Vol. 02, No. 02, pp. 52 –58 (2021)
ISSN: 2708-0757

JOURNAL OF APPLIED SCIENCE AND TECHNOLOGY TRENDS

www.jastt.org

Multimodal Emotion Recognition using Deep Learning

Sharmeen M.Saleem Abdullah [1], Siddeeq Y. Ameen [2], Mohammed A. M.sadeeq[3], Subhi R. M. Zeebaree[4]

[1] *Duhok Polytechnic University, Duhok, Iraq, sharmeenbarwary2020@gmail.com*
[2] *Duhok Polytechnic University, Duhok, Iraq, siddeeq.ameen@dpu.edu.krd*
[3] *Duhok Polytechnic University, Duhok, Iraq, mohammed.abdulrazaq@dpu.edu.krd*
[4] *Duhok Polytechnic University, Duhok, Iraq, subhi.rafeeq@dpu.edu.krd*

Abstract

New research into human-computer interaction seeks to consider the consumer's emotional status to provide a seamless human-computer interface. This would make it possible for people to survive and be used in widespread fields, including education and medicine. Multiple techniques can be defined through human feelings, including expressions, facial images, physiological signs, and neuroimaging strategies. This paper presents a review of emotional recognition of multimodal signals using deep learning and comparing their applications based on current studies. Multimodal affective computing systems are studied alongside unimodal solutions as they offer higher accuracy of classification. Accuracy varies according to the number of emotions observed, features extracted, classification system and database consistency. Numerous theories on the methodology of emotional detection and recent emotional science address the following topics. This would encourage studies to understand better physiological signals of the current state of the science and its emotional awareness problems.

Keywords: *Emotion recognition, Facial recognition, Physiological signals, Deep Learning*

Received: February 19, 2021 / Accepted: April 14, 2021 / Online: April 17, 2021

I. INTRODUCTION

Recognition of emotions is a dynamic process that targets the emotional state of the person, which means that the emotions corresponding to each individual's actions are different [1, 2]. Human beings, in general, communicate their feelings in different ways[3, 4]. To ensure meaningful communication, accurate interpretation of these emotions is important [5, 6]. However, emotional recognition in our everyday lives is important for social contact, and emotions play an important role in deciding human actions[7, 8].

The various ways people communicate their feelings, both verbally and nonverbally, including expressive speech, facial gestures, body languages, etc [9, 10]. Therefore, emotional signals from multiple modalities may be used to predict the emotional state of a subject [11-13]. However, the single modal model cannot easily judge the consumer's emotion[15 ,14] . With cannot decide someone is emotional simply by looking at a particular entity or occurrence in front of our eyes [16, 17]. This is one reason why emotional recognition should be treated as a multimodal problem [18, 19].

Affective computation is implemented in the world of school [20, 21], smart cards, and the scene of motor vehicles [22], entertainment to health-care [23]. This applies in human-computer interface architecture, intelligent robotics, safety control [24, 25] etc.

For the last few years, deep learning [23, 26], relying on the most modern system, has achieved great success in many areas, such as signal processing, artificial intelligence, and emotion detection. Deep belief networks (DBN) [27, 28], convolution neural network (CNN) [29-31], recurrent neural networks (RNN) [32], these are the most commonly used approaches for deep learning [33].

In this paper, we present a review of the recent advancements in emotion research using multimodal signals; Feature extraction and classification methodologies using deep learning, particularly for emotion elicitation stimuli. This review aims to access and upgrade real-time emotion detection systems to know the latest progress in this technology. The latest 2019 and 2020 observations are

doi: 10.38094/jastt20291

discussed in our review of the topics and contributions.

The other portion of this paper is structured as follows: Section 2 outlines the concepts of recognizing human feelings. Section 3 explains how multimodal emotional processing signals are combined and provides a summary of recent research work. Section 4 offers a dispute and comparisons, then a general conclusion, in Section 5 of this emotional recognition study.

