Grenze International Journal of Engineering and Technology, June Issue

Enhancing Movie Genre Classification through Emotional Intensity Detection: An Improvised Machine Learning Approach

Mohit Mehta[1], Somnath Sinha[2], Neha Singhal[3] and Sudhakar T[4]
[1-4]Christ (Deemed to be University), Bangalore, India
Email: mohit.mehta@mca.christuniversity.in, {somnath.sinha, neha.singhal, sudhakar.t}@christuniversity.in

***Abstract*— Movie Genre Classification through Emotion Intensity is a computer vision technique used to identify facial emotion through a sequential neural network model and to get the genre of the movie with it. This paper delves into latest advancements in Emotion Detection, particularly emphasizing neural network models and leveraging face image analysis algorithms for emotion recognition.**

***Index Terms*— Face perception, Facial emotion recognition, Software evaluation, Image processing, Mental health (MH), AI/ML.**

I. INTRODUCTION

In today's era of digital entertainment, understanding the emotional undercurrents in movies is pivotal for creating engaging and resonant storytelling experiences. Movie Genre Classification (MGC) typically relies on textual and visual features, often overlooking the crucial aspect of emotional intensity. This research dives into the importance and advantages of integrating Emotional Intensity Detection into MGC, unraveling a novel approach to enrich genre classification with emotional nuances.

Sentiment Analysis is a cornerstone for this research, acting as the lens through which emotional cues are extracted from movie scripts, dialogues, and audiovisual elements. Our objective is to measure the emotional intensity in movie material by utilizing sentiment analysis tools, which will provide a more comprehensive knowledge of viewer experiences. By using sentiment analysis, genre categorization becomes more detailed and captures the emotional aspects that greatly influence audience preferences.

Movie Genre Classification through Emotional Intensity Detection addresses a critical gap in existing methodologies. Conventional genre classification frequently ignores the complex and dynamic nature of audience participation. This study recognizes the changing tastes of viewers who look for emotionally charged content in addition to particular genres. By solving this issue, we want to open the door for more precise and customized movie suggestions that take into account the varied emotional tastes of different audiences.

The methodology hinges on advanced machine learning techniques integrating sentiment analysis and deep learning models. Convolutional Neural Networks (CNNs) are employed to extract intricate emotional features from textual and audiovisual elements. These features are then seamlessly integrated
into the genre classification framework. Dropout layers are strategically utilized to mitigate overfitting, ensuring the model's robustness.

Grenze ID: 01.GIJET.10.2.178
© Grenze Scientific Society, 2024

This research contributes to the fields of both machine learning and film studies, offering a paradigm shift in the way we approach Movie Genre Classification. Combining sentiment analysis and deep learning improves genre classification accuracy and creates new opportunities for a deeper comprehension of the emotional narrative environment. The latter parts of this paper explore the nuances of the suggested methodology, including experimental findings and analysis that highlight the value of our strategy in improving movie genre classification via emotional intensity detection.

II. LITERATURE REVIEW

In paper [1], they comprehensively analyze the latest bias in facial emotion recognition (FER) using facial image analysis with neural networks. The authors conducted a systematic review to gather and analyze relevant literature in this field. Their methodology involved conducting a thorough search of databases and selecting papers based on pre-defined inclusion criteria. The results of their review revealed that deep neural net-works, particularly CNN-based architectures, have shown remarkable performance in FER tasks, outperforming traditional machine learning methods. The authors high-lighted the significance of diverse and representative datasets for training and evaluating FER systems. However, the paper also acknowledged some limitations. One limitation is the potential bias in the selected literature due to publication bias or language restrictions. Additionally, the review focused on image-based FER and did not extensively cover multimodal approaches that incorporate other modalities like audio or text.

In paper [2], they introduce RCL-Net, a novel approach for recognizing facial ex-pressions in real-world settings. The method addresses the challenges posed by occlusion, illumination variations, and pose changes, often resulting in low recognition rates and accuracy. RCL Net combines a residual network and a hybrid attention mechanism to emphasize local detail feature information, enabling the model to capture significant facial expression characteristics. The proposed method was evaluated using popular datasets, including FER2013, FERPLUS, CK+, and RAF-DB. The experimental results demonstrate the superior capability and robustness of RCL-Net compared to recent methods, both in laboratory controlled and field environments. Overall, the paper con-tributes valuable insights into facial expression recognition in challenging real-world scenarios.

