Journal of Theoretical and Applied Information Technology
15th October 2022. Vol.100. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EMOTION DETECTION USING CONTEXT BASED FEATURES USING DEEP LEARNING TECHNIQUE

SOHIT AGARWAL[1], DR. MUKESH KUMAR GUPTA[2]

[1]Research Scholar, Department Of Computer Engineering and Information Technology, Suresh Gyan Vihar University, Jaipur, Rajasthan, India,

[2]Professor, Department of Electrical Engineering, Suresh Gyan Vihar University, Jaipur, Rajasthan, India

E-mail: [1]sohit.agarwal@gmail.com

ABSTRACT

Detection of emotion from various aspect is the most important thing in social media environment. Numerous research has gone into giving robots the ability to recognize emotions. Most prior computer vision attempts focused on assessing facial expressions and, in certain circumstances, body position. This strategy works nicely in certain situations. Their performance is restricted in natural settings. Studies suggest that the scene setting, together with facial expression and body stance, helps us perceive people's emotions. However, owing to a paucity of data, the context processing for automated emotion identification has not been fully studied. In this paper we proposed a system for detection of emotion using real time visual and context-based features. The major objective of this research to identify the sentiment from large video based on context. To achieve this functionality, it needs to extract numerous frames from video and validate all frames using proposed algorithms. This research basically carried out the extraction of hybrid features from real time image as well as video dataset and build classifier to selective features. The hybrid deep learning classification algorithm has used for predict the correct sentiment with Hybrid Convolutional Neural Network (H-CNN) and Recurrent Neural Network (RNN). In extensive experimental analysis evaluation has done in terms of accuracy which obtains better than traditional machine learning and deep learning classification algorithms.

Keywords: *Context Aware Emotion Detection, Feature Extraction, Classification, Sentiment Classification, Machine Learning*

1. INTRODUCTION

The Social interactions are facilitated by emotional cues, universal signs that humans use to communicate throughout ordinary activities. To show others that they are happy, individuals utilise facial expressions such as a broad grin. Psychologists believe that people get emotional signals from their social partners and mix them with their own experiences to detect emotions and make appropriate actions. Furthermore, emotion detection, particularly facial expression identification, has long been vital in a human-computer interface (HCI). Several scientific investigations on the recognition of facial expressions have recently been done to create solutions based on new technologies in computer vision and pattern recognition. This study has several uses, including advertising, continuous monitoring, efficient surveillance cameras, and robotic interface development [1].

The complexity and dynamism of human emotional displays make emotion detection difficult. Emotions evolve throughout time, are multi-modal in structure, and differ in physiological and communication [2]. Also, using facial cues, important emotional indicators, might be difficult due to differences in head postures and lighting [3]. Noise and obstruction also impact body expressions and speech tone. In certain circumstances, emotions need context [4]. Face expression representation in video-based emotion identification commonly contains start, peak, and offset [5,6]. The onset and offset phases are generally shorter than the apex time. The lack of a defined time boundary between phases causes issues, as can spontaneous emotions.

However, issues with head posture, illumination, and the complexity of facial expression representation owing to unplanned expression make video-based emotion identification difficult in the field. Emotions are contextualised. For example, in a dark setting or when the subject of interest is small, we may distinguish emotions based on our previous experiences with relevant items such as objects, people, and portions of the scene. A multilayer emotional feature map is required to cope with ambiguous emotion temporal bounds. Traditional and deep learning approaches frequently concentrate on facial expressions, which fluctuate in response to stressful states, underlying intents, and social interactions. Ekman et al. recommend dividing emotions into six main categories: anger, disgust, fear, happiness, sorrow, and surprise [7].

2. PROBLE STATEMENT

The various emotion detection technique has developer by previous researchers such as face emotion [1] [2], text emotion [32], action emotion [15], body movement [15] etc. These systems basically carried out the internal features for detection of sentiment or emotions. But sometimes external features should be impact on actual sentiment class, and no any existing approaches has evaluated such external features. In this research we proposed a system for detection of emotion from internal as well as extract features such as image context features for detection of sentiment.

