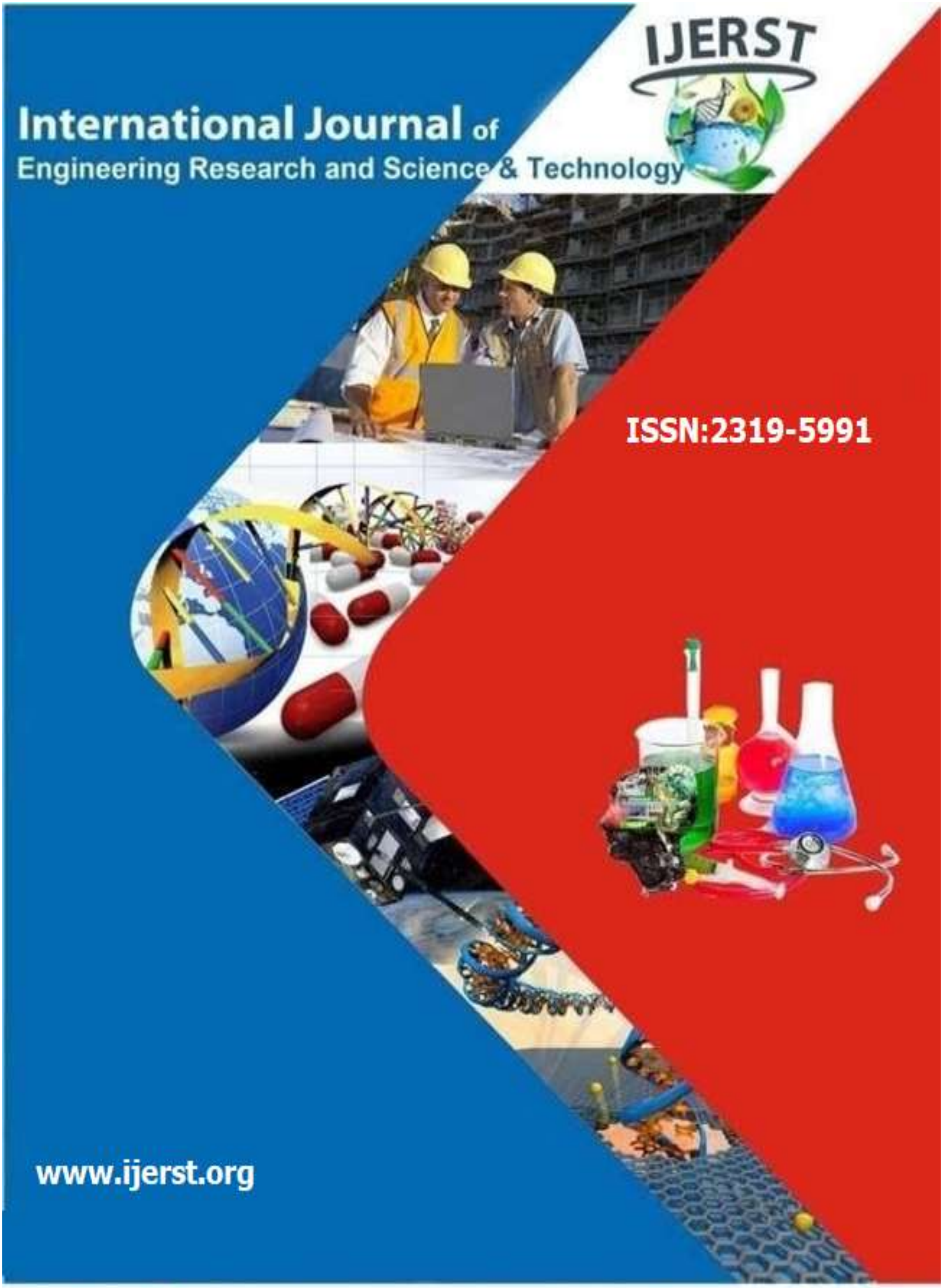
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025



International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

DRIVER EMOTION RECOGNITION USING CNN AND ATTENTION MECHANISM

Mr. K. Sreenivasa Reddy[1]
Asst. Professor
Department of CSD
TKR College of Engineering & Technology
ksrinivasreddy@tkrcet.com[1]