A key component of human and computer interaction is facial expression identification, which has attracted a lot of interest lately. Paper [3] presents a comprehensive investigation of the applications of deep convolutional neural networks (DCNNs) for this task. The study's methodology involves training a DCNN model using the Amsterdam Dynamic Facial Expression Set-Bath Intensity Variations (ADFES-BIV) dataset, merging intensity levels, and extracting multiple frames per emotion. Evaluation is per-formed using the Warsaw Set of Emotional Facial Expression Pictures (WSEFEP) dataset. Preprocessing steps include frame extraction, grayscale conversion, histogram equalization, face detection, and region of interest determination. The outcomes show how successful the suggested strategy is, surpassing chance levels for all. However, limitations such as the small training dataset and the need for validation on larger and more diverse datasets should be addressed. Overall, the paper contributes to the field of facial emotion recognition, but further research is needed to enhance the robustness and generalizability of the approach.

The Paper [4] addresses the challenge of dynamic facial expression recognition (DFER) in real world. The authors propose a novel approach that includes a global convolution-attention block to rescale the channels of feature maps and an intensity-aware loss to help the network distinguish low-intensity expression samples. The GCA block enhances critical channels in low-intensity sequences while suppressing less useful channels. The paper highlights the importance of considering expression intensity in the discrete label DFER task and provides insights into improving recognition accuracy in real-world scenarios. However, it does not address the challenge of annotating intensity-related priors in low-cost datasets, which could limit its applicability to datasets without continuous intensity labels.

The paper [5] investigated sex differences in facial emotion recognition's accuracy and response latencies using short video stimuli. The study included 111 participants, both adolescents and adults, who viewed 1-second videos depicting ten different facial emotions at three intensity levels. The findings demonstrated that, compared to men, women were more accurate and quicker to recognize facial expressions. This durability of the female advantage was observed because it did hold true for various emotion categories and expression intensity levels. However, when identifying neutral faces, there was no discernible difference between males and females. In examining sex differences in facial emotion identification, the report emphasized the significance of employing dynamic stimuli and considering expression intensity fluctuations. The study offers insights into the variables impacting face emotion perception and advances our knowledge of how sex differences appear in everyday social interactions.

The paper [6] explores the performance of three different Facial Expression Recognition in classifying standardized and non-standardized emotional facial expressions. Based on standardized facial expressions, the results show that all three systems can effectively recognize fundamental emotions, with FaceReader having the highest accuracy. However, all algorithms perform much worse when evaluated on non-standardized facial expressions taken from emotionally charged movie sequences, misclassifying certain emotions as neutral. The study highlights the potential of automated FER as an alternative to human emotion recognition but also points out the need for further empirical evaluation to improve accuracy, particularly for specific facial expressions.

In the paper [7], the author provides a comprehensive literature review on facial emotion recognition (FER) using machine learning (ML) as well as deep learning (DL) approaches. The author collects relevant literature published during the current decade and discusses various FER datasets and their benchmark results for evaluation. The paper highlights the challenges in FER, such as variations in gender, age, and image quality. The methodology involves a detailed review and comparison of ML and DL techniques. The results demonstrate that DL-based approaches achieve higher accuracy but require more training time and processing capabilities. The work serves as a useful manual for both novice and seasoned researchers, and it ends by highlighting the need for more study in the FER field.

The paper [8] presents a comprehensive review of the perception of facial expressions of emotion. The authors discuss two models in cognitive science and neuroscience. The continuous model represents face expressions as feature vectors in a multidimensional space, allowing for different intensities and modes of expression. On the other hand, the categorical model consists of classifiers tuned to emotion categories. The authors propose a revised model that combines these approaches by using distinct continuous spaces to recognize compound emotion categories. They emphasize the im-portance of precise detection of facial landmarks rather than recognition for the classification of facial expressions. The paper provides an overview of the supporting model, discusses the role of configural and shape features, and highlights the need for advancements in face detection algorithms. The proposed model has implications for machine learning, computer vision, and studies on human perception, social interactions, and disorders. There is also discussion of the shortcomings of the current models and the difficulties facing the industry. All things considered, this work provides an extensive analysis of the prior research, introduces a fresh methodology, and makes recommendations for future approaches.