2.1 Contribution of Research

This research basically focusses on identification of sentiment using context-based features, which consider object features and outsider features. The generation of caption for input image and detection the sentiment score using hybrid deep learning algorithm. During the execution we extracts numerous features from input image such as color features, shape features, texture features, histogram, object detection, autoencoder etc. For getting better accuracy.

3. REVIEW OF LITERATURE

Asghar et al. [8] developed layered support vector machines to categories emotions across cultures. The SVM results were associated with the SVM costume probability distributions. A naive Bayes predictor decides if an emotion exists. Combining face pictures from JAFFE, TFEID, KDEF, CK+, and Radboud creates the multi-culture dataset. Hasani et al. [9] suggested a 3D CNN model for FER in videos. This innovative network system combines three-dimensional ResNet with temporal information from several video snippets. In addition, the network uses facial landmark points as input, emphasising facial components over locations that may only play a minor part in forming facial emotions. The suggested method was evaluated on four publicly accessible facial expression datasets.

Fan et al. [10] built a framework that combines discriminative features from convolutional neural networks with handcrafted data from form and appearance to improve the FER system's resilience and accuracy. This technique outperforms past models on FER-2013, CK+, FERG, and JAFFE datasets [11]. The proposed system encodes shape, appearance, and dynamic data. Face texture information is also obtained to enhance texture extraction discrimination. The system outperforms current FER approaches on the CK+ dataset. A low-resolution video or depth sensor data may be used to analyse gait, according to [12]. There are two widely acknowledged emotion representation models in research. The Pleasure Dominance Arousal model portrays emotions as a continuous spectrum, whereas the Distinct Categories model treats emotions as discrete categories. This article uses the Distinct Categories approach with four emotion classes: angry, joyful, sad, and neutral. In 2018, Ahmed et al. [13] proposed a technique to extract 17 Laban Movement Analysis (LMA) characteristics from input gait sequences. A Genetic Algorithm was used to choose a subset of characteristics for four classifiers. The classifiers' outputs were merged using the score and rank-level fusion to boost accuracy.

Ahmed et al. [14] refined this approach. The authors found six key GER characteristic groupings. The feature set was then refined using ANOVA and Multivariate ANOVA before being re-select using a binary GA. Lastly, the final set of characteristics was used to categories emotions using the score and rank-level fusion. Research employing traditional machine learning algorithms for feature extraction and classification was robust and contributed to the area by defining several characteristics that are favorable for GER. However, handmade features had a restricted range of domain-specific characteristics. Recent deep learning techniques also use data-driven algorithms that focus on prominent characteristics and work with significantly bigger training datasets.

. Consequently, [15] misclassify more gaits as cheerful, explaining the low class and mean average accuracy scores. The proposed LSTM and MLP architecture use robust sequential neural networks to extract temporal and spatial features, tapered neural networks to consolidate those features, handcrafted features to reduce sensitivity to class imbalances in the dataset, and a light architecture to infer emotions. Using a multi-objective weighted voting ensemble classifier, Onan et al. SVM and NB are the basic learners in their proposed system, and their performance in terms of sensitivity and security determines the weighted adjustment [16]. This system outperforms standard ensemble learning models in applications including sentiment analysis, software defect prediction, spam filtering, credit risk modelling, and semantic mapping. The job of detecting software defects on a dataset of laptops achieves the best accuracy of 98.86 percent. The suggested model outperformed all other individual emotions, with class AP scores exceeding 0.9 for furious, happy, and sad classes. Notably, implementing batch normalization increased the suggested architecture's neutral class emotion recognition from 0.46 to 0.65. With just 0.18 past results published in HAPAM [17], this emotional class has proved to be the hardest to distinguish for all models. The suggested architecture's average accuracy of 0.65 for the neutral class is three times better than the best previous method's 0.18. The previous best technique [17] is heavily influenced by the dataset's uneven class distribution, with the furious class outperforming the happy, sad, and neutral classes. Because it is solely data-driven, it is more sensitive to class differences in the data. The proposed bi-modular neural network reduces this behavior by using robust handmade features and batch normalization for network regularization. Graph NNs and handcrafted and deep feature combinations favor the happy class. Randhavane et al. [18] created synthetic gait sequences using a DNN encoder-decoder. The network's encoder was then used to extract deep characteristics paired with posture- and movement-based gait features. Finally, a random forest classifier identified four distinct moods.

