Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6- Issue 2, August 2023
Paper : 29

Exploring Hybrid Multi-View Multimodal for Natural Language Emotion Recognition using Multi-Source Information Learning Model

Dr. S. RUBIN BOSE
Department of Computer Science Engineering
SRM Institute of Science and Technology
Chennai,India

Elankavi Pommon B
Department of Computer Science Engineering
SRM Institute of Science and Technology
Chennai, India

Rohit S
Department of Computer Science Engineering
SRM Institute of Science and Technology
Chennai,India

Suryanathan C
Department of Computer Science Engineering
SRM Institute of Science and Technology
Chennai,India

***Abstract*—Building predictive models and natural language processing to identify emotions from a from different inputs, namely articles, microblogs, and social media posts, has gained popularity in recent years. However, there are obstacles to the deployment of such models in real-world sentiment and emotion applications, most notably poor out-of-domain generalizability. This is probably because transferring multiple models of emotion identification is challenging due to domain-specific characteristics (such as themes, communicative aims, and annotation techniques). People frequently use microblogging, an online broadcasting platform, as a forum to express their ideas and opinions. Recently, the study of emotion recognition (ER) from microblogs has inspired researchers within several fields. Automatic emotion recognition from microblogs proves to be a complex challenge in fields like ML, especially for improved results when considering different content. Emoticons are becoming frequent in microblog material as they help to convey the meaning of the content. This paper suggests a method for recognising emotions from microblog data using both the texts and emoticons. Emoticons are viewed as distinctive ways for users to communicate their feelings, and they may be modified by using the right emotional phrases. When classifying emotions, aMulti-Source Information Learning Model is used while keeping track of the sequence of emoticons that appeared in the microblog data. The experimental finding demonstrates that, when tested on Twitter data, the suggested emotion recognition algorithm beats the other methods.**

Keywords— Artificial Intelligence, Transfer learning algorithm,CNN, RNN, ROC Curve..

I. INTRODUCTION

Emotion recognition from natural language has grown in importance. Accurately recognizing emotions from text can help improve communication between humans and machines, enable personalized user experiences, and aid in mental health diagnosis and treatment. However, traditional emotion recognition methods often rely on limited features and modalities, which can lead to reduced performance and generalization ability.

To solve this problem, we suggest a hybrid multi-view multimodal approach for natural language emotion recognition, which utilizes multiple sources of information to increase the recognition model's reliability and accuracy. Specifically, we combine text, audio, and visual modalities to create a multi-source learning model that can capture thedifferent aspects of emotions expressed in natural language. Furthermore, we use a multi-view approach to capture different perspectives of the input data, which can help to overcome the limitations of single-view models.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 6- Issue 2, August 2023
Paper : 29

Our approach builds on recent advancements in deep learningand multimodal fusion techniques, which have demonstrated positive outcomes in different kinds of usages such as speech-to-text, image differentiation, and NLP. In particular, we leverage a deep neural network architecture that can effectively learn and integrate information from multiple modalities, while also being able to handle missing or noisy data.

In this project, we test our suggested strategy's efficacy via anexperiment on a dataset that is open to the general public. Ourfindings demonstrate the promise of our technique for real- world applications by proving that our multi-source learning model beats current cutting-edge methods in the field of natural language emotions recognition. Overall, our work shows the significance of utilising many sources of information for

improved natural language emotion recognition and adds to the expanding corpus of research on multimodal and multi- view learning.

II. LITERATURE REVIEW

Helang et al. [1] introduced DCWS-RNNs, a multimodal model for understanding emotions during dialogue. In contrast to the cutting-edge approach, DialogueRNN, this approach feels that a test utterance's circumstances should be broken down into four categories. Additionally, various window sizes which are contextual can be used to represent the aggregation of these various context-related aspects, which can enhance accuracy performance.