II. Emotion Recognition and Deep learning

A. Facial expression recognition

Facial gestures are important ways of expressing feelings in nonverbal contact. Facial expression recognition plays an important role in many applications, including human-computer interaction and health care [34, 35]. Mehrabian observed that 7% of knowledge moves between people through writing, 38% through voice, and 55% through facial expression [36]. Ekman et al. [37] defined six basic emotions: happiness, sadness, surprise, fear, and anger. He proved that human perceive these emotions regardless of their cultures [32]. Emotions could be expressed using two orthogonal dimensions: valence and arousal, as Feldman et al. [38]. He suggested that everyone has different ways of communicating their feelings. Moreover, there are strong variations in peoples' feelings when asked to express periodic emotions [39, 40]. The valence can be positive to negative, and the arousal can be calm to excited [41]. This work would categorize the input into its variations of valence and arousal [42]. Early methods of extracting facial expressions had been developed manually by developers by developing algorithms for extracted functions. such as Garbor wavelet, Weber Local Descriptor (WLD), Local Binary Pattern (LBP) [43], multi-feature fusion, etc. These features are not resilient against imbalances in subjects and may have a high loss of texture features from the original image [44]. The Deep Neural Network model application to facial expression analysis is the hottest subject these days in facial recognition [45]. FER also has a wide variety of social life uses, such as smart protection, lie detection, and smart medical practice[46],[47]. [27] reviewed facial expression recognition models based on deep learning techniques, including DBN, deep CNN, Long Short Term Memory (LSTM) [48, 49] and their combination.

B. Speech emotion recognition

T Speech expression recognition is one of the main elements of human-computer interface systems. They will communicate their feelings employing voice and face [50]. The speech recognition system is being widely used for the detection of emotion [12, 45]. The earliest experiments on emotion recognition in speech considered extraction of hand-crafted features of speech for classification. Liscombe et al. (2003) extracted a series of continuous speech features based on fundamental pitch, amplitude, and spectral tilt and evaluated its relationship with various emotions [51]. Various algorithms to recognize feelings in human speaking have been proposed [52] over the years. Many people proposed machine learning algorithms like Support vector machines, hidden Markov models, Gaussian mixture models, etc. Deep learning has been widely used in several speech domains, including speech recognition [50]. Convolution neural network has also been used for speech emotion recognition by [53]. [54] shows that using RNN bidirectional- (Bi-LSTM) is better for extracting essential speech characteristics for better speech recognition performance [55].

C. Multimodal emotion recognition

Multimodal emotion processing continues to see the widespread application in science [9]. This expansion would help better understand emotions with the experience of other related modalities of the study (video, audio, sensor data, etc. Many different approaches and strategies are integrated to meet the study goal. Many of them use big data techniques, semantic principles and deep learning [30].

III. Multimodal emotion recognition

Emotions are dynamic psychophysiological processes that occur nonverbally, which makes emotion identification complicated. Multimodal learning is a lot more efficient form of learning than unimodal one [56]. Studies also tried integrating signals from different modalities for better efficiency and precision, such as facial expressions and audio, audio and written text, physiological signals, and various combinations of these modalities [57]. At present, This technique was supplemented to increase the precision of emotion detection further. Multimodal fusion model can achieve emotion detection results by integrating physiological signals in various ways [58]. With the recent developments in Deep Learning (DL) architectures, deep learning has been applied [59] in multimodal emotion recognition. Deep learning techniques include deep belief net, deep Convolutional neural network, LSTM [60],support vector machine (SVM) [61],and their combination [27].

A. Multimodal emotion recognition combining (audio and text, image and text)

[Pan, Zexu et al.] [62] studied a hybrid fusion process, referred to as a multimodal attention network (MMAN), to make use of visual and textual signals in speech emotion recognition. They suggest a new multimodal focus mechanism, cLSTM-MMA, which promotes attention across three modalities and fuses information selectively. During late fusion, cLSTM-MMA is fused with other uni-modal subnetworks. The tests demonstrate that identifying speech emotions profits immensely from visual and textual signals. The suggested cLSTM-MMA alone is as successful in terms of precision as other fusion approaches but with a much more compact network structure.

[Siriwardhana et al.][63] searched the use of the pre-trained "BERT-like" architecture for self-supervised learning (SSL) to both represent language and text modalities in order to recognize the multimodal language emotions. They demonstrate that a basic fusion mechanism (Shallow-Fusion) simplifies the overall structure and strengthens complex fusion mechanisms.

[Priyasad et al.] [64] presented a deep learning-based approach to protect codes that are characteristic emotion. Through a SincNet layer, band-pass filtering technique and neural net, the researchers managed to extract acoustic

Identify applicable funding agency here. If none, delete this text box.

features from raw audio, and the output of said band-pass filters is then applied to the input to DCNN. A set of representations on the N-gram level is first determined in a bidirectional recurrent neural network, then in another recurrent neural network using cross attention before being combined as a final score.

[Krishna et al.] [50] introduced a new way of combining a raw-waveform-based convolutional neural network with cross-modal focus. They use raw audio processing by using one-dimensional convolutional models and attention processes between the audio and text feature to obtain the enhanced emotion detection. Their prototypes demonstrate the proposed architectures accomplish state of the art emotional classifications.