The research by Osman et al. [19] addressed the data-sparsity problem of recommender systems by integrating a sentiment-based analysis. Their work was applied to the Internet Movie Dataset (IMDb) and Movie Lens datasets, but improvements in sentiment analysis have been made since the paper was published. In particular, when only sparse rating data are available, sentiment analysis can play a key role in improving the quality of recommendations. This is because recommendation algorithms mostly rely on users' ratings to select the items to recommend. Such ratings are usually insufficient and very limited. On the other hand, sentiment-based ratings of items that can be derived from reviews or opinions given through online news services, blogs, social media, or even the recommender systems themselves are seen as being capable of providing better recommendations to users.

In addition, some recommendation approaches leverage item metadata to deal with problems mainly associated with collaborative filtering methods [20]. Among such data, social tags have become an important input to recommender systems for streaming platforms. Many efforts have been addressed to unify tagging information to reveal behavior and extract the latent semantic relations among items [21]. This system proposed a method for automatic generation of social tags for music recommendation. The purpose is to avoid the cold-start problem common in such systems, when a user or an item is newly added to the system and as a result has few ratings. Instead of relying on ratings in a music recommendation method, social tags may be used to improve music recommender systems by calculating the similarity between music pieces by combining both tag and rating, in the same way that other item attributes, such as movie genres or music audio features, are used to classify items or establish item similarity [22].

They suggested a voting based ensemble classification algorithms has used sentiment analysis in [23]. The researchers studied three types of tweets (positive, negative, and neutral). The "twitter-airline-sentiment" dataset was used to test ML classifiers. While the [24] describes twitter sentiment analysis using supervised classification approach. Their research looked at the impact of TF, TF-IDF, and word2vec on classification accuracy. LSTM, a deep learning model, performed less accurately than ML models. The voting classifier's accuracy is 78.9% using TF and TF-IDF feature extraction. Umer et al. [25] used a CNN and LSTM ensemble to analyze tweet sentiment. Because ML classifiers struggle with large datasets, they recommended a Deep Learning-based ensemble system. They tested their method on three datasets. They use word2vec and TF-IDF to extract features. The CNN-LSTM outperformed other

Journal of Theoretical and Applied Information Technology
15th October 2022. Vol.100. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

classifiers. They also compared the suggested model's performance to other deep learning models, validating the proposed technique.

Stjanovski et al. [26] tested deep CNN sentiment analysis using Twitter data. We employed the dropout layer, softmax layer, two fully connected layers, and numerous changing windowed filters to train the proposed CNN on top pre-trained word embeddings from big text corpora. On Twitter corpora, the pre-trained word vectors perform quite well for unsupervised learning. They utilized the 2015 Twitter dataset and got an F1 Score of 64.85%. Jianqiang et al. [27] proposed a deep learning-based method to categories tweets. The system's global vector (Glove) depth CNN (DCNN). The authors created sentiment features by concatenating pre-trained N-gram and word embedding features. They also employed a recurrent structure to collect contextual data and CNN to represent text. On the STSGd dataset, their suggested approach obtained 85.97 percent accuracy. Santos et al. [28] suggested using a deep convolutional neural network to classify sentiment in short texts. They employed two datasets and got 86.4% accuracy on the ST corpus. Ishaq et al. [29] recommended a deep neural network-based model for hotel visitor reviews. A multi-class classification with 3 or 10 classes was examined by the authors. LSTM achieves 97 percent accuracy in binary class categorization.