In paper [9], Authors propose a new facial emotion recognition model using a convolutional neural network (CNN) called ConvNet. The classification model is developed by fusing data extracted from facial expression images using Local Binary Pattern (LBP) along with region-based Oriented FAST and CNN. The authors trained the model using the FER2013 database and tested it on the JAFFE and CK+ datasets. The authors provide the research community with publicly accessible materials. However, the paper acknowledges limitations such as the need for a large amount of data for successful deep learning models and challenges with long training durations and less recognition rates in difficult environments. Overall, the study demonstrates the effectiveness of ConvNet in facial emotion recognition and highlights the significance of data diversity in improving model performance.

The paper [10] highlights the importance of facial emotion recognition (FER) systems in counseling and similar tasks. The study aims to build an FER system to support practitioners in understanding their patients' emotions. They discuss the potential of FER systems in recognizing both macro and micro-expressions and highlight the ad-vantages of camera-based emotion recognition. The paper presents the methodology used in the study, where an existing algorithm with high accuracy and speed is chosen and coded as a web application. The study findings indicate that the mood matrix has significant use for guidance counselors, and the students were comfortable during testing. However, the paper acknowledges limitations such as the lack of publicly available source code for some FER models and the need to consider patient comfort levels in future research.

The paper [11] presents a method for estimating expression intensity using deep CNNs. The results show the effectiveness of the deep CNN method compared to handcrafted facial features. Still, challenges remain regarding estimation accuracy for facial expressions with small data and the issue of overfitting due to data shortages. They employ the intraclass correlation coefficient (ICC) to measure the variability of collected labels. The deep CNN model used for estimation is based on the VGG-Face architecture. The researchers extend the range of targeted facial expressions and estimate intensity through regression. The limitations of the study include the scarcity of avail-able data and potential variability in the evaluation of expression intensity labels by different raters.

The paper [12] presents a ways for collecting as well as modeling multidimensional expression annotation’s from human annotators. The authors aim to capture the ambiguity and intensity of facial expressions, as the perception of expressions can vary across observers. The main contributions of this paper is an efficient method for collecting reproducible annotations, a benchmark dataset for facial expression prediction algorithms. The paper highlights the multidimensional nature of facial expressions and emphasizes the importance of considering ambiguity in expression annotations. How-ever, it does not directly contribute to algorithm development.

In paper [13] author provides a literature review on recent advancements in automatic emotion recognition using deep learning. The authors focus on studies utilizing facial images exclusively as a channel of information in human communication. The review highlights the contributions, architectures, and databases used in these studies, as well as the progress achieved. Conventional approaches for FER, such as geometric and appearance-based features, are discussed alongside their limitations. The paper emphasizes the success of deep learning techniques in improving FER performance. It also mentions the availability of FER databases for training deep neural networks. However, the paper does not cover how could be the reliance on facial images exclusively, which may not capture the full range of multimodal cues present in human communication. Another limitation could be the generalizability of the proposed architectures and methods across diverse populations, as facial expressions and emotional cues can vary across cultures and individuals.
The paper [14] investigates the recognition of facial emotion along with its role in aging. The study conducted an extensive cross-sectional analysis involving more than four hundred participants across a wide age range (5-96 years). The results showed that dynamic expressions were more accurately recognized across the lifespan than static expressions. The advantage of dynamic stimuli was particularly prominent in the elderly population, where performance for static images significantly decreased. The findings highlighted the use of dynamic stimuli in assessing facial expression recognition in older adults and caution against relying solely on static images.
The paper [15] shows a comparative study on the recognition of facial emotion in-tensities using machine learning algorithms. The authors address the need for encoding the intensity of observed facial emotions and propose a methodology that involves feature extraction, followed by classification. The study utilizes various databases and achieves accurate emotion recognition and intensity estimation. The results shows the potential application of this approach in real-time behavioral facial emotion recognition. The paper provides an extensive literature review, discusses the challenges and motivations, and outlines the experimental techniques, including face detection and emotion recognition algorithms.

III. METHODOLOGY/ IMPLEMENTATION

In this study, we are trying to predict the genre and type of the movie using Emotion Detection. Firstly, convolutional neural networks (CNNs) was used for emotion recognition, utilizing the FER-2013 dataset, featuring grayscale images with dimensions of 48x48 pixels. The Convolutional Neural Network (CNN) architecture was meticulously crafted, integrating multiple convolutional layers with varying filter sizes to effectively capture hierarchical features in the input images. Max-pooling layers were strategically inserted to downsample spatial dimensions, while spatial dropout layers were employed to prevent overfitting during training [Figure 1].

