iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** I **Month of publication:** January 2025

DOI: https://doi.org/10.22214/ijraset.2025.65640

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue I Jan 2025- Available at www.ijraset.com

Context-Based Emotion Recognition using the EMOTIC Dataset

Mansi Borawake[1], Dr. Madhuri Bhalekar[2]
School of Computer Engineering and Technology, MIT World Peace University, Pune, India

Abstract: ***Emotion detection is an essential aspect of human-computer interaction, empowering systems to interpret and react more accurately to human emotions. Traditional emotion recognition methods often rely on facial expressions or speech cues alone, which may not capture the full complexity of emotional states. However, context - such as the surrounding environment, objects, and interactions - Serves a key function in accurately identifying emotions. This paper introduces an innovative method for emotion recognition that incorporates contextual information by utilizing the EMOTIC dataset, which contains images annotated with both apparent emotions and their corresponding contextual features. By leveraging deep learning models that combine conventional facial emotion detection with contextual cues, we significantly improve the precision and dependability of emotion recognition systems. The proposed model outperforms baseline methods, demonstrating the critical importance of context in enhancing emotion recognition performance.***

Keywords: galvanic skin response (GSR), photo plethysmography (PPG), and electro encephalography, Image Processing and Deep Learning.

I. INTRODUCTION

motion recognition enhances user experience by enabling systems to respond appropriately to users' emotional states, making interactions more intuitive and personalized. Examine the constraints of conventional emotion recognition techniques that depend only on facial expressions or vocalizations. In recent years, computerized emotion detection via artificial intelligence and computer vision has become indispensable. Advancements in deep neural networks enable this technology to include environmental, social, and cultural elements, in addition to face emotions. Our objective is to develop more empathic systems for many applications, ranging from healthcare to analyzing emotional exchanges on social media. We used genuine photos from many databases, including EMOTIC, to train our models for the development of this technology. We created two advanced algorithms using deep learning methodologies. We examine context and body language to improve our comprehension of human emotions in photographs by optimizing our models' hyperparameters. To ascertain emotions in context, we combine the 26 distinct emotional categories with the three continuous emotional dimensions. We finalize the suggested pipeline by integrating our models.

II. LITERATURE REVIEW

According to [1] The framework you're discussing enhances music recommendation systems by utilizing data from wearable devices and sensors to assess users' moods. It gathers Bodily signals, such as galvanic skin response (GSR), photoplethysmography (PPG), and electroencephalography (EEG), along with facial expression data from a camera. This blend allows the recommendation engine to gain a deeper insight into the user's emotional state as it occurs. For instance, if the system detects increased stress levels through GSR and EEG readings, it can suggest calming music to help alleviate that stress. By incorporating this diverse array of data, the system can provide more tailored and precise music suggestions, which enhance user experience and engagement.

The article [2] discusses a commercially available EEG Bluetooth headset that monitors changes in brain waves, including alpha and beta waves. This device allows for safe data transmission to mobile devices via Bluetooth. The EEG signals gathered can offer valuable insights into various cognitive disorders and conditions, serving as a useful reference for medical professionals in treatment planning. Additionally, the categorization and classification of depression levels derived from EEG signals can form a foundation for assessing music therapy. Consequently, the article proposes implementing a music recommendation system that generates a list of relaxing music options tailored to users based on their specific symptoms and emotional states. The article[3] describes the creation of an attention device that uses electroencephalographic (EEG) technology to monitor and assess an individual's attention levels based on their music preferences. Initially, the recorded brainwave data is filtered to remove any inaccuracies before being processed by a Support Vector Machine (SVM) classifier. This classifier differentiates between two categories of brainwave data: attention and non-attention. The result is a hybrid music recommendation model that leverages this attention data alongside the user's musical preferences, allowing for a more personalized music selection experience tailored to the user's engagement levels.