The structure of the data affects deep learning sentiment categorization. For the creation of a maximal sentiment categorization system, three CNN-based and five RNN-based deep neural networks were examined. The research indicated that bigger training data sets improved model accuracy. They also looked at the data input structure, and found that a word-level structure learned hidden patterns more correctly than a character-level structure [30]. Research proposes a hybrid sentiment classification algorithm that combines word embedding with deep learning [31]. The authors used FastText embedding with character embedding to create a suggested combination of CNN and BiLSTM that scored 82.14 percent accuracy. Another research used CNN-LSTM deep learning to analyse Twitter sentiment [32]. Their strategy used unlabelled data to pre-train word embeddings with remote supervision and fine-tuning. Their suggested technique is based on the number of CNN and LSTM network ensembles utilised to classify tweets. The suggested method was evaluated using the SemEval-2017 twitter dataset. They got 74.8% accuracy using 10 CNN and 10 LSTM networks.

Most current approaches aim to get the images emotionally local region and handle it separately, without addressing if the emotional small communities are interrelated. This is not favourable to sophisticated scenario analysis. Contextual information is commonly employed in scene interpretation and feature extraction tasks. According to APCNet [33], number of co, adaptable, and global communications guidance context aspects are important in interpreting complicated scenarios. Other scales lose information while capturing things of a single scale. Thus, most visual activities need multi-scale. The adaptive property allows a pixel in the context feature to not only correlate with local pixels but also infer relationships with other areas in the global context, capturing long-distance dependencies. Thus, adaptable qualities may help us find relationships between areas. We first use multi-scale adjustable context features in sentiment classification analysis applications, inspired by APCNet. Dual capabilities could incorporate Features and simple items may be described by low-level characteristics, whereas sophisticated products can be described by high-level features.

4. PROPOSED SYSTEM DESIGN

The below Figure 1 describes a hybrid feature selection for detection of emotion using deep learning classification. The CNN has used for classification of sentiment and hybrid feature extracted, such as luminance, chrominance, histograms-based features, binary features, sobel features, autoencoder, alfa features, beta features, gamma features, epsilon, etc., in the convolutional layer and feed to the pooling layer. In dense layer, classification has done as an image sentiment.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

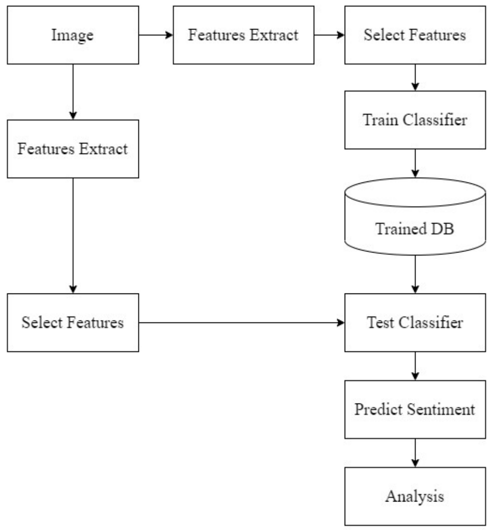

Figure 1 : Proposed System Architecture For Emotion Detection

Pre-processing: In the image there are several kinds of noises. To remove noise, the eminence of the images is expanded by pre-processing methods. This technique's main goal is to remove the constant noise ratio, adjust the visual aspect of the input image, remove noise and unwanted parts in the context, smooth the internal components, and keep the edge.

Feature Extraction: The process of extracting higher-level knowledge about an object, such as structure, texture, colour, and comparison, is known as feature extraction. The visual processing time schedules and the machine learning system both use texture analysis. Finding and analysing textures can help with diagnosis at various stages of tumour detection. It's being used to enhance a diagnostic method's precision by choosing statistical features such as mean, contrast, energy, entropy, standard deviation, skewness, etc. The image segmentation is a method to distribute an image object into minor parts to examine and distinguish the significant evidence of a image. It creates several no of pixels in the image and given label to share particular feature information.

Convolutional layer: A CNN's base layer is a convolutional layer. The convolution process operates on a tiny local region of the information using a convolutional kernel of a specific dimension. A convolutional kernel is a weighted sum that can be learned. The convolutional layer's information is fed via an objective function, and then binaries feature map is produced. The feature map may be used as the input for the convolution layers that follows. As a result, after stacking multiple convolutional layers layer by layer, more complex characteristics may be retrieved. Furthermore, the cells in each feature map contribute the strength of a convolutional kernel in an activation function, ensuring that the dimensionality in the structure does not substantially grow even as the quantity of convolutional layers expands, lowering the model's memory footprint. As a result, this model may aid in the formation of a more complex network structure.