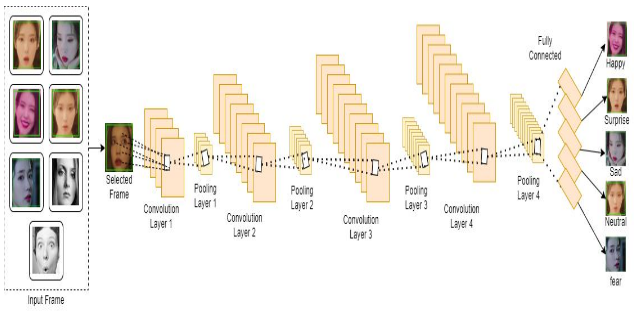


Figure 1. Training process of CNN model for face emotion detection

To enhance the model's generalization capabilities, data augmentation was introduced through the Keras Image Data Generator. This augmentation included rotational transformations (up to 20 degrees), horizontal and vertical shifts (up to 20% of the im-age size), and horizontal flips. These augmentations created a diverse training set, enabling the model to learn robust and invariant features.
The training process spanned 125 epochs, and a batch size of 255 was utilized for efficiency. A crucial aspect of our training strategy was the implementation of a learning rate scheduler. This scheduler dynamically adjusted the learning rate throughout training, following a specified decay schedule. This adaptive learning rate facilitated effective convergence, ensuring the model learned representations efficiently throughout the training epochs.
The accuracy [Figure 2] we achieved was 71.64948453608247 after 125 epochs since one epoch took roughly 240 seconds on a system with an i7-9750 CPU and GTX 1650 GPU. The training process took around 8 hours and 30 minutes to train the model.

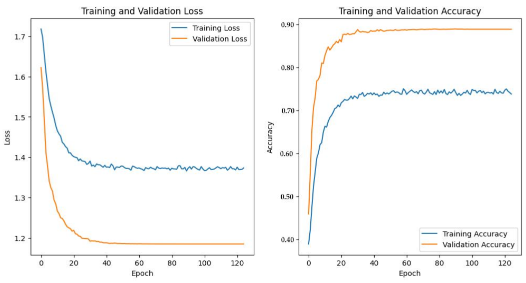


Figure 2. Model Accuracy and Loss Score we achieved during Training the model with above mentioned algorithm.

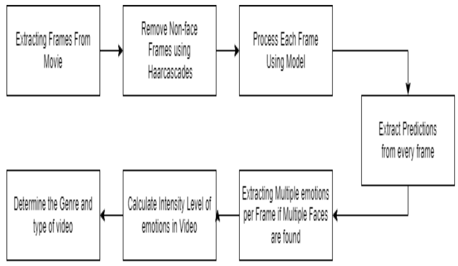


Figure 3. Work flow to determine the emotion intensity in a video

After the model [Figure 3] is trained, we Extract frames from a movie. As any standard video has 24 frames per second, we take every 12th frame as that is a frame at every 0.5 second. Then, using the Haarcascades library of OpenCV, we remove the frames where no faces are found and keep only those frames where any faces are detected to save unnecessary time that frames with no faces might use. Then, we process each frame through the model to detect the emotions of the faces (it can be one face or multiple faces) found in it. Then, we clear out multiple emotions in every frame and keep only distinct ones in every frame. As per that, we can calculate the intensity of any emotion at any point in time based on how many seconds that emotion is being held at any point in time. Then, using the intensity level we got from above for the emotion being held throughout the video, firstly, we classify the video type based on the emotion with the most intensities as it can help us read people's emotions like they have happy, sad, angry, scared, surprised or disgusted face majorly and based on that we can classify the genre of the movie like happy emotion depicts comedy or romance genre, sad emotion depicts drama, fear emotion depicts horror or thriller genre, surprise emotion depicts mystery or suspense genre, angry emotion depicts action or crime genre, and lastly disgust, fear, and surprise emotion depicts it is a psychological genre, and by this we can predict the genre and type of the movie. As per the emotion and their combination, we have classified various genres like such, and we can further extend it by classifying combinations of emotions per genre along with emotion being held in the maximum number of frames and the intensity they have to get better accuracy.

IV. Results

The analysis of emotion distribution across the video revealed insightful patterns in the expression of affective states. Figure 4 illustrates the frequency of each emotion category observed throughout the video frames. The multicolored bar graph provides a clear visual representation, with each color corresponding to a distinct emotion category. Notably, it can be observed that certain emotions dominate the overall emotional landscape, while others exhibit varying degrees of occurrence. The y-axis denotes the count of frames associated with each emotion, providing a quantitative measure of the prevalence of specific affective states. This analysis lays the foundation for a nuanced understanding of the emotional dynamics present within the video.