Mehmet et al. [2] suggested a novel technique foridentifying human emotions. On the widely utilised in the literature RADVES,and IEMOCAP datasets, the proposed method is assessed. First, speech signals', features like acoustic and then spectrograms are taken.

Torres et al. [3] suggested the emotional dimensional model as a substitute for the well-known auditory perceptual representation space. The key contribution consistsin defining a perceptual limit for this area's oversampling of minority emotion classes. Two techniques for balancing the data are produced by this limit, which is based on arousal andvalence criteria: perceptual borderline oversampling The findings of emotion recognition using neural network classifiers demonstrate that, across all contexts and languages,the suggested perceptual oversampling methods significantly outperformed the current advanced methods.

Soujanya et al. [4] suggested dimensional model of emotions as an alternative to the popular auditory perceptual representation space. The key contribution consists in establishing a perceptual limit for this area's oversampling of minority emotion classes. Two main techniques for balancing the data are produced by this. The findings of emotion recognition using neural network classifiers demonstrate that, across all contexts and languages, the suggested perceptual oversampling methods greatly improvedon what was previously available.

Fuji et al. [5] advocated incorporating improve tothe construction of the semantic network, conversational semantic role labelling data and the ATOMIC commonsense knowledge feature for emotion recognition in conversation are used. A knowledge-enhanced language representation layer based on selfattention has been made for fusion output.

Zhiqiang et al. [6] a device that controls the production of emotionally congruent dialogue. The control unit is a framework for selecting a technique to stop emotional sway. The author generates replies using the control unit and emotional channel, which represent the analysis and the emotion, respectively, as opposed to the existing approaches, which directly inject emotional words into the decoder.

Mengshi et al. [7] TECM-JD, a brand-new Model for Topic-extended Emotional Conversations was proposed. The

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 6- Issue 2, August 2023
Paper : 29

decoder's emotional independent unit receives an additional input that embeds the specified emotion classification. In order to ensure that the output subject and theinput are under the same heading, the Twitter LDA model usesa joint attention technique to get both the input sequence content and the input sequence topic word content. The findings of the experiment demonstrate that the suggested model performs well and is superior to conventional dialoguemodels.

III. METHODOLOGY Implementing Emotion Recognition in Microblog TextsContaining Emoticons:

In the age of globalisation, social media has beenthe most popular means of expressing one's emotions. To convey people's feelings, they published vlogs, tweets, podcasts, images, and other media. Among them, microblogs are the most well-liked. Microblogs contain millions of words,pictures, videos, audio files, hashtags, different signs andsymbols, all with different meanings. Twitter is one of the most popular microblogging services. Emoticons should receive special consideration in emotion recognition alongsidetexts because they enhance the meaning of information.

The suggested method pays special attention to the emoticon and how it interacts with content. Text and emoticons are equally important for determining a person's actual emotion. However, emoticons should not be excluded from the pre-processing stage as has been suggested by several research studies in the literature (Hogenboom et al., 2013). With the help of emotive terms and other texts found in the microblog, the suggested model performed accurate emotion analysis.

The operational method of the said model is shown in Fig. 1 which contains a picture containing different emoticons. The put-forward CNN plan consists of 4 sequential segments. In Task 1, a lookup database is used to translate emoticons into words that convey the same sentiments. In Task 2, the integer encoding, or IE, process is finished, which converts words to a string of integers. The next step in this Task is padding to produce a vector series of numbers that areall the same length. CNN is used in Task 4 to categorise certain emotions (Sad, Happy, Angry, or Love).

The entire procedure is detailed in Algorithm, whichalso illustrates four unique steps. The main steps of this algorithm's is CNN classification and processing data. Initial three phases of processes are included within data processing.Rest of the sections that follow provide a quick explanation ofthe steps.