1. Pooling layer: After such a convolution operation, a combining layer is usually used. The greatest, intermediate, and randomised pooling layers are examples of general pooling layers. The highest and averaged pooling algorithms identify the greatest and averaged values of neighbouring neurons, correspondingly, whereas the randomised pooling algorithm chooses values from neurons based on likelihood. Other types of convolution layer, such as overlaying pooling and geometric pyramid pooling, are frequently superior to the conventional pooling layers. Max pooling, independent of which kind is employed, seeks to collect features but is unconcerned with their exact positions, ensuring that the connectivity can learn important features even if the input layer shifts a little amount. Furthermore, a wavelet transform does not change the amount of feature maps in the preceding layer, but it lowers their spatial complexity and retains the critical data in the feature vector, decreasing supervised learning computation even more. The feature has been extracted such as binary features, Sobel features, autoencoder features, histogram features and some GLCM base features.

Classification algorithms: This method has no of image processing techniques for image segmentation. Supervised is the simplest way to classify data. It is very useful for large image but has a poor contrast [16]. The classification method first trained data using supervised learning model with labelled dataset and validate the test dataset accordingly [18]. During the module it extracts various features of training data and generates feature vectors of those selected features. Once training has done similar feature extraction have been applied on testing dataset and classify test data accordingly).

5. RESULTS AND DISCUSSION

Execution using Intel i7 CPU 2.7 GHz with 16 GB RAM has used for experimental analysis. 5G network has been tested using the RESENT

Journal of Theoretical and Applied Information Technology
15th October 2022. Vol.100. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(32,50,101,152) version. The several deep learning frameworks have been examined on both datasets, and the accuracy of system is shown in Table 1.

Table 1: System Performance With Various Deep Models On EMOTIC Dataset

Data samples	PNN	RCNN	FCNN	H-CNN
1000	92.60%	96.80%	97.40%	96.10%
2000	93.85%	96.70%	95.60%	97.30%
5000	92.10%	95.60%	94.80%	96.25%
10000	91.30%	96.50%	96.20%	96.40%

The above Table 1 describes an emotion classification for Emotic images dataset for all deep models using TensorFlow for different data samples.

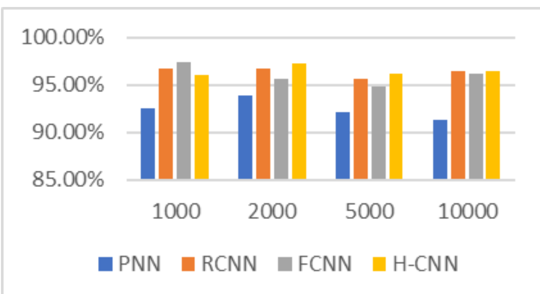

Figure 2 : Performance Evaluation Of Emotion Classification Using Various Deep Learning Techniques

The above Figure 2 demonstrates emotion classification techniques using various deep learning classification algorithms. The REGNET-100 deep learning

Findings

- The generation of captions for input image plays major role for identification sentiment that provides significant accuracy with any deep learning classifiers.
- This research produces high er accuracy over the conventional deep learning classifiers such as PNN, CNN, FCNN and RCNN etc.
- The proposed H-CNN produces 2.5% higher emotion detection accuracy for image and video dataset than conventional deep learning classifiers.
- The RESNET frameworks has used for evaluation of proposed research but RESNET-101 provides higher accuracy than others.
- This research effective when it deals image dataset with generation of caption while little slower for video objects.