©IJRASET: All Rights are Reserved

The[4] study investigates how music tracks in English and Urdu affect human stress levels by analyzing brain signals. Twenty-seven participants, comprising 14 men and 13 women aged 20 to 35, all fluent in Urdu, volunteered for the research. Their electroencephalograph (EEG) signals were recorded using a four-channel MUSE headset while they listened to various music tracks. To assess their stress levels, participants completed a state and trait anxiety questionnaire, which required them to subjectively rate their stress. The effect of musical style on stress varies based on individual taste, emotional reactions, and context, yet each genre provides unique calming qualities that can aid in managing stress.

The [5] The study categorizes six fundamental emotions—anger, disgust, fear, joy, sadness, and surprise—into two classes: "emotion" and "no emotion," while analyzing different machine learning and deep learning approaches for this classification.This categorization is based on physiological data and subjective assessments of valence, arousal, and dominance sourced from the DEAP dataset (Dataset for Emotion Recognition utilizing EEG, physiological measurements, and video signals).The research explores the effectiveness of these techniques both with and without feature selection. The results demonstrate high classification accuracies for each emotion, with ninety-eight point zero two percent for anger, one hundred percent for joy, ninety-six percent for surprise, ninety-five percent for contempt, ninety point seven five percent for fear, and ninety point zero eight percent for sadness.

The[6] study utilizes ScientoPy to conduct a scientometric analysis of significant publications on recommendation approaches found in scientific databases, including Elsevier's Scopus and Clarivate Web of Science. Over the past two decades, there has been substantial research focused on emotion-based tourism recommendation systems. The review emphasizes the importance of collecting, processing, and extracting features from data gathered through sensors and wearable devices to identify emotions effectively. The research suggests several key topic areas, including recommendation systems, emotion detection, wearable technology, and machine learning, as critical components in advancing the field.

The [7] study explores the role of music as a therapeutic element in digital treatment programs aimed at enhancing mental health and well-being. It highlights that music can evoke emotional responses in listeners, leading to measurable changes in brain activity, which can be captured using electroencephalography (EEG). A scoping review was performed to identify recent research exploring the impact of music on brain activity and emotional states within these digital therapy frameworks. Out of 585 publications that met the study's criteria, six relevant papers were selected for further analysis. The [8] study introduces a noninvasive brain-computer interface (BCI) that utilizes EEG technology to control a robotic arm, enabling users to perform tasks involving reaching and gripping multiple targets while maneuvering around obstacles using hybrid control strategies. Findings from a group of seven participants showed that training in motor imagery (MI) can change brain rhythms, and six of these participants were able to successfully carry out tasks performed online with a hybrid-control robotic arm system. This system showcases strong performance by merging MI-based EEG with computer vision, gaze detection, and partially autonomous guidance. Such integration significantly enhances task precision and lessens the cognitive strain associated with prolonged mental activity. The [9] study outlines a process for classifying EEG data into normal and seizure categories using a Long Short-Term Memory (LSTM) network. The preprocessing phase involves three key steps: normalizing the EEG data, applying appropriate filters to extract relevant sections of the data, and managing the dataset effectively. Once the data is preprocessed, it is utilized to train the LSTM network. Following the training process, the Softmax function is applied to classify the input data, differentiating between normal activity and seizure events. This structured approach enhances the accuracy of seizure detection in EEG data. The [10] study discusses a method for collecting data from users' social media texts through the Internet of Things (IoT) for the purpose of emotion detection. Once the emotions are identified, two music suggestion approaches are proposed. The first approach is an expert-based method, where specialists curate Music suggestions tailored to the identified emotions. The second approach is a feature-based method that operates without expert intervention, using the rhythm and articulation of the music to make Suggestions tailored to the user's emotional state. Additionally, a feedback mechanism is implemented in the music suggestion process, allowing the algorithm to refine its recommendations based on user responses. This dual approach enhances the relevance and personalization of music suggestions.

A. Research Gaps

1) The CNN-GAN and RNN-GAN methods for detecting sentiment have a low level of accuracy.