[Caihua] [24] through tests, he found that the SVM-based approach of machine learning is powerful for voice consumer sentiment analysis. He suggested an SVM-based multimodal speech emotion recognition. The experimental findings then reveal that the SVM algorithm has advanced dramatically by applying this SVM approach to the common database classification problem. Finally, he applies the approach to understand the emotional expression and produces an emotional recognition result with successful speech.

[Lee et al.] [65] developed a multimodal deep learning model that utilizes facial images and textual details explaining the circumstances. To classify the characters' facial images in the Korean TV series 'Misaeng: The Incomplete'1 into seven emotions: Rage, Disgust, Terror, Joyful, Neutral, Sad, and Surprise, Using photographs and text, they developed two multimodal models to identify emotions. The experiment findings indicated that using text definition of the behaviour of the characters dramatically increases recognition efficiency.

[Liu, Gaojun et al.] [66] A new multimodal music emotion grouping system was developed based on music audio quality and text for music lyrics. Use of the LSTM network for classification is suggested in terms of audio, and the classification effect is greatly increased relative to other machine learning methods. It is proposed to use Bert in terms of lyrics to describe the emotions of lyrics, which essentially addresses long-term dependency. LSFM is suggested in lyrics in terms of multimodal fusion. The emotion dictionary is used to alter lyrics' emotional classification outcomes. The neural network is implemented based on linear weighted decision-making stage fusion, which increases efficiency and precision.

Table I compares several works in terms of deep learning algorithms/ techniques, accuracy, and dataset used.

B. *Multimodal emotion recognition combining (facial and body physiological)*

[DHAOUADI et al.] [67] evaluate the introduction of (LSTM) and Deep Neural (DNN) networks in young gamers for real-time tension monitoring. The research was focused on their reactions to the body. Physiological signals such as electrocardiography (ECG), electrodermal activity (EDA) and electromyography (EMG), calculated by non-invasive wearable sensors, are used for this function. It can be seen from the results obtained that, in this case, the LSTM model was more effective and reliable than the DNN for the parameters they fixed.

TABLE I. COMBINING SIGNALS FROM AUDIO AND TEXT, IMAGE AND TEXT

Author	Neural network architecture and deep learning technique (algorithms)		Accuracy	Data set used
	classification	***method***		
[Zexu et al.] **[62]**	LSTM	MMAN ,Fusion method	73.98%	IEMOCAP
[Siriwardhana, et al]**[63]**	SSL modle	Speech-BERT, RoBERT Shallow fusion	—	IEMOCAP, CMU-MOSEI, CMU-MOSI),
[Priyasad, et al] **[64]**	DCCN with a SincNet layer, RNN	band-pass filters	80.51%	IEMOCAP
[Krishna et al] **[50]**	1D CNN	cross-modal attention	1.9% improvement	IEMOCAP
[Caihua] **[24]**	SVM	Fusion method	72.52%,	Berlin Emotional DB
[Lee et al.] **[65]**	CNN	Natural Language Processing (NLP)	—	Asian Character from the TV drama series
[Liu, et al] **[66]**	LSTM	Bert model, LSFM	5.77% improvement	777 songs(Music Mood Classification Data Sets)

[Yang et al.] [68] proposed a method of emotion detection based on electrocardiogram (ECG) and photoplethysmogram (PPG) signals as target sources of data entry. Three states of emotions (positive, neutral, negative) were described as outputs of classification. To efficiently map the subject's emotions with the extracted features from both ECG and PPG signals, a convolution neural network (CNN) was developed. With just two signals tracked, the output is equal or better than similar works.

[Zhang et al.] [69] proposed a revolutionary multimodal fusion system with regularization focused on a new kernel matrix perspective and a deep network architecture. In representation learning, they used the deep network architecture's superior efficiency to transform the native space of a predefined kernel into a task-specific function space. They adopted a shared presentation layer to learn the representation of fusion, which implies the tacit combination of many kernels. Simultaneously, in the loss function, a new

regularization term was added to model relationships between representations to enhance multimodal fusion efficiency.

[Nakisa et al.] [70] suggest a temporal fusion approach with deep learning model to capture the non-linear emotional association and enhance emotion classification efficiency. Using two separate fusion approaches: early fusion and late fusion, the proposed model's efficiency is evaluated. Specifically, after learning modality using a single deep network, they use a CNN (ConvNet) (LSTM) model to combine the EEG and BVP signals to learn and investigate the strongly coupled representation of feelings through modalities. The findings revealed that the temporal multimodal deep learning models effectively grouped human emotions into four quadrants of dimensional emotions based on early and late fusion approaches.

