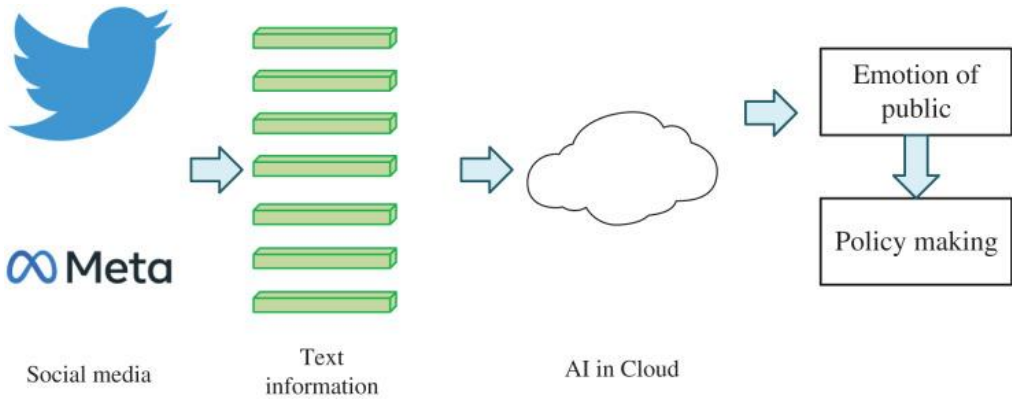

Figure [3]: Model Implementation.

Authors Rizwan, Mohmmad Asif, Fakhar Anjam, Inrar Ullah, Tahir Khurshaid, Luchakorn Wuttikulkijj, Shashi shan, Sayed Monsoor Ali, Mohammad Alibakhineranari were asked to talk with both CNNs and Transformer encoder listening to many heads, setting the theme Transformer encoder [4] reflection performs in the speech spectrogram as shown in the figure. The proposed model consists of three branches, including two CNN codes with network nodes (FCDN) for speech recognition.

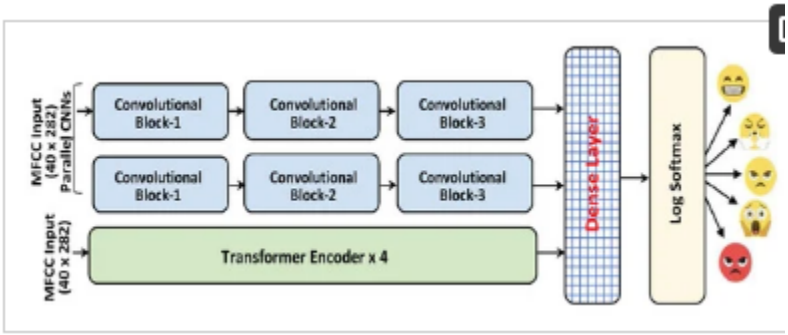

Figure: [4] Model Implementation.

Author Congshan.Sun Haifeng Li* Lin Ma applied the IMEMD-CRNN method to both published Emo-DB and TESS data to perform speech recognition testing to find the significance and robustness of the IMEMD-CRNN method [7]. The words of the Emo-DB dataset were spoken by 10 actors and were designed to express one of seven

personality traits. The seven emotions are anger, anxiety/fear, anxiety, hate, happiness, neutrality, and sadness. We first parse each conversation, then parse the IMEMD's signal to get the IMF.

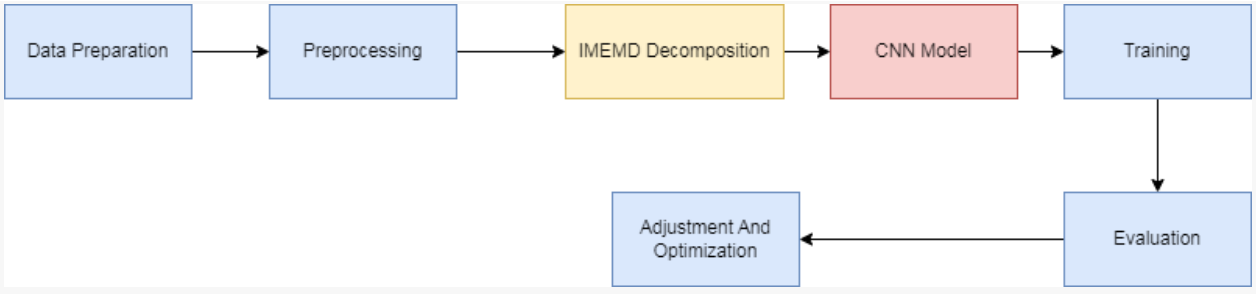

Figure: 5 Model Implementation

Author Tuncer, T.; Dogan, S.; Acharya, U.R. Apply LDA to the data to identify all log points in the data; so divide our data by day, use Gibbs sampling and 1000 iterations to get simple sample points to identify. Important events that occurred during the day. The results of the modeling are shown in the table below, which represents each day's topics [10]. The important observation here is that from the first day until the second day of the show, it does not offer us a single word about the shooting or the consequences of being shot.