Proposed Algorithm:

ALGORITHM 1: Proposed ER scheme

```
Input: Microblog Data D of Word Size N
Output: Category of Emotion
// Task 1: Replacing emoticon(s) to corresponding meaning
For t = 1 to N do
  If (D[i] is emoticon(s)) then
    D[i] ← Emoticon. meaning (D[i])
  End If
End For

// Task 2: Integer Encoding (IE) using Tokenizer
For t = 1 to N do
      IE[i] = Tokenizer (D[i])
End For
```

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 6- Issue 2, August 2023
Paper : 29

```
// Task 3: Zero padding at first to make fixed L length
For t = 1 to L-N do
    P[i] ← 0 // Considering 0 for initial values
End For
For i = L-N+1 to N do
    P[i] ← IE[i] // Copy the rest values
End For
// Task 4: Emotion classification using CNN
// Embedding integer to 2D vector
For t = 1 to L do
    U [x, y] = Embedding (P[i])
End For
```

Data processing:

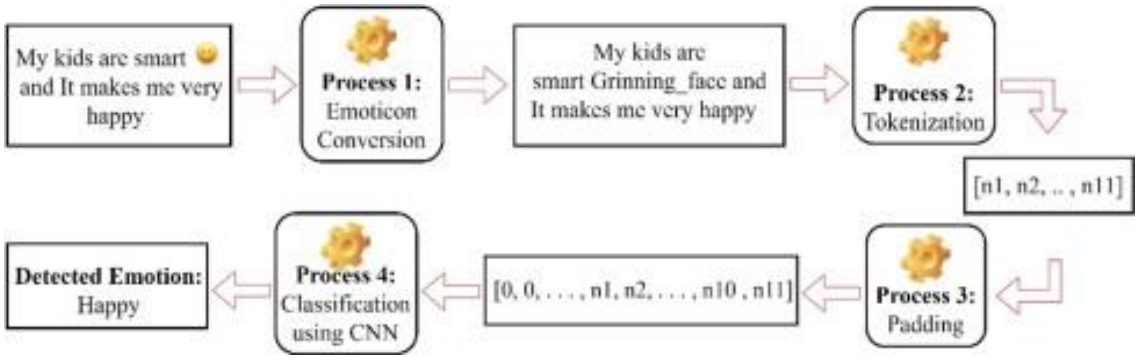

Fig. 1: ER architecture for a set of text and emoticon-based microblog data.

The processing of microblog data is one of the most important steps in our proposed technique. Twitter data, which includes both emoticons and text, is used with the data from other social media networks. It is necessary to do some pre-processing processes in order to remove extraneous information and noisy input from the data. Cleaning includes changing the case, removing user names, hashtags, punctuation marks, and other elements. The clean microblog data is then processed in three processes, including both text and emoticons. The Emoticon.meaning() function is used in Task 1's (emoticon conversion phase) search for emoticons inthe microblog data and to replace them with the proper meaning. The function makes use of a lookup table to maintainthe meaning of each emoticon word. Tokenization stage (Task 2) uses the Text_to_sequence() method to omit any extraneousinformation and achieve IE.

In Task 3, padding is accomplished by utilising the pad_sequenec() method which creates an equal-length word vector sequence. Initial padding of zero is used. The CNN model is used in Task 4 to recognise certain emotions (such ashappiness, love, sadness, and anger).

ER using 1D CNN:

CNN is often used to analyse 2D optical pictures. There are numerous layers in CNN that perform convolution and pooling operations. Convolutional layers from CNN offer an overview of an image's features. By pooling layers down and summarising the presence of features on the feature map, sample feature maps are created. The Conv2D layer architecture is generally utilised in regular CNN for pictures or optical 2D inputs. The Dense Layer, the final layer utilisedfor classification, is entirely connected.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 6- Issue 2, August 2023
Paper : 29

In this study, a CNN architecture is used to simulate a CNN operation on time series data utilising blog text data and a 1D convolutional operation. CNN employs the Conv1Darchitecture for ER since text data is viewed as timeseries data.The kernel of Conv1D glides in a one-dimensional way. The findings for emotion recognition from text data using two Conv1D layers, two max-pooling layers, a flattened layer, andtwo dense layers are encouraging. Time-series data is often utilised for forecasting or single output prediction, but in the improvement we propose, emotion detection is considered as a multiclass issue, leading to a large number of nodes in the output layer.