2) The ensemble framework is unable to recognize all of the components that make up an object in an image, which has an influence on training. Furthermore, the results show that the module can only recognize large objects

3) Ensemble is unable to identify new characteristics as compared to RNN-GAN and CNN- GAN; it can only collect unique characteristics and detect them as part of the ensemble.

4) There are certain colored objects that cannot be recognized by either the algorithms or the ensemble model in the image.

B. Existing Datasets

Multiple sources of data can generate sentiment analysis datasets. The conventional method for assembling a human-labeled dataset involves soliciting input from a substantial number of participants. In the field of sentiment classification, we can gather extensive opinion data by extracting information from popular social media platforms like Snapchat, Flickr, Facebook, and Instagram, and from websites that focus on business and customer feedback like Amazon, Snap deal, eBay, etc. Moreover, individuals in contemporary culture often articulate their thoughts and share their feelings via social media platforms on the internet.

III. CONTRIBUTIONS

The suggested study used a deep learning methodology to categorize picture sentiment. This study demonstrates a variety of strategies for feature extraction and selection from visual objects and develops the training knowledge appropriately. We have employed various feature selection techniques for extraction parameters like as a shape, texture, alpha, density, and color characteristics. Researchers often utilize text metadata to evaluate the emotions associated with each image. Normalizing the dataset is expected to significantly improve classification accuracy.

The preceding literature analysis demonstrates that several systems use deep convolutional neural networks, using diverse boosting techniques to enhance both temporal and spatial complexity.

1) Researchers have employed region-based convolutional neural networks (R-CNN) and some other neural network techniques, such as CNNs, to evaluate the sentiment of various image datasets.
2) We used the ImageNet library to extract features and construct the training model accordingly. Occasionally, exceeding the precision of CNN and comparable models may prove difficult.
3) The most often used datasets include the Flickr dataset, the Twitter testing dataset, and the ImageNet dataset.
4) The combination of a rapid recurrent neural network and a convolutional neural network may provide optimal accuracy with minimal temporal complexity, achieving the highest precision.

IV. RESEARCH CONTRIBUTION

1) We have outlined recent progress in image sentiment analysis, utilizing both traditional machine learning and deep learning techniques. This information will be greatly beneficial for other
2) A recent survey has guided our gap analysis, highlighting areas that need further exploration. These gaps will provide important direction for upcoming research initiatives.

V. PROPOSED METHODOLOGY

In proposed research work, evaluate the entire system using supervised learning algorithms and first collect data from the EMOTIC dataset. In order to extract various features from inputs, the CNN algorithm sets and generates the trained model appropriately. In the testing system, classify each input EMOTIC dataset with the appropriate labels and show the system's efficiency. In feature extraction, the objective is to lower the feature count in a dataset by deriving new features from the current ones. On the other side, feature selection focuses on minimizing the number of input variables when developing a predictive model. Classification involves categorizing a given set of data into distinct classes and can be applied to both structured and unstructured data.

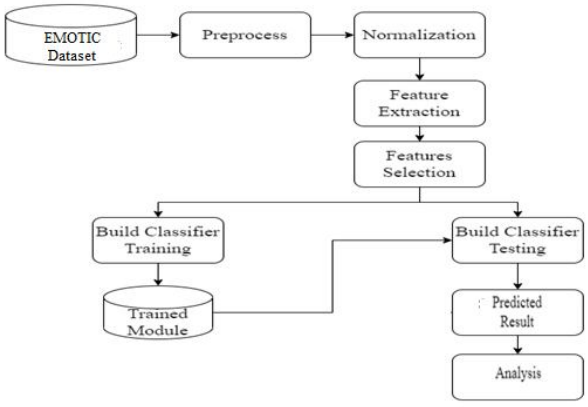


Figure 1: System architecture

A. *Module 1: Dataset Description*

Provide a detailed description of the EMOTIC dataset, including the types of images, emotion categories, and contextual labels. Discuss any data preprocessing steps taken to prepare the dataset for model training.