[Asghar et al.] [71] employed features from four previously trained DNN architectures to identify EEG data. Then they generated the data for network training by transforming the captured data into two-dimensional images. They suggested an innovative and powerful form of emotion detection, which is a high-quality feature collection. Using DFC-based functionality, they are shortening the training time for the network. The results suggested that the model has significantly improved the extracted features and feature classification efficiency.

[Nie et al.] [72] proposed a multi-layer LSTM architecture to retrieve the system emotion recognition's multimodal input function. There will be major changes by neural-network-based add-on ideas in utterance-level. They use LSTM to analyze video in order to extract the global features of speech. The experimental findings produced outstanding and important efficiency enhancements over the baseline.

Table II compares several works in terms of deep learning algorithms/ techniques, accuracy, and dataset used.

IV. Discussion

Several scientific experiments on emotional recognition have been carried out during the last decade using multimodal Mixing multiple modal signals such as facial and audio gestures, audio and written language, physiological signals and different variations in those modalities.

In the proposed multimodal emotion recognition and after reviewing previous studies, everyone used more than one modal to identify emotions using different methods and techniques combined with deep learning. Where deep learning also has multiple algorithms, methods and architectures in classification and extraction features. These deep learning algorithms directly affect achieving a higher rate of deeper understanding to improve precision, trust, and performance.

In this paper, after reviewing the latest research in multimodal emption recognition, we divided them into two groups. The first group combined signals from audio and text or image and text. The second group combined signals from facial and body physiological. Table I and Table II above show the specifics of the work that has been undertaken to identify the different feelings using multimodal so far.

TABLE II. Combining signals from physiological signals

Author	Neural network architecture and deep learning technique (algorithms)		Accuracy	Data set used
	classification	*method*		
[Zexu et al.] **[62]**	LSTM	MMAN ,Fusion method	73.98%	IEMOCAP
[Siriwardhana, et al]**[63]**	SSL modle	Speech-BERT, RoBERT Shallow fusion	—	IEMOCAP, CMU-MOSEI, CMU-MOSI),
[Priyasad, et al] **[64]**	DCCN with a SincNet layer, RNN	band-pass filters	80.51%	IEMOCAP
[Krishna et al] **[50]**	1D CNN	cross-modal attention	1.9% improvement	IEMOCAP
[Caihua] **[24]**	SVM	Fusion method	72.52%,	Berlin Emotional DB
[Lee et al.] **[65]**	CNN	Natural Language Processing (NLP)	—	Asian Character from the TV drama series
[Liu, et al] **[66]**	LSTM	Bert model, LSFM	5.77% improvement	777 songs(Music Mood Classification Data Sets)

The Table II literature includes seven researchers from the above-mentioned first party. Previously used, but with new data sets, techniques were paired with deep education. We find that some of these studies fairly break down their results, while others are less straightforward, making the comparison a difficult challenge. In [37] CNN based raw waveform was used following cross-modal focus between audio and text characteristics, also in [30], deep learning (CNN) methodology was developed to manipulate and to merge text and acoustic data to understand emotions, both of them achieved good results and improved performance with accuracy reaching 80.5% on the IEMOCAP data set. In [24], the author used SVM-based multimodal voice emotion recognition. Results show that the SVM algorithm has modified a lot to speech emotion recognition with the accuracy of 72.52 % using Berlin Emotional DB. In [50] and [56], the authors used LTSM for classification, and both of them record significant improvement in performance and accuracy reaches 73.98% and (5.77% improvement) respectively, using the different database. In [51], they combine speech and text.

In contrast, in [44] they combine image and text as signal to recognition. In the two studies, they benefited from the deep learning, as they noticed a performance improvement. However, the accuracy rate was not mentioned.

Table II, Which contains six researchers from the second group mentioned above. We can see that emotion elicitation techniques offer assistance in emotion classification. In [57], [61], the author both used LTSM and deep learning to classifications features but using different physiological signals. When three physiological signals(ECG, EDA,EMG) are considered, the classification accuracy was 95% using LTSM. However, when only two physiological signals (EEG,BVP) are considered, the classification accuracy was 71.61%. In [58],[62], they both used CNN, and two physiological signals are considered, but in [62], more than one classification algorithms are used, which had a great impact on the results with an accuracy rate of more than 97% on SEED database. From here, we noticed that with more than one algorithm, the results are better in some cases. In [60], to boost the efficiency of the multimodal fusion, a new algorithm and new architecture using the kernel matrix and a deep neural network are proposed., this method generates different accuracy on different posed databases, to approximately 63% when dealing with DEAP data set while with DECAF data set the accuracy is less about 57%.