Examining consecutive intensity levels provides a nuanced perspective on the temporal dynamics of emotions within the video. Figure 5 illustrates the distribution of consecutive intensity levels for specific emotions, including 'happy,' 'angry,' 'disgust,' 'fear,' 'sad,' and 'surprise.'

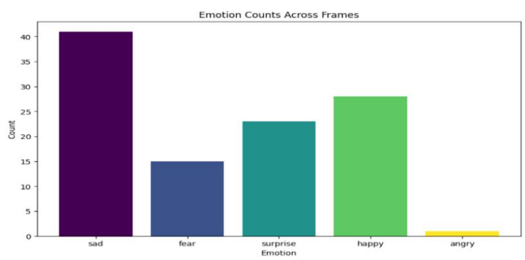
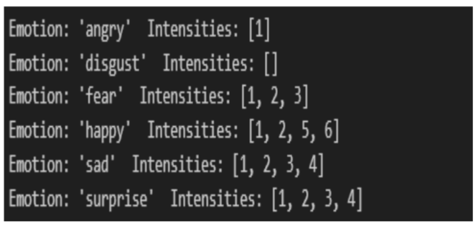


Figure 4. The emotion level observed throughout the sample video



```
Emotion: 'angry'  Intensities: [1]
Emotion: 'disgust'  Intensities: []
Emotion: 'fear'  Intensities: [1, 2, 3]
Emotion: 'happy'  Intensities: [1, 2, 5, 6]
Emotion: 'sad'  Intensities: [1, 2, 3, 4]
Emotion: 'surprise'  Intensities: [1, 2, 3, 4]
```

Figure 5. Intensity levels of the Video.

The graph in the Figure 6 illustrates unique emotion, and the corresponding counts of consecutive frames that exhibit similar intensity levels. Notably, the x-axis labels de-note the emotion category followed by the intensity level (e.g., 'happy - L1' signifies the count of consecutive frames with low intensity for happiness). The numeric value assigned after L shows for how many seconds that emotion is being held without any emotion change in the frame. The y-axis quantifies the number of occurrences for each distinct intensity level. This analysis unveils varying patterns in the duration and intensity of emotional expressions, shedding light on the nuanced temporal characteristics of affective states within the video content.

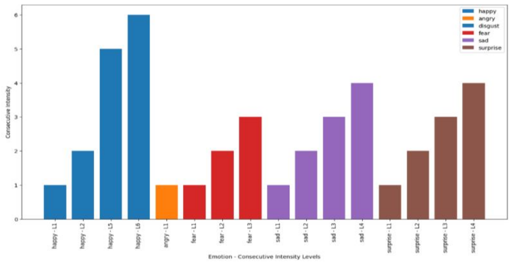
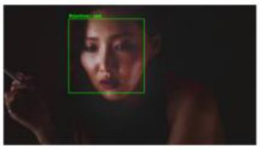
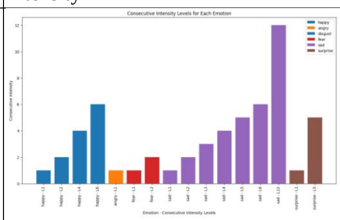
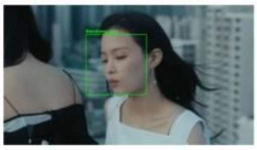
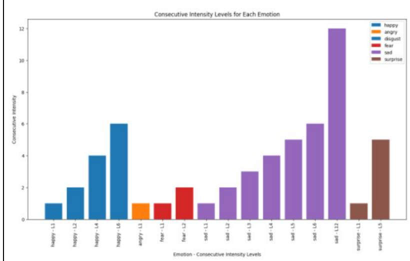


Figure 6. Intensity level of Emotions throughout the video based on fps.

Figure *7*. Prediction made by the algorithm.

Figure 7 shows prediction made by the algorithm based on the that which emotion is being held longest and which emotion has the highest intensity throughout the video, that what is type and genre of the video.