B. *Module 2: Feature Extraction*

The features extracted from images can be categorized into visual and contextual elements. Visual features include facial expressions, which are key indicators of emotions reflected through changes in facial muscles, eye movements, and mouth shapes, as well as body language, which encompasses posture, gestures, and movements that convey emotions or intentions. On the other hand, contextual features consist of the background, providing insight into the setting or environment surrounding the subject, and objects, which are specific items or elements within the image that may influence or relate to the emotions expressed. Various feature extraction techniques are employed to capture these intricate details. Local Binary Patterns (LBP) effectively analyze local pixel patterns for texture information, while the Histogram of Oriented Gradients (HOG) focuses on gradient orientations to detect shapes and patterns. Additionally, deep features obtained from Pre-trained CNN models like as VGG-Face and ResNet are utilized to extract rich representations of both facial expressions and contextual elements, resulting in a comprehensive understanding of the emotions conveyed in images.

C. *Module 3: Training and Validation*

The process of training a deep learning model for emotion recognition begins by organizing the data, which includes splitting the dataset into training, validation, and test sets to achieve reliable generalization. At this stage, data augmentation techniques are employed to diversify the training dataset by generating altered versions of the original images, and normalization is applied to ensure uniformity across the input data. These steps are vital for improving the model's accuracy and effectiveness in identifying emotions.

During training, images are fed into the model, which extracts features through a forward pass. A loss function assesses the gap between predicted and true labels and through a backward pass, the model adjusts its weights using techniques like Adam or SGD. To enhance performance, methods such as learning rate scheduling, early stopping, and dropout are employed. For validation, the model’s predictions are assessed using metrics like accuracy (percentage of correct predictions), precision (percentage of true positive predictions), recall (percentage of actual positives detected), and F1-score (a combined metric of precision and recall), providing a comprehensive evaluation of its effectiveness.

D. *Module 4: Classification*

Describe the deep learning model used for emotion recognition, such as convolutional neural networks (CNNs) or transformers. Explain how the model integrates both facial and contextual features to make emotion predictions. We approach emotion recognition as a problem of multi-label classification. Based on both facial expressions and contextual cues, the EMOTIC dataset classifies each image into one or more of several predefined emotion categories. The EMOTIC dataset includes a range of emotion categories, such as happiness, sadness, anger, fear, and surprise, among others. Each image may exhibit multiple emotions simultaneously, reflecting the complexity of human emotional states. In addition to facial expressions, the classification model considers contextual information present in the image, such as the surrounding environment, objects, and other people, which can significantly influence the perceived emotion. The introduced model leverages deep learning techniques, like convolutional neural networks (CNNs) and transformer-based models, to extract features from both the subject's facial expressions and the surrounding context. Once these features are obtained, they are employed to classify the image into relevant emotion categories. This method enables a holistic comprehension of the emotional elements portrayed in the images, successfully capturing visual signals along with contextual information.

E. *Module: Evaluation Metrics*

We evaluate the classification model's performance across various metrics, such as accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC). We pay particular attention to the model's capacity to address the multilabel aspect of the problem, ensuring that it can effectively classify multiple emotions simultaneously. Additionally, we evaluate the influence of contextual factors on classification accuracy, as these elements can significantly impact the model's overall performance in emotion recognition tasks.