V. Conclusion

In this paper, we analyzed and discussed the various multimodal identification of human emotions using deep learning. The findings indicate that emotion recognition can be done with greater precision and enhancement using a multimodal approach from biological signals to identify emotional states.

Emotion modulates nearly all human speech styles, such as facial expression, movements, stance, voice tone, sentence collection, breathing, skin temperature and clamminess, etc. Emotions will alter the message significantly: often, it is not what has been said that is most relevant, nor how it has been said. Faces tend to be the most obvious means of contact between feelings. However, as opposed to the voice and other modes of speech, they are often easily manipulated in reaction to various social circumstances.

However, emotional variations can be observed in physiological signals for a very short period of around 3-15seconds [73]. Therefore, it would provide better results to extract the details on the moment of emotional response. Throughout the process of the different physiological signals, this will entail a window dependent strategy.

Furthermore, it will improve a consumer emotion detection system with higher classification accuracy by using a comprehensive and novel extraction feature, feature collection and classification techniques.

We also note that researchers have learned and checked their pattern in many datasets to assess the proposed neural network structure. The recognition rates differ from one database to another in compliance with the same deep learning pattern.

References

[1] N. Perveen, D. Roy, and K. M. Chalavadi, "Facial Expression Recognition in Videos Using Dynamic Kernels," IEEE Transactions on Image Processing, vol. 29, pp. 8316-8325, 2020.

[2] S. Bateman and S. Ameen, "Comparison of algorithms for use in adaptive adjustment of digital data receivers," IEE Proceedings I (Communications, Speech and Vision), vol. 137, pp. 85-96, 1990.

[3] H. I. Dino and M. B. Abdulrazzaq, "Facial expression classification based on SVM, KNN and MLP classifiers," in 2019 International Conference on Advanced Science and Engineering (ICOASE), 2019, pp. 70-75.

[4] O. F. Mohammad, M. S. M. Rahim, S. R. M. Zeebaree, and F. Y. Ahmed, "A survey and analysis of the image encryption methods," International Journal of Applied Engineering Research, vol. 12, pp. 13265-13280, 2017.

[5] V. Shrivastava, V. Richhariya, and V. Richhariya, "Puzzling Out Emotions: A Deep-Learning Approach to Multimodal Sentiment Analysis," in 2018 International Conference on Advanced Computation and Telecommunication (ICACAT), 2018, pp. 1-6.

[6] D. A. Zebari, H. Haron, S. R. Zeebaree, and D. Q. Zeebaree, "Enhance the Mammogram Images for Both Segmentation and Feature Extraction Using Wavelet Transform," in 2019 International Conference on Advanced Science and Engineering (ICOASE), 2019, pp. 100-105.

[7] L. Chen, Y. Ouyang, Y. Zeng, and Y. Li, "Dynamic facial expression recognition model based on BiLSTM-Attention," in 2020 15th International Conference on Computer Science & Education (ICCSE), 2020, pp. 828-832.

[8] M. Wu, W. Su, L. Chen, W. Pedrycz, and K. Hirota, "Two-stage Fuzzy Fusion based-Convolution Neural Network for Dynamic Emotion Recognition," IEEE Transactions on Affective Computing, 2020.

[9] A. Clark, S. Abdullah, and S. Ameen, "A comparison of decision-feedback equalizers for a 9600 bit/s modem," Journal of the Institution of Electronic and Radio Engineers, vol. 58, pp. 74-83, 1988.

[10] S. Ammen, M. Alfarras, and W. Hadi, "OFDM System Performance Enhancement Using Discrete Wavelet Transform and DS-SS System Over Mobile Channel," ed: ACTA Press Advances in Computer and Engineering, 2010.

[11] J. Liang, S. Chen, and Q. Jin, "Semi-supervised Multimodal Emotion Recognition with Improved Wasserstein GANs," in 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2019, pp. 695-703.

[12] A. A. Yazdeen, S. R. Zeebaree, M. M. Sadeeq, S. F. Kak, O. M. Ahmed, and R. R. Zebari, "FPGA Implementations for Data Encryption and Decryption via Concurrent and Parallel Computation: A Review," Qubahan Academic Journal, vol. 1, pp. 8-16, 2021.