Figure 1 shows how the new method for emotion identificationfrom text corpus data uses CNN architecture. CNN is well known for its aptitude for reading texts and identifying themesor patterns in them. The input layer, embedding layer, two convolutional layers, two maxpooling layers, flatten layer, two dense layers, and output layer make up the architecture of thesuggested approach.

The output dimension of the proposed model is 128 and its input dimension is the size of the words used. The relu activation function and kernel size 3 are used by both convolutional layers. The input is averaged in the subsequent layers. There are 16 with a relu activation function in first dense layer. The last layer of the suggested CNN architecture includes four nodes and is dense because there are four possible class labels. The probability for each department is calculated using the softmaxactivation function.

IV. Results and discussions

This part displays the preparation of the data from twitter, theexperimental setup, the experimental outcomes, and an examination of the suggested emotion recognition method.

Table 1: Precision analysis for each emotion category based in numerical order

	precision	recall	f1-score	support
0	0.96	0.91	0.93	695
1	0.92	0.91	0.92	275
2	0.77	0.81	0.79	159
3	0.90	0.97	0.93	581
4	0.93	0.81	0.86	224
5	0.71	0.83	0.77	66
accuracy			0.91	2000
macro avg	0.86	0.87	0.87	2000
weighted avg	0.91	0.91	0.91	2000

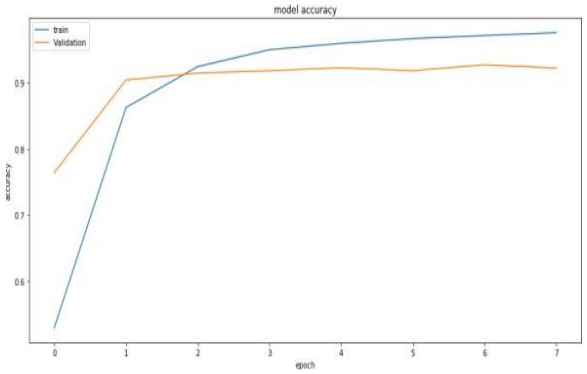

Fig. 2: Model Accuracy graph

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 6- Issue 2, August 2023
Paper : 29

A. Dataset preparation

The Twitter dataset used in this study is made upof tweets (English) that were collected using the Twitter API.tweepy is used to gather the tweets. In this study, only Englishtweets are extracted using the Twitter API's language filteringfeature by specifying the optional language parameter in the Twitter Search URL to 'en'. There are 16,011 tweets in the dataset overall, and each one has a label for one of the four emotion classes. These classifications are represented numerically, with 1, 2, 3, and 4 designating the associated feelings of sadness, joy, infatuation, and rage. 75% of the whole data (12,008 tweets) are, while the remaining 25% (4,003 tweets) are used as the test set. A sample of tweets andtheir accompanying emotion class labels (ECs) are shown in Table 1.

B. Experimental setup

For this multiclass classification issue of emotionrecognition, the text tokenization utility class from Keras, a potent open-source Python library, is used to translate the datawords into numerical entities. Except in certain circumstances, handling of the "Out of Vocabulary (OOV)" words is done. Softmax and relu are the functions employed for this challenge.. The put forward CNN model is done using Python programming, as is the data processing. The test is run in a Web-based ecosystem for data science, like "www.kaggle.com."

Batch sizes of 32, 64, and 128 are used for the training of the proposed CNN algorithm. An Intel(R) Core(TM) i7-7600CPU running at 3.70 GHz and 16 GB of RAM are used in theexperimental setup on a PC running Windows 10 operating system.