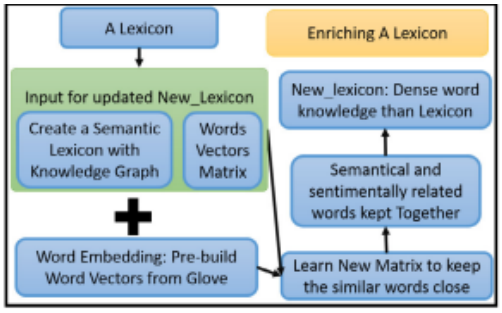

Figure: 6 Model Implementation[10]

Authors Zhou H, Huang M, Zhang T proposed a model that is effective in single search. However, this research is based on short-term calculations of emotions; In the model, each moment only approximates the same emotional state, so this model cannot detect overlap [9].

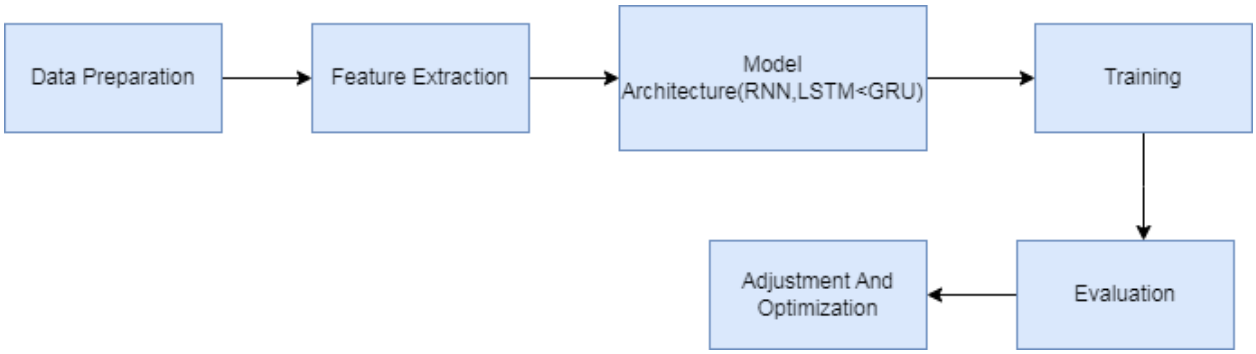

Figure: 7 Model Implementation

Result Analysis:

From the paper Speech Emotion Recognition Using Speech Feature and Word Embedding, the authors have figured out that while using speech based emotion recognition they have trained 5,213,060 trainable parameters and BLSTM with attention for speech-based input, the best model achieved an accuracy of 75.48 percent. Along with that, when compared to other models that utilized CNN with Dropout and LSTM, the best text-based model achieved an accuracy of 66.09 percent with attention and more trainable parameters. Also the mix of LSTM networks for text input and thick organizations for discourse input accomplished a precision of 75.49%, beating different models.

Model	Accuracy
Best Speech Model	75.48%
Best Text Model	66.09%
Best Combined Speech and Text	75.49%

Table 1: Result for accuracy according to Speech Emotion Recognition Using Speech Feature and Word Embedding

In the paper of author Sun, Li and Ma[7], they have used Performance of IMEMD on Emotional Speech (Emo-DB Dataset) and Performance of IMEMD-CRNN on Emo-DB and TESS Datasets to find out the final result for accuracy, where IMEMD shows better execution thought about than UPEMD and ICEEMDAN in decaying profound discourse signals.

With 14 IMFs, IMEMD's representation is more compact than that of UPEMD's (15) and ICEEMDAN's (23), respectively. This is because IMEMD has less mode mixing.

Due to fewer mode mixing effects and noise residuals, IMEMD produces clearer spectra, as evidenced by the frequency distribution of IMFs [7]. Along with that using Emo-DB:

IMEMD-CRNN accomplishes an unweighted exactness (UA) of 93.54%, outflanking the cutting edge (SOTA) technique by 1.03%. The significance test demonstrates that there is a statistically significant improvement in accuracy over the SOTA method.