TABLE I. SOME SAMPLE TEST-CASES

Video Name	Intensity	Prediction
 [Hwa Sa - by Maria]	 Happy; Intensity Levels: [1, 2, 4, 6] Angry; Intensity Levels: [1] Fear; Intensity Levels: [1, 2] Sad; Intensity Levels: [1, 2, 3, 4, 5, 6, 12] Surprise; Intensity Levels: [1, 5]	This is a sad video with happy ending
 [LEE HI - by HOLO]	 Happy; Intensity Levels: [1, 3, 4] Fear; Intensity Levels: [1] Sad; Intensity Levels: [1, 2, 3, 11, 19]	This is a sad video with happy ending

V. CONCLUSIONS

In this study, we have delved into the temporal intricacies of emotion within video content, leveraging a robust Convolutional Neural Network (CNN) model for emotion recognition to predict the genre of the movie correctly. Our exploration of overall emotion distribution across frames has provided a foundational understanding of prevalent affective states while analyzing consecutive intensity levels, unveiling nuanced temporal patterns. Integrating data augmentation, learning rate scheduling, and meticulous dataset curation enhances the reliability of our findings. Beyond contributing to emotion recognition and genre prediction, this research opens avenues for refined model architectures and multimodal data exploration. Our work advances emotion dynamics comprehension and paves the way for context-aware emotion recognition systems, promising impactful applications across various domains, from entertainment to mental health diagnostics. However, it is crucial to acknowledge that the potential for broader applications and more profound insights exists, constrained only by the limited access to expansive datasets and hardware limitations. As we navigate these constraints, our findings lay the groundwork for future research endeavors, suggesting that further expansion and refinement of our approach hold promise for uncovering even more profound insights into the intricate interplay of emotions over time.

REFERENCES

[1] Cîrneanu A-L, Popescu D, Iordache D. New Trends in Emotion Recognition Using Image Analysis by Neural Networks, A Systematic Review. Sensors. 2023; 23(16):7092.

[2] Liao J, Lin Y, Ma T, He S, Liu X, He G. Facial Expression Recognition Methods in the Wild Based on Fusion Feature of Attention Mechanism and LBP. Sensors. 2023; 23(9):4204.

[3] Abdulsalam WH, Alhamdani RS, Abdullah MN. Facial emotion recognition from videos using deep convolutional neural networks. Int. J. Mach. Learn. Comput. 2019;9(1):14-19.

[4] Li H, Niu H, Zhu Z, Zhao F. Intensity-aware loss for dynamic facial expression recognition in the wild. In Proceedings of the AAAI Conference on Artificial Intelligence 2023; Vol. 37, No. 1, pp. 67-75.

[5] Wingenbach TS, Ashwin C, Brosnan M. Sex differences in facial emotion recognition across varying expression intensity levels from videos. PLoS one. 2018;13(1): e0190634.

[6] Küntzler T, Höfling TT, Alpers GW. Automatic facial expression recognition in standardized and non-standardized emotional expressions. Frontiers in psychology. 2021; 12:1086.

[7] Khan AR. Facial emotion recognition using conventional machine learning and deep learning methods: current achievements, analysis and remaining challenges. Information. 2022 May 25;13(6):268.

[8] Martinez A, Du S. A model of the perception of facial expressions of emotion by humans: research overview and perspectives. Journal of Machine Learning Research. 2012;13(5).

[9] Debnath T, Reza MM, Rahman A, Beheshti A, Band SS, Alinejad-Rokny H. Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity. Scientific Reports. 2022 Apr 28;12(1):6991.

[10] Gom-os DF, Yong KY. An empirical study on the use of a facial emotion recognition system in guidance counseling utilizing the technology acceptance model and the general comfort questionnaire. Applied Computing and Informatics. 2022 Oct 4.

[11] Kawashima T, Nomiya H, Hochin T. Facial Expression Intensity Estimation using Deep Convolutional Neural Network. InProceedings of the the 8th International Virtual Conference on Applied Computing & Information Technology 2021, pp. 7-12.

[12] Bryant DA, Deng S, Sephus N, Xia W, Perona P. Multi-Dimensional, Nuanced and Subjective-Measuring the Perception of Facial Expressions. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022, pp. 20932-20941.

[13] Maurya V, Dwivedi M, Singh R, Maurya D. Facial Emotion Recognition. Vol. 11 Issue IV Apr 2023, 4361-4366.

[14] Richoz AR, Lao J, Pascalis O, Caldara R. Tracking the recognition of static and dynamic facial expressions of emotion across the life span. Journal of vision. 2018;18(9):5;1-27.

[15] Mehta D, Siddiqui MFH, Javaid AY. Recognition of Emotion Intensities Using Machine Learning Algorithms: A Comparative Study. Sensors. 2019; 19(8):1897.