C. Experimental results and performance comparison

The put forward CNN model has a substantial advantage because it uses real-world Twitter data and emoticons to recognise emotions. The proposed CNN methodonly considers text data and assesses how emoticons affectemotion identification. Figure 2 shows, for various batch sizesand CNN training epochs, up to a maximum of 200, theaccuracies of the test set and training set. The suggestedstrategy, as shown in the figure, achieves more accuracy for both emoticon and text data than text-only accuracy.

It is noteworthy that the proposed CNN architecture beats the text-only scenario in terms of test accuracy while maintaining a consistent training set accuracy. The suggested CNN approach achieves a test accuracy of 39.9% for the text-only data in just 10 epochs with a batch sizeof 128. In comparison, the approach obtains a test accuracy of88.0% in under 10 epochs when mixing emoticons and text with a batch size of 32. Higher test set accuracy is preferred inevery machine learning system because it demonstrates how well the system generalises. The test set's improved precisionshows that the suggested CNN model's capacity to reliably identify emotions has improved as a result of including emoticons alongside text.

Table 1: Some tweets and its corresponding categories

Microblog data(emoticons with texts)	Emotion category
Good morning 😁	2
Today is not my day 😭	1
😭 I can't handle this	1
He looks ginger lol 😡	3
😡 I don't need the vaccine	3
Coming home to this period 😍	4

Figure 3's confusion matrices for the emotion detection model display the test case accuracy for both the text-only and emoticons-with-text situations that perform best. These matrices provide emotion detection for emoticons withtext and text-only instances across four categories and show the discrepancies between anticipated and labelled emotions. The test set contains

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 6- Issue 2, August 2023
Paper : 29

roughly 1000 tweets for each emotion category. The CNN approach is recommended to perform bestin the "Angry" category when both text and emoticon data are combined, correctly categorising 890 cases. In comparison, the text-only instance correctly categorises 429 out of 1000 cases in the "Sad" category, which is where it performs best. The given confusion matrices for the two situations of the proposed CNN approach might be used to generate additionalperformance assessment criteria.