[13] A. A. Salih and M. B. Abdulrazaq, "Combining best features selection using three classifiers in intrusion detection system," in 2019 International Conference on Advanced Science and Engineering (ICOASE), 2019, pp. 94-99.

[14] H. Dino, M. B. Abdulrazzaq, S. R. Zeebaree, A. B. Sallow, R. R. Zebari, H. M. Shukur, et al., "Facial Expression Recognition based on Hybrid Feature Extraction Techniques with Different Classifiers," TEST Engineering & Management, vol. 83, pp. 22319-22329, 2020.

[15] S. S. R. Zeebaree, S. Ameen, and M. Sadeeq, "Social Media Networks Security Threats, Risks and Recommendation: A Case Study in the Kurdistan Region," International Journal of Innovation, Creativity and Change, vol. 13, pp. 349-365, 2020.

[16] S. Y. Ameen and S. W. Nourildean, "Coordinator and router investigation in IEEE802. 15.14 ZigBee wireless sensor network," in 2013 International Conference on Electrical Communication, Computer, Power, and Control Engineering (ICECCPCE), 2013, pp. 130-134.

[17] M. R. Al-Sultan, S. Y. Ameen, and W. M. Abduallah, "Real Time Implementation of Stegofirewall System," International Journal of Computing and Digital Systems, vol. 8, pp. 498-504, 2019.

[18] E. Chandra and J. Y.-j. Hsu, "Deep Learning for Multimodal Emotion Recognition-Attentive Residual Disconnected RNN," in 2019

International Conference on Technologies and Applications of Artificial Intelligence (TAAI), 2019, pp. 1-8.

[19] M. B. Abdulrazzaq and J. N. Saeed, "A comparison of three classification algorithms for handwritten digit recognition," in 2019 International Conference on Advanced Science and Engineering (ICOASE), 2019, pp. 58-63.

[20] J. Chen, Y. Lv, R. Xu, and C. Xu, "Automatic social signal analysis: Facial expression recognition using difference convolution neural network," Journal of Parallel and Distributed Computing, vol. 131, pp. 97-102, 2019.

[21] M. R. Mahmood, M. B. Abdulrazzaq, S. Zeebaree, A. K. Ibrahim, R. R. Zebari, and H. I. Dino, "Classification techniques' performance evaluation for facial expression recognition," Indonesian Journal of Electrical Engineering and Computer Science, vol. 21, pp. 176-1184, 2021.

[22] L. Hu, W. Li, J. Yang, G. Fortino, and M. Chen, "A Sustainable Multi-modal Multi-layer Emotion-aware Service at the Edge," IEEE Transactions on Sustainable Computing, 2019.

[23] E. Ghaleb, M. Popa, and S. Asteriadis, "Multimodal and Temporal Perception of Audio-visual Cues for Emotion Recognition," in 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), 2019, pp. 552-558.

[24] C. Caihua, "Research on Multi-modal Mandarin Speech Emotion Recognition Based on SVM," in 2019 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS), 2019, pp. 173-176.

[25] M. B. Abdulrazzaq and K. I. Khalaf, "Handwritten Numerals' Recognition in Kurdish Language Using Double Feature Selection," in 2019 2nd International Conference on Engineering Technology and its Applications (IICETA), 2019, pp. 167-172.

[26] K. B. Obaid, S. Zeebaree, and O. M. Ahmed, "Deep Learning Models Based on Image Classification: A Review," International Journal of Science and Business, vol. 4, pp. 75-81, 2020.

[27] T. D. Nguyen, "Multimodal emotion recognition using deep learning techniques," Queensland University of Technology, 2020.

[28] S. Y. Ameen, F. M. Almusailkh, and M. H. Al-Jammas, "FPGA Implementation of Neural Networks Based Symmetric Cryptosystem," in 6th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications May, 2011, pp. 12-15.

[29] K. Mohan, A. Seal, O. Krejcar, and A. Yazidi, "Facial Expression Recognition using Local Gravitational Force Descriptor based Deep Convolution Neural Networks," IEEE Transactions on Instrumentation and Measurement, 2020.

[30] C. Marechal, D. Mikolajewski, K. Tyburek, P. Prokopowicz, L. Bougueroua, C. Ancourt, et al., "Survey on AI-Based Multimodal Methods for Emotion Detection," ed, 2019.

[31] E. S. Hussein, U. Qidwai, and M. Al-Meer, "Emotional Stability Detection Using Convolutional Neural Networks," in 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 2020, pp. 136-140.