Limitation of Research

- Object detection from input image and video when image having high crowd density, that also impact on detection accuracy.
- Sometime noise data also impact on accuracy, the conventional filtration techniques are not able to eliminate various custom noise from image due to this issue it generates high error rate or overfitting problem.
- Time computation is another limitation of this research when it deals with large videos, because it needs to generate various frames and then identify the class sentiment.
- High performance computation required for training the heterogeneous or large datasets

6 CONCLUSIONS

This research has proposed an image context-based sentiment lexicon with multiscale perception and several emotional levels. To learn the sentiment correlation degree of distinct areas at different scales in the picture, a novelty-level context H-CNN is presented. In the picture sentiment analysis challenge, the adaptive context architecture is originally introduced, which helps the model grasp difficult circumstances. The suggested H-CNN component's multi-scale characteristic is independently of model structure and may mix multiple scales according to various data sources to mine contextual information from images. The model analyses two level feature fusion techniques to better employ multiple sense of emotional representation to comprehend semantic items. The extraction of internal and external features from input images and video gives better detection accuracy than conventional clasisifcation and detection techniques. Even not a system which focused on extraction of context-based features, and identify the class sentiment according to both features. As a conclusion this research fulfil the outcome to detection the emotion using context features on hetrogenious

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

dataset including synthetic as well as real time streaming dataset. The future work for this system to implement a ensemble deep learning classification with heterogenous dataset with collaboration various features.

REFERENCES

[1] Corneanu, C.A.; Simon, M.O.; Cohn, J.F.; Guerrero, S.E. Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-Related Applications. IEEE Trans. Pattern Anal. Mach. Intell. 2016, 38, 1548–1568.

[2] Bänziger, T.; Grandjean, D.; Scherer, K.R. Emotion recognition from expressions in face, voice, and body: The Multimodal Emotion Recognition Test (MERT). Emotion 2009, 9, 691–704.

[3] Martinez, B.; Valstar, M.F. Advances, Challenges, and Opportunities in Automatic Facial Expression Recognition. In Advances in Face Detection and Facial Image Analysis; Springer International Publishing: Cham, Switzerland, 2016; pp. 63–100.

[4] Wieser, M.J.; Brosch, T. Faces in Context: A Review and Systematization of Contextual Influences on Affective Face Processing. Front. Psychol. 2012, 3.

[5] Koelstra, S.; Pantic, M.; Patras, I. A Dynamic Texture-Based Approach to Recognition of Facial Actions and Their Temporal Models. IEEE Trans. Pattern Anal. Mach. Intell. 2010, 32, 1940–1954. [PubMed]

[6] Bernin, A.; Müller, L.; Ghose, S.; Grecos, C.; Wang, Q.; Jettke, R.; von Luck, K.; Vogt, F. Automatic Classification and Shift Detection of Facial Expressions in Event-Aware Smart Environments. In Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference, Corfu, Greece, 26–29 June 2018; pp. 194–201.

[7] Ekman, P.; Friesen, W.V. Facial Action Coding System: A Technique for the Measurement of Facial Movement; Consulting Psychologists Press: Palo Alto, CA, USA, 1978.

[8]. Asghar, K.; Sadddique, M.; ul Haq, I.; Nawaz-ul Ghani, M.A.; Ali, G. Stacked Support Vector Machine Ensembles for Cross-Culture Emotions Classification. IJCSNS 2019, 19, 23.

[9] Hasani, B.; Mahoor, M.H. Facial expression recognition using enhanced deep 3D convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 30–40.

[10] Fan, X.; Tjahjadi, T. Fusing dynamic deep learned features and handcrafted features for facial expression recognition. J. Vis. Commun. Image Represent. 2019, 65, 102659.

[11] Minaee, S.; Minaei, M.; Abdolrashidi, A. Deep-emotion: Facial expression recognition using attentional convolutional network. Sensors 2021, 21, 3046. [PubMed]

[12] Xu, S.; Fang, J.; Hu, X.; Ngai, E.; Guo, Y.; Leung, V.; Cheng, J.; Hu, B. Emotion Recognition From Gait Analyses: Current Research and Future Directions. arXiv 2020, arXiv:2003.11461.

[13] Ahmed, F.; Sieu, B.; Gavrilova, M.L. Score and rank-level fusion for emotion recognition using genetic algorithm. In Proceedings of the IEEE 17th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), Berkeley, CA, USA, 16–18 July 2018; pp. 46–53.