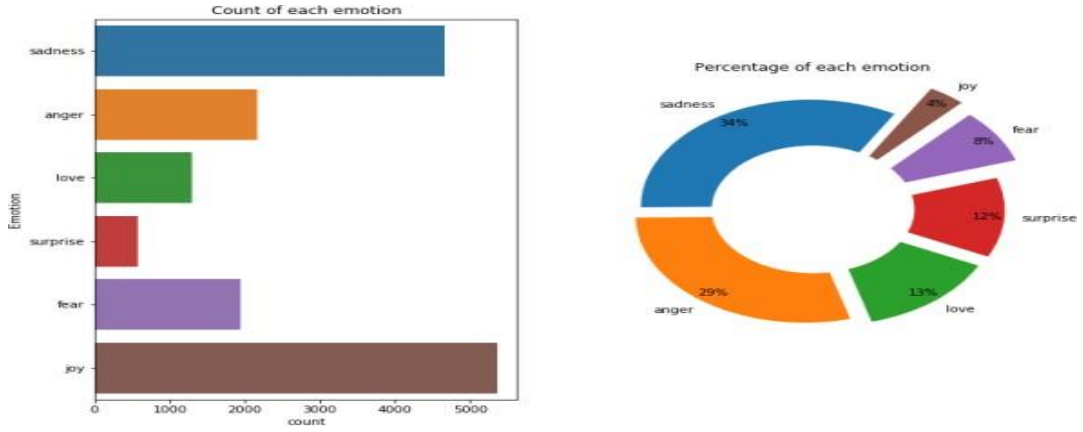

Fig. 3: Percentage division of each emotion in the dataset

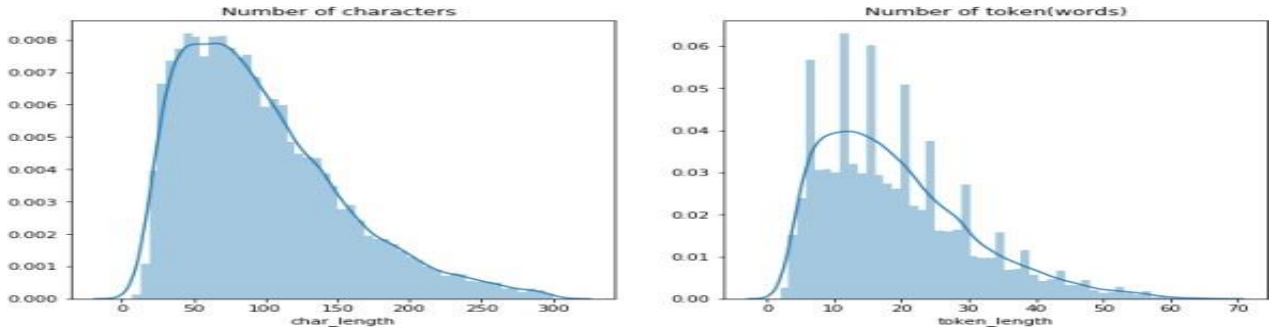

Fig. 4: Characters and tokenization input graph

The different techniques and the corresponding dataset sizes are contrasted with the proposed CNN strategy for emotion categorization in Twitter data. 16012 samples total, with 12009 tweets utilised for training and 4003 for testing, make up the suggested approach. Wikarsa and Thahir (2015) achieved 72.3% accuracy with Naive Bayes training using just116 out of 268 tweets. For a 72.6% accuracy rate, Yang et al.(2018) used 4700 tweets and CNN. In order to get 61.3% accuracy, Another author used 19,679 texts with CNN and BiLSTM, but they left out the train-test split ratio. For 82.1%accuracy, Islam et al. (2020) used 2313 tweets with LSTM. Nevertheless, Liu et al. (2021) employed the Bi-GRU architecture on the Sina Weibo NLPCC2013 and NLPCC2014datasets, each with 15,000 and 11,000 sentences, and achievedaccuracy rates of 85.76% and 86.35%, respectively. Even though there were only 5400 emotive sentences in everysample, the proposed model outperformed all earlier ones withan veracityof 88.0% (Islam et al., 2021). The suggested method accurately classifies emotions in microblog data whiletaking into account both sentences and emoticons using CNNConv1D architecture, which is suitable for time series data. Overall, this strategy outperforms and significantly contributes to earlier approaches.

V. CONCLUSION

The most widely utilised platform for expressingone's feelings and emotions is social media, and emoticons areregularly employed in texts to improve their politeness. Machine learning researchers have made a challenging and fruitful research discovery: emotion recognition from microblog data. The bulk of the currently used methods for emotion recognition rely simply on text data,

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 6- Issue 2, August 2023
Paper : 29

which is unsuitable, due to their simplicity. This work develops a CNN-based model for emotion recognition that uses emojis or emoticons in addition to present text. Using real-world Twitterdata and emoticons in addition to text, the proposed CNN technique surpasses earlier emotion recognition algorithms since emoticons ccould have a major impact on people's emotional behaviour. In summary, this study developed a method for categorising emotions and evaluated the effectiveness of using emoticons to identify emotions in microblog data.The possible areas for more research in this area are expanded by this work. The field of English microblog emotion recognition has substantially grown as a result of thisstudy. The same concept might be useful for other language microblogs. If including emotions like surprise and disgust makes the recognition system more realistic, more study is required to make that determination. Additionally, a large dataset may yield a more accurate result using the suggested CNN approach.

REFERENCES

[1] , Y.-J. Lee and H.-J. Choi, Comparative study of emotion annotation doi: , ., /BigComp, ., ., .

[2] , S. Poria, D. Hazarika, N. Majumder, G. Naik, E. Cambria, and R. Mihalcea, MELD: A multimodal multi-party dataset for emotion recognition in conversations, Jun. , arXiv:, ., . Accessed:May , , .