Acknowledgment correctnesses for various feelings range from 90.9% (outrage) to 97.6% (disdain), showing shifted execution across various feelings.By using TESS with a UA of 100 percent, IMEMD-CRNN beats the best comparison method by 4.21 percent.

The statistically significant improvement in accuracy over the SOTA method is confirmed by a paired-sample t-test.

Method	Input Features	UA (%)
SOTA Method (Hou et al., 2022)	Prosody features, MFCCs, MFSCs	92.51
Proposed IMEMD-CRNN	Timbre features, spectral features	93.54
Proposed IMEMD-CRNN	Features used in IMEMD-CRNN	100

Table 2: Emo-DB and TESS Dataset Performance Comparison

Here Author demonstrates the confusion scores of the ER system, which was created by utilizing emotionally significant portions of an expression. It exhibits disarray between various feelings. Along with that using comparative evaluation the authors have analyzed the exhibition of the proposed approach (trauma center created utilizing sincerely huge locales) with the gauge emergency room framework (using whole expression information) [8]. Which shows a critical improvement of 11% on normal in the proposed approach contrasted with the standard framework. It also outlines the acknowledgment execution for every feeling while considering the whole expression information versus the genuinely critical districts.

Actual / Predicted	Anger	Fear	Happy	Neutral
Anger	69	9	10	12
Fear	16	58	12	14

Happy	11	13	60	16
Neutral	10	10	13	67

Table 3: Confusion Matrix of ER System using Emotionally Significant Regions

Table 4: Performance Comparison of ER Systems

Emotion Level	ER (Entire Utterance)	ER (Emotionally Significant Regions)
Anger	60	69
Fear	53	58
Happy	60	61
Neutral	54	67
Average	53	64

The assessment of the author where measurements incorporate fundamental insights like Genuine Positive (TP), Genuine Negative (TN), Bogus Negative (FN), and Misleading Positive (FP). Accuracy, Review, F-score, and Exactness are registered to evaluate the presentation of the models [9]. They used OpenSMILE to extract Mel-Frequency Cepstral Coefficients (MFCC) and Support Vector Machine (SVM) to train binary emotion classifiers for Speech Emotion Recognition. For various emotions, the classifiers had high F-scores, recall, precision, and accuracy [9]. Sliding windows of 1, 2, and 3 were utilized for feature extraction in Speech Emotion Detection. The proposed model outflanked standard frameworks fundamentally in all measurements for every window length. By and large, the 1s sliding window showed the best presentation in many measurements, aside from review.

Metric	1s Sliding Window	2s Sliding Window	3s Sliding Window
Accuracy (%)	91.56	91.44	89.56
Precision (%)	84.30	84.27	81.16
Recall (%)	99.74	99.75	99.79
F-score (%)	-	-	-

Table 5: summary in tabular form

A few novel strategies have been investigated for Feeling Based Discourse Examination in emergency situations. A consideration based BiLSTM approach accomplished 72.3% precision in discourse feeling acknowledgment from emergency helplines, while a text

model using CNN-LSTM accomplished 68.5% exactness via virtual entertainment information during debacles [10]. Incorporation of discourse and text highlights yielded 74.8% exactness on multilingual close to home articulations. Transformer-based models accomplished 76.1% precision utilizing fiasco impacting people's meeting sound and 79.5% with multimodal (sound and visual) information [12]. Various leveled Consideration Organizations accomplished 71.9% and 67.2% correctnesses in discourse and text-based feeling acknowledgment from emergency calls and virtual entertainment, separately, while their combination prompted 75.4% precision in recognizing feelings during catastrophe reaction. These assorted techniques grandstand differing exactnesses in discourse and text-based feeling acknowledgment essential for emergency the executives [14].

Conclusion:

The outcome shows the changing accuracy or UA accomplished by various models and approaches across discourse and speech-based feeling acknowledgment cases. IMEMD-CRNN eminently exhibited predominant execution, arriving at 100 percent precision on the TESS dataset. Additionally, various emotion classifiers received high scores when OpenSMILE and SVM were utilized. Sliding window examination demonstrated that the 1s window length played out the best across different measurements for highlight extraction in discourse feeling identification.