[14] Ahmed, F.; Bari, A.H.; Gavrilova, M.L. Emotion recognition from body movement. IEEE Access 2019, 8, 11761–11781.

[15] Yan, S.; Xiong, Y.; Lin, D. Spatial temporal graph convolutional networks for skeleton-based action recognition. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, LO, USA, 2–7 February 2018; pp. 1–9.

[16] Bhattacharya, U.; Mittal, T.; Chandra, R.; Randhavane, T.; Bera, A.; Manocha, D. Step: Spatial temporal graph convolutional networks for emotion perception from gaits. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 1342–1350.

[17] Bhattacharya, U.; Roncal, C.; Mittal, T.; Chandra, R.; Kapsaskis, K.; Gray, K.; Bera, A.; Manocha, D. Take an emotion walk: Perceiving emotions from gaits using hierarchical attention pooling and affective mapping. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 145–163.

[18] Randhavane, T.; Bhattacharya, U.; Kapsaskis, K.; Gray, K.; Bera, A.; Manocha, D. Identifying emotions from walking using

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

affective and deep features. arXiv 2019, arXiv:1906.11884

[19] Osman, N.A.; Noah, S.A.M. Sentiment-based model for recommender systems. In Proceedings of the 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP), Kota Kinabalu, Malaysia, 26–28 March 2018; pp. 1–6.

[20] Oramas, S.; Nieto, O.; Sordo, M.; Serra, X. A deep multimodal approach for cold-start music recommendation. In Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems, Como, Italy, 27 August 2017; pp. 32–37.

[21] Advani, G.; Soni, N. A Novel Way for Personalized Music Recommendation Using Social Media Tags. IJSRD-Int. J. Sci. Res. Dev. 2015, 2, 404–410.

[22] Eck, D.; Lamere, P.; Bertin-Mahieux, T.; Green, S. Automatic generation of social tags for music recommendation. Adv. Neural Inf. Process. Syst. 2007, 20, 385–392.

[23] Onan, A.; Koruko ˘glu, S.; Bulut, H. A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. Expert Syst. Appl. 2016, 62, 1–16.

[24] Rustam, F.; Ashraf, I.; Mehmood, A.; Ullah, S.; Choi, G.S. Tweets classification on the base of sentiments for US airline companies.Entropy 2019, 21, 1078.

[25] Umer, M.; Ashraf, I.; Mehmood, A.; Kumari, S.; Ullah, S.; Sang Choi, G. Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model. Comput. Intell. 2021, 37, 409–434.

[26] Stojanovski, D.; Strezoski, G.; Madjarov, G.; Dimitrovski, I. Twitter sentiment analysis using deep convolutional neural network. In Proceedings of the International Conference on Hybrid Artificial Intelligence Systems, Bilbao, Spain, 22–24 June 2015; pp. 726–737.

[27] Zhao, J.; Gui, X.; Zhang, X. Deep convolution neural networks for twitter sentiment analysis. IEEE Access 2018, 6, 23253–23260.

[28] Dos Santos, C.; Gatti, M. Deep convolutional neural networks for sentiment analysis of short texts. In Proceedings of the COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, Dublin, Ireland, 23–29 October 2014;pp. 69–78.

[29] Ishaq, A.; Umer, M.; Mushtaq, M.F.; Medaglia, C.; Siddiqui, H.U.R.; Mehmood, A.; Choi, G.S. Extensive hotel reviews classification using long short term memory. J. Ambient. Intell. Humaniz. Comput. 2020, 1–11.

[30] Seo, S.; Kim, C.; Kim, H.; Mo, K.; Kang, P. Comparative study of deep learning-based sentiment classification. IEEE Access 2020, 8, 6861–6875.

[31] Salur, M.U.; Aydin, I. A novel hybrid deep learning model for sentiment classification. IEEE Access 2020, 8, 58080–58093.

[32] Cliche, M. BB_twtr at SemEval-2017 task 4: Twitter sentiment analysis with CNNs and LSTMs. arXiv 2017, arXiv:1704.06125. 3

[33] He, J.; Deng, Z.; Zhou, L.; Wang, Y.; Qiao, Y. Adaptive pyramid context network for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019