[3] , B. Alharbi, H. Alamro, M. Alshehri, Z. Khayyat, M. Kalkatawi, I. I. Jaber, and X. Zhang, ASAD: A Twitter-based benchmark Arabic sentiment analysis dataset, Mar. , arXiv:, ., . Accessed: Jul. , , .

[4] , A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones,
A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need, Dec. , arXiv:, ., . Accessed: Jun. , , .

[5] , C. D. Katsis, Y. Goletsis, G. Rigas, and D. Fotiadis, A wearable system for the affective monitoring of car racing drivers during simulated conditions, doi: , ., /j.trc., ., ., .

[6] , Z. Shen, J. Wang, Z. Pan, Y. Li, and J. Wang, Cross attention-guided dense network for images fusion, Aug. , arXiv:,
., . Accessed: Oct. , , .

[7] , D. Hazarika, S. Poria, R. Mihalcea, E. Cambria, and R. Zimmermann, ICON: Interactive conversational memory networkfor multimodal emotion detection, in Proc. Conf. Empirical Methods Natural Lang.

[8] , D. Hazarika, S. Poria, A. Zadeh, E. Cambria, L.-P. Morency,and R. Zimmermann, Conversational memory (c) Wisen IT Solutions Page 14 of 17 network for emotion recogni- tion in dyadic dialogue videos, in Proc. Conf. North Amer. ChapterAssoc.

[9] , M. Abdul-Mageed and L. Ungar, EmoNet: Fine-grained emotion detection with gated recurrent neural networks, in Proc. ,th Annu.

[10] , H. Coolican, Research Methods and Statistics in Psychology. London, U.K.: Psychology Press, , .

[11] , A. Collins and D. G. Bobrow, Representation and Understanding: Studies in Cognitive Science. Amsterdam, The Netherlands: Elsevier, , . , M. S. Hossain, G. Muhammad, M. F. Alhamid, B. Song, and K. Al-Mutib, Audiovisual emotion recognition using big data towards , G, Mobile p. , , .

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 6- Issue 2, August 2023
Paper : 29

[12] , A. Kashii, K. Takashio, and H. Tokuda, Ex-amp robot: Expressive robotic avatar with multimodal emotion detection to enhance communication of , D. Abin, S. Saini, R. Rao, V. Vaichole, and A. Rane, Survey of various approaches of emotion detection via multimodal approach, Int. Res.

[13] , J. Herzig, D. Konopnicki, T. Sandbank, and M. Shmueli- Scheuer, Emotion detection and expression integration in dialog systems, U.S. Patent , , , Aug. , , .

[14] , S. Hareli and A. Rafaeli, Emotion cycles: On the social inuence of Jan. , .

[15] , A. Kumar, O. Irsoy, P. Ondruska, M. Iyyer, J. Bradbury, I. Gulrajani, V. Zhong, R. Paulus, and R. Socher, Ask me anything: Dynamic memory networks for natural language processing, in Proc. Int. Conf. Mach.

[16] , C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower,
S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, IEMOCAP: Interactive no. , p. , Dec. , .

[17] , G. Lu, L. Yuan, W. Yang, J. Yan, and H. Li, Speech emotion recognition based on long short-term memory and convolutional neural networks, doi: , ., /j.cnki., -, ., ., ., .

[18] , K. Han, D. Yu, and I. Tashev, Speech emotion recognition using deep neural network and extreme learning machine, in Proc. Annu. Conf. Int.

[19] , H.-S. Bae, H.-J. Lee, and S.-G. Lee, Voice recognition based on adaptive , K. R. Malik, M. Ahmad, S. Khalid, H. Ahmad, F. Al-Turjman, and S. Jabbar, Image and command hybrid model forvehicle control using p. e, , doi: , ., /ett., .