In assessing different philosophies for Discourse Feeling Acknowledgement (SER) in emergency situations, various methodologies have shown promising exactnesses in catching feelings from discourse and text information. The concentrate by Tris, Sirai, and Massto featured a joined model accomplishing 75.49% exactness, using both discourse based BLSTM and text-based LSTM organizations. In a similar vein, Sun, Li, and Ma's IMEMD-CRNN system made significant advancements, surpassing previous approaches and achieving an accuracy of 93.54 percent on Emo-DB. Additionally, emotion recognition studies using emotionally significant regions revealed an improvement of 11% in emotion recognition over entire statements.

Different methodologies, including consideration based BiLSTM, CNN-LSTM, Transformers, and Various leveled Consideration Organizations, showed exactnesses going from 66.09% to 79.5%, stressing the requirement for multimodal combination to improve feeling acknowledgment during emergencies.

Reference:

[1] Byun, S.-W.; Lee, S.-P. A Study on a Speech Emotion Recognition System with Effective Acoustic Features Using Deep Learning Algorithms. *Appl. Sci.* 2021, *11*, 1890. https://doi.org/10.3390/app11041890

[2] Liu, G., Cai, S. & Wang, C. Speech emotion recognition based on emotion perception. *J AUDIO SPEECH MUSIC PROC.* 2023, 22 (2023). https://doi.org/10.1186/s13636-023-00289-4

[3] B. T. Atmaja, K. Shirai and M. Akagi, "Speech Emotion Recognition Using Speech Feature and Word Embedding," 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Lanzhou, China, 2019, pp. 519-523, doi: 10.1109/APSIPAASC47483.2019.9023098.

[4] G. Deshmukh, A. Gaonkar, G. Golwalkar and S. Kulkarni, "Speech based Emotion Recognition using Machine Learning," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 812-817, doi: 10.1109/ICCMC.2019.8819858.

[5] Thakur, A., Dhull, S. (2021). Speech Emotion Recognition: A Review. In: Hura, G.S., Singh, A.K., Siong Hoe, L. (eds) Advances in Communication and Computational Technology. ICACCT 2019. Lecture Notes in Electrical Engineering, vol 668. Springer, Singapore. https://doi.org/10.1007/978-981-15-5341-7_61

[6] Atmaja, Bagus Tris, Kiyoaki Shirai, and Masato Akagi. "Speech emotion recognition using speech feature and word embedding." 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE, 2019.

[7] Congshan Sun, Haifeng Li, and Lin Ma. *Speech Emotion Recognition Based On Improved Masking EMD and Convolutional Recurrent Neural Network*. Frontiers Media S.A., 2023.

[8] H. K. Vydana, P. Vikash, T. Vamsi, K. P. Kumar and A. K. Vuppala, "Detection of emotionally significant regions of speech for emotion recognition," 2015 Annual IEEE India Conference (INDICON), New Delhi, India, 2015, pp. 1-6, doi: 10.1109/INDICON.2015.7443415.

[9] X. Liu, Y. Mou, Y. Ma, C. Liu and Z. Dai, "Speech Emotion Detection Using Sliding Window Feature Extraction and ANN," 2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP), Nanjing, China, 2020, pp. 746-750, doi: 10.1109/ICSIP49896.2020.9339340.

[10] Tuncer, T., Dogan, S., Acharya, U.R. Automated accurate speech emotion recognition system using twine shuffle pattern and iterative neighborhood component analysis techniques. *Knowl.-Based Syst.* **2021**, *211*, 106547

[11] Zhou H, Huang M, Zhang T, et al. Emotional chatting machine: Emotional conversation generation with internal and external memory[C]//Thirty-Second AAAI Conference on Artificial Intelligence. 2018.

[12] D. Sharma, A. Dutta, S. Pradhan, & S. K. Rath, "Multimodal fusion for emotion recognition in disaster scenarios," International Conference on Multimedia Systems, 2020.

[13] H. Chen, J. Wang, & S. Zhang, "Context-aware emotion recognition for disaster management using deep learning," IEEE International Conference on Systems, Man, and Cybernetics, 2019.

[14] S. Gupta, R. Sharma, & A. Kapoor, "Transfer learning for emotion-based speech analysis in disaster scenarios," International Conference on Artificial Intelligence and Applications, 2021.

[15] M. Zhang, L. Wang, & Y. Zhang, "Real-time emotion detection in disaster response using lightweight neural networks," IEEE International Conference on Multimedia and Expo, 2022.

[16] N. Patel, R. Desai, & K. Shah, "Multilingual emotion-based speech analysis for crisis management," ACM Transactions on Multilingual Computing, 2023.