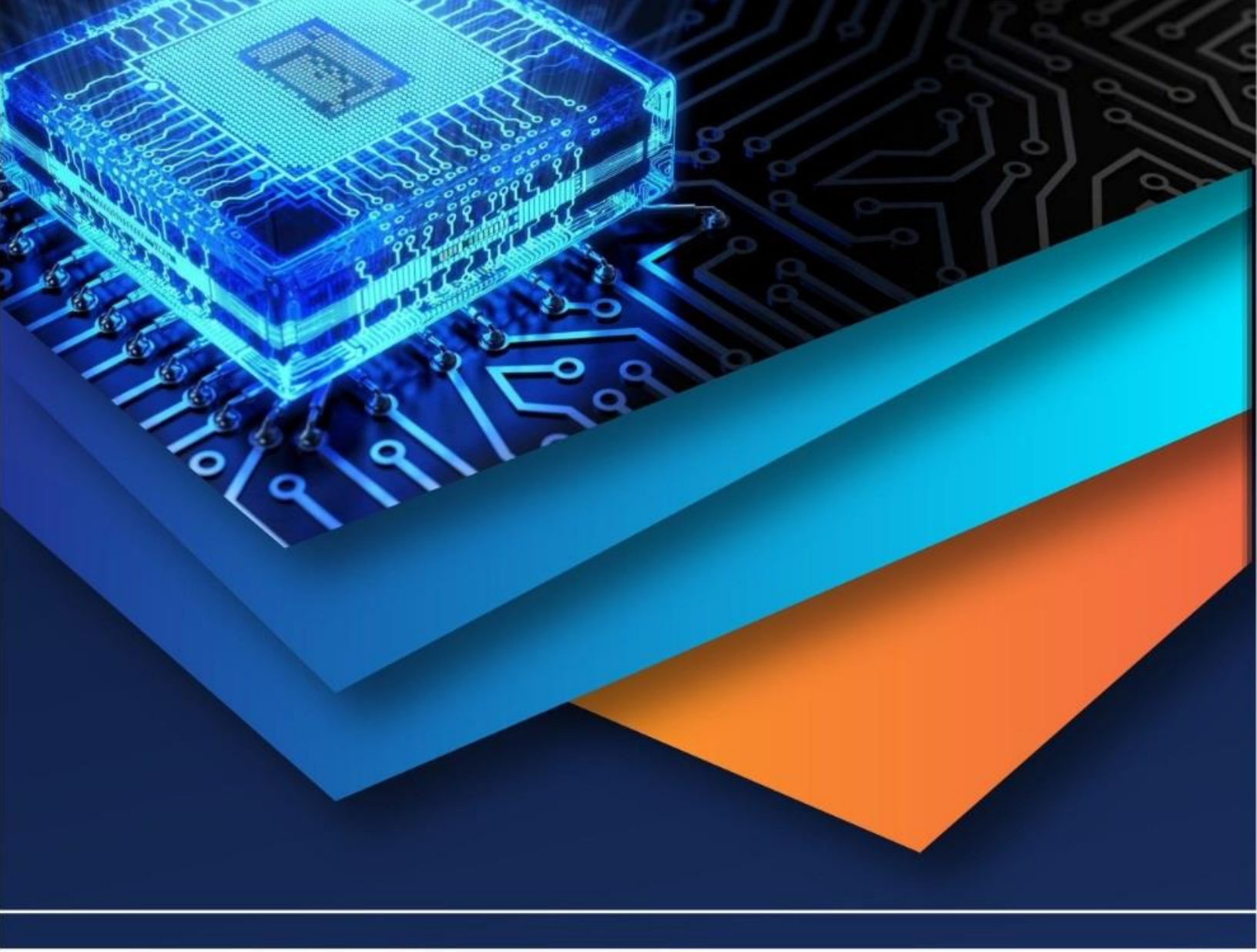
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

VI. CONCLUSION

Use the proposed approach to generate substantial emotion for a diverse range of photos. The system may extract hybrid characteristics, including picture and body information, to aid in sentiment creation. The system is capable of functioning with both picture and text datasets. Detecting emotions via contextual analysis of photos presents considerable hurdles in learning methodologies. Multiple attributes, such as color, form, texture, and other encoder characteristics, might affect the creation of visual captions during the classification process. Numerous studies indicate that situational context is a significant influence in the proper interpretation of others' emotions. The absence of data has impeded a thorough study on context processing for automated emotion recognition. One proposed solution is an emotion-detection system that utilizes real-time visual and contextual data. At its core, this study extracts hybrid features from real-time image datasets and develops a classifier that selects characteristics based on these attributes. The hybrid deep learning classifier, which employs a Deep Convolutional Neural Network (DCNN), effectively predicts the emotion associated with a given text using a Recurrent Neural Network (RNN). This holistic approach enhances the system's emotion analysis capabilities by integrating insights derived from visual inputs along with textual information.