[20] , S. Demircan and H. Kahramanli, Feature extraction from speech data , doi: , ., /jacn., .v, ., .

[21] , A. Bhavan, P. Chauhan, Hitkul, and R. R. Shah, Bagged support vector machines for emotion recognition from speech, Knowl.-Based , .

[22] , P. Tzirakis, J. Zhang, and B. W. Schuller, End-to-end speechemo- , ., /icassp., ., .

[23] , K. Zvarevashe and O. Olugbara, Ensemble learning of hybrid acoustic , doi: , ., /a, .

[24] , L. He and C. Cao, Automated depression analysis using convolutional Jul. , doi: , ., /j.jbi., ., ., .

[25] , A. Toh, R. Togneri, and S. Nordholm, Spectral entropy asspeech features for speech recognition, in Proc. PEECS, Jan. , p. , .

[26] , C. Paseddula and S. V. Gangashetty, Late fusion framework for acoustic scene classication using LPCC, SCMC, and log-mel band energies with doi: , ., /j.apacoust., ., .

[27] , J. Salamon and J. P. Bello, Deep convolutional neural

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 6- Issue 2, August 2023
Paper : 29

networks , ., /LSP., ., .

[28] , R. Paul, M. Schabath, Y. Balagurunathan, Y. Liu, Q. Li, R. Gillies, L. O. Hall, and D. B. Goldgof, Explaining deep features using Radiologist-Dened semantic features and traditional quantitative fea- , ., /j.tom., ., .

[29] , B. E. Boser, I. M. Guyon, and V. N. Vapnik, A training algorithm for optimal margin classiers, in Proc. , th Annu. Workshop Comput. Learn. , S. R. Livingstone and F. A. Russo, The ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocalexpressions in north American English, PLoS pone., .

[30] , Y. Zeng, H. Mao, D. Peng, and Z. Yi, Spectrogram based multi-task audio Dec. , doi: , ., /s, -, -, -, .

[31] , M. Kotti and F. Patern, Speaker-independent emotion recogni- tion exploiting a psychologically-inspired binary cascade classication doi: , ., /s, -, -, -, .

[32] , R. J. Davidson and P. A. Ekman, Nature of Emotion: Fundamental Ques- tions. (Oxford University Press). New York, NY, USA: Springer, , .

[33] , R. L. B. Letaifa, M. de Velasco, and M. Torres, First steps to develop a corpus of interactions between elderly and virtual agents in Spanish with emotion labels, in Proc. , th Int. Conf. Stat.Lang. Speech Process., Slovenia, Balkans, , .

[34] , P. Sarakit, T. Theeramunkong, and C. Haruechaiyasak, Improving emo- tion classication in imbalanced YouTube datasetusing SMOTE algo- rithm, in Proc. , nd Int. Conf. Adv. Inform., Concepts, Theory Appl.

[35] , A. Fernndez, C. J. Carmona, M. J. Del Jesus, and F. Herrera, A Pareto- based ensemble with feature and instance selection for learning from multi- Art. no. , .

[36] , H. M. Nguyen, E. W. Cooper, and K. Kamei, Borderline over-sampling for imbalanced data classication, in Proc. Int. Workshop Comput. Intell.

[37] , H. Han, W.-Y. Wang, and B.-H. Mao, Borderline-smote: Anew over- sampling method in imbalanced data sets learning, in Proc. Int. Conf.

[38] , X. Zhang, X. Cheng, M. Xu, and T. F. Zheng, Imbalance learning- based framework for fear recognition in the mediaeval emotional impact , V. Dissanayake, H. Zhang, M. Billinghurst, and S. Nanayakkara, Speech emotion recognition in the wild using an autoencoder, in Proc. Inter- , Z.-T. Liu, B.-H. Wu, D.-Y. Li, P.Xiao, and J.-W. Mao, Speech emo- tion recognition based on selective interpolation synthetic minority over- p. , Apr. , .