[32] H. I. Dino and M. B. Abdulrazzaq, "A Comparison of Four Classification Algorithms for Facial Expression Recognition," Polytechnic Journal, vol. 10, pp. 74-80, 2020.

[33] H. Miao, Y. Zhang, W. Li, H. Zhang, D. Wang, and S. Feng, "Chinese Multimodal Emotion Recognition in Deep and Traditional Machine Leaming Approaches," in 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia), 2018, pp. 1-6.

[34] D. Liu, X. Ouyang, S. Xu, P. Zhou, K. He, and S. Wen, "SAANet: Siamese action-units attention network for improving dynamic facial expression recognition," Neurocomputing, vol. 413, pp. 145-157, 2020.

[35] S. Zhang, X. Tao, Y. Chuang, and X. Zhao, "Learning deep multimodal affective features for spontaneous speech emotion recognition," Speech Communication, 2020.

[36] W. Mellouk and W. Handouzi, "Facial emotion recognition using deep learning: review and insights," Procedia Computer Science, vol. 175, pp. 689-694, 2020.

[37] P. Ekman and W. V. Friesen, Unmasking the face: A guide to recognizing emotions from facial clues: Ishk, 2003.

[38] L. A. Feldman, "Valence focus and arousal focus: Individual differences in the structure of affective experience," Journal of personality and social psychology, vol. 69, p. 153, 1995.

[39] S. R. Zeebaree, O. Ahmed, and K. Obid, "CSAERNet: An Efficient Deep Learning Architecture for Image Classification," in 2020 3rd International Conference on Engineering Technology and its Applications (IICETA), 2020, pp. 122-127.

[40] M. B. Abdulrazaq, M. R. Mahmood, S. R. Zeebaree, M. H. Abdulwahab, R. R. Zebari, and A. B. Sallow, "An Analytical Appraisal for Supervised Classifiers' Performance on Facial Expression Recognition Based on Relief-F Feature Selection," in Journal of Physics: Conference Series, 2021, p. 012055.

[41] S. Y. Ameen and M. R. Al-Badrany, "Optimal image steganography content destruction techniques," in International Conference on Systems, Control, Signal Processing and Informatics, 2013, pp. 453-457.

[42] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. W. Shalaby, "A 3D-convolutional neural network framework with ensemble learning techniques for multi-modal emotion recognition," Egyptian Informatics Journal, 2020.

[43] I. A. Khalifa, S. R. Zeebaree, M. Ataş, and F. M. Khalifa, "Image steganalysis in frequency domain using co-occurrence matrix and Bpnn," Science Journal of University of Zakho, vol. 7, pp. 27-32, 2019.

[44] A. Chen, H. Xing, and F. Wang, "A Facial Expression Recognition Method Using Deep Convolutional Neural Networks Based on Edge Computing," IEEE Access, vol. 8, pp. 49741-49751, 2020.

[45] S. Dou, Z. Feng, X. Yang, and J. Tian, "Real-time multimodal emotion recognition system based on elderly accompanying robot," in Journal of Physics: Conference Series, 2020, p. 012093.

[46] G. Wen, T. Chang, H. Li, and L. Jiang, "Dynamic Objectives Learning for Facial Expression Recognition," IEEE Transactions on Multimedia, 2020.

[47] I. Lasri, A. R. Solh, and M. El Belkacemi, "Facial Emotion Recognition of Students using Convolutional Neural Network," in 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS), 2019, pp. 1-6.

[48] S. Rajan, P. Chenniappan, S. Devaraj, and N. Madian, "Novel deep learning model for facial expression recognition based on maximum boosted CNN and LSTM," IET Image Processing, vol. 14, pp. 1373-1381, 2020.

[49] J. Saeed and S. Zeebaree, "Skin Lesion Classification Based on Deep Convolutional Neural Networks Architectures," Journal of Applied Science and Technology Trends, vol. 2, pp. 41-51, 2021.

[50] D. Krishna and A. Patil, "Multimodal Emotion Recognition using Cross-Modal Attention and 1D Convolutional Neural Networks," Proc. Interspeech 2020, pp. 4243-4247, 2020.

[51] J. Liscombe, J. Venditti, and J. Hirschberg, "Classifying subject ratings of emotional speech using acoustic features," in Eighth European Conference on Speech Communication and Technology, 2003.

[52] R. Ibrahim, S. Zeebaree, and K. Jacksi, "Survey on Semantic Similarity Based on Document Clustering," Adv. sci. technol. eng. syst. j, vol. 4, pp. 115-122, 2019.