REFERENCES

[1] Bhagya, C., &Shyna, A.: An Overview of Deep Learning Based Object Detection Techniques. In 1st International Conference on Innovations in Information and Communication Technology (ICIICT),pp. 1-6. IEEE,2021.
[2] Ragusa, E., Cambria, E., Zunino, R., &Gastaldo, P.: A Survey on Deep Learning in Image Polarity Detection: Balancing Generalization Performances and Computational Costs. Electronics, p.783.,2021.
[3] Islam, J., & Zhang, Y.: Visual sentiment analysis for social images using transfer learning approach. In International Conferences on Big Data and Cloud Computing , Social Computing and Networking, Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom)pp.124130.IEEE.,2019.
[4] Yuhai, Y., Hongfei, L., Meng, J., and Zhao, Z.: Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks., Algorithms 9, pp. 2, 2016.
[5] You, Q., Luo, J., Jin, H. and Yang, J.: Robust image sentiment analysis using progressively trained and domain transferred deep networks., In Twenty-ninth AAAI conference on artificial intelligence, 2015.
[6] Alharbi, A.S.M.; de Doncker, E. Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information. Cogn. Syst. Res. 2019, 54, 50–61.
[7] Kraus, M.; Feuerriegel, S. Sentiment analysis based on rhetorical structure theory: Learning deep neural networks from discourse trees. Expert Syst. Appl. 2019, 118, 65–79.
[8] Do, H.H.; Prasad, P.; Maag, A.; Alsadoon, A.J. Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review. Expert Syst. Appl. 2019, 118, 272–299.
[9] Abid, F.; Alam, M.; Yasir, M.; Li, C.J. Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter. Future Gener. Comput. Syst. 2019, 95, 292–308.
[10] Yang, C.; Zhang, H.; Jiang, B.; Li, K.J. Aspect-based sentiment analysis with alternating contention networks. Inf. Process. Manag. 2019, 56, 463–478.
[11] Wu, C.; Wu, F.; Wu, S.; Yuan, Z.; Liu, J.; Huang, Y. Semi-supervised dimensional sentiment analysis with variation auto encoder. Knowl. Based Syst. 2019, 165, 30–39.
[12] Costa, Willams, et al. "A survey on datasets for emotion recognition from vision: Limitations and in-the-wild applicability." Applied Sciences 13.9 (2023): 5697.
[13] Han, Xiao, Fuyang Chen, and Junrong Ban. "Music emotion recognition based on a neural network with an Inception-GRU residual structure." Electronics 12.4 (2023): 978.
[14] Huang, Yibo, et al. "Emotion recognition based on body and context fusion in the wild." Proceedings of the IEEE/CVF international conference on computer vision. 2021.
[15] Agarwal, S. O. H. I. T., and DR Mukesh Kumar Gupta. "Emotion detection using context based features using deep learning technique." J. Theor. Appl. Inf. Technol. 100.19 (2022): 1-12.
[16] Onim, Md Saif Hassan, Himanshu Thapliyal, and Elizabeth K. Rhodus. "Utilizing Machine Learning for Context-Aware Digital Biomarker of Stress in Older Adults." Information 15.5 (2024): 274.
[17] Zhou, Siwei, et al. "Emotion recognition from large-scale video clips with cross-attention and hybrid feature weighting neural networks." International Journal of Environmental Research and Public Health 20.2 (2023): 1400.
[18] Alsabhan, Waleed. "Human–computer interaction with a real-time speech emotion recognition with ensembling techniques 1D convolution neural network and attention." Sensors 23.3 (2023): 1386.
[19] Chinnalagu, Anandan, and Ashok Kumar Durairaj. "Context-based sentiment analysis on customer reviews using machine learning linear models." PeerJ Computer Science 7 (2021): e813.
[20] Abdullah, Sharmeen M. Saleem Abdullah, et al. "Multimodal emotion recognition using deep learning." Journal of Applied Science and Technology Trends 2.01 (2021): 73-79.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)