[53] D. Bertero and P. Fung, "A first look into a convolutional neural network for speech emotion detection," in 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP), 2017, pp. 5115-5119.

[54] J. Lee and I. Tashev, "High-level feature representation using recurrent neural network for speech emotion recognition," in Sixteenth annual conference of the international speech communication association, 2015.

[55] D. A. Hasan, B. K. Hussan, S. R. Zeebaree, D. M. Ahmed, O. S. Kareem, and M. A. Sadeeq, "The Impact of Test Case Generation Methods on the Software Performance: A Review," International Journal of Science and Business, vol. 5, pp. 33-44, 2021.

[56] Y.-T. Lan, W. Liu, and B.-L. Lu, "Multimodal Emotion Recognition Using Deep Generalized Canonical Correlation Analysis with an Attention Mechanism," in 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1-6.

[57] P. Bhattacharya, R. K. Gupta, and Y. Yang, "The Contextual Dynamics of Multimodal Emotion Recognition in Videos," arXiv preprint arXiv:2004.13274, 2020.

[58] H. Zhang, "Expression-EEG Based Collaborative Multimodal Emotion Recognition Using Deep AutoEncoder," IEEE Access, vol. 8, pp. 164130-164143, 2020.

[59] F. Al-Naima, S. Y. Ameen, and A. F. Al-Saad, "Destroying steganography content in image files," in IEEE Proceedings of Fifth International Symposium on Communication Systems, Networks and Digital Signal Processing, University of Patras, Patras, Greece, 2006.

[60] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, "Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting," Energies, vol. 13, p. 391, 2020.

[61] M. Verma, S. K. Vipparthi, G. Singh, and S. Murala, "LEARNet: Dynamic imaging network for micro expression recognition," IEEE Transactions on Image Processing, vol. 29, pp. 1618-1627, 2019.

[62] Z. Pan, Z. Luo, J. Yang, and H. Li, "Multi-modal Attention for Speech Emotion Recognition," arXiv preprint arXiv:2009.04107, 2020.

[63] S. Siriwardhana, A. Reis, R. Weerasekera, and S. Nanayakkara, "Jointly Fine-Tuning" BERT-like" Self Supervised Models to Improve Multimodal Speech Emotion Recognition," arXiv preprint arXiv:2008.06682, 2020.

[64] D. Priyasad, T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Attention Driven Fusion for Multi-Modal Emotion Recognition," in ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 3227-3231.

[65] J.-H. Lee, H.-J. Kim, and Y.-G. Cheong, "A Multi-modal Approach for Emotion Recognition of TV Drama Characters Using Image and Text," in 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), 2020, pp. 420-424.

[66] G. Liu and Z. Tan, "Research on Multi-modal Music Emotion Classification Based on Audio and Lyirc," in 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2020, pp. 2331-2335.

[67] S. DHAOUADI and M. M. B. KHELIFA, "A multimodal Physiological-Based Stress Recognition: Deep Learning Models' Evaluation in Gamers' Monitoring Application," in 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 2020, pp. 1-6.

[68] C.-J. Yang, N. Fahier, W.-C. Li, and W.-C. Fang, "A Convolution Neural Network Based Emotion Recognition System using Multimodal Physiological Signals," in 2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan), 2020, pp. 1-2.

[69] X. Zhang, J. Liu, J. Shen, S. Li, K. Hou, B. Hu, et al., "Emotion Recognition From Multimodal Physiological Signals Using a Regularized Deep Fusion of Kernel Machine," IEEE Transactions on Cybernetics, 2020.

[70] B. Nakisa, M. N. Rastgoo, A. Rakotonirainy, F. Maire, and V. Chandran, "Automatic Emotion Recognition Using Temporal Multimodal Deep Learning," IEEE Access, 2020.

[71] M. A. Asghar, M. J. Khan, M. Rizwan, R. M. Mehmood, and S.-H. Kim, "An Innovative Multi-Model Neural Network Approach for Feature Selection in Emotion Recognition Using Deep Feature Clustering," Sensors, vol. 20, p. 3765, 2020.

[72] W. Nie, Y. Yan, D. Song, and K. Wang, "Multi-modal feature fusion based on multi-layers LSTM for video emotion recognition," Multimedia Tools and Applications, pp. 1-10, 2020.

[73] H. Gunes and M. Pantic, "Automatic, dimensional and continuous emotion recognition," International Journal of Synthetic Emotions (IJSE), vol. 1, pp. 68-99, 2010.