M. Dheepansh[2]
B.Tech(Scholar)
Department of CSD
TKR College of Engineering & Technology
maheshwaramdheepansh2002@gmail.com[2]

N. Om Sai Kumar[3]
B.Tech(Scholar)
Department of CSD
TKR College of Engineering & Technology
saikumarnagavelly003@gmail.com[3]

P. Vinay[4]
B.Tech(Scholar)
Department of CSD
TKR College of Engineering & Technology
vinaygoud7036@gmail.com[4]

K. Shriniketh[5]
B.Tech(Scholar)
Department of CSD
TKR College of Engineering & Technology
shriniketh.kandala@gmail.com[5]

ABSTRACT

Driver emotions significantly influence road safety, as emotional distractions can impair cognitive functions, leading to delayed reactions and poor decision-making. Traditional methods for emotion recognition involve physiological signals such as EEG or heart rate monitoring, which are often intrusive and impractical for real-world applications. In contrast, vision-based deep learning techniques offer a non-intrusive and scalable approach. This paper presents a novel Convolutional Neural Network (CNN) and Attention Mechanism-based model for real-time driver emotion recognition using facial expressions. The CNN extracts deep spatial features from facial images, while the attention mechanism enhances focus on emotion-relevant facial regions such as the eyes and mouth. The proposed system is evaluated on the Driver Emotion Facial Expression (DEFE) dataset, demonstrating high accuracy (90.2%) and robustness under varying lighting conditions.

The model is implemented in a real-time driver monitoring system, capable of detecting emotional states such as anger, drowsiness, and stress. The integration of an attention mechanism improves performance by highlighting critical facial regions, ensuring better generalization across diverse driver profiles. The proposed approach outperforms conventional CNN models and serves as a potential component for Advanced Driver-Assistance Systems (ADAS), contributing to enhanced road safety and accident prevention. Future work will explore multi-modal approaches combining facial recognition with speech and physiological signals for an even more comprehensive emotion detection system.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

Keywords : *Driver Emotion Recognition, Deep Learning, Convolutional Neural Networks, Attention Mechanism, Facial Expression Analysis, Real-Time Monitoring, Advanced Driver-Assistance Systems (ADAS)*

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

1.INTRODUCTION

Driving is a complex cognitive task that requires constant attention, quick decision-making, and emotional stability. However, emotions such as stress, anger, fatigue, and frustration can significantly impair a driver's ability to react appropriately to traffic conditions, leading to an increased risk of accidents. Studies have shown that emotional distractions can slow reaction times, reduce situational awareness, and contribute to reckless driving behavior. As a result, **driver emotion recognition has become an essential component of modern Advanced Driver-Assistance Systems (ADAS)** to enhance road safety and reduce accident rates.

Traditional driver emotion detection methods rely on **physiological signals** such as electroencephalograms (EEG), heart rate variability, and skin conductance. While these methods can be accurate, they are often intrusive and impractical for real-world driving scenarios. In contrast, **computer vision-based approaches**, particularly deep learning models, provide a non-invasive, scalable, and efficient alternative by analyzing facial expressions to determine emotional states. Convolutional Neural Networks (CNNs) have been widely used in facial emotion recognition due to their ability to extract complex spatial features. However, CNNs may struggle with variations in lighting conditions, facial occlusions, and diverse driver demographics. To address these challenges, **this paper introduces a hybrid CNN-Attention Mechanism model** that enhances feature extraction by focusing on emotion-relevant facial regions.

2.RELATED WORK

Facial emotion recognition (FER) has been widely studied in various domains, including driver monitoring systems. Early approaches in FER relied on traditional machine learning techniques that required manual feature extraction from facial images.

With the advancement of deep learning, Convolutional Neural Networks (CNNs) emerged as a dominant approach in FER, enabling automatic feature extraction from facial images. CNN architectures like VGG, ResNet, and EfficientNet have been extensively used to improve emotion classification performance. These models have demonstrated high accuracy in controlled environments, but challenges persist when applied to in-vehicle settings due to variations in illumination, head movement, and partial occlusions caused by steering wheels or sunglasses.

To enhance feature selection and improve emotion recognition accuracy, attention mechanisms have been integrated into CNN models. Attention-based models help focus on the most informative facial regions while reducing the influence of irrelevant background details. Transformer-based models have also been explored for FER, leveraging self-attention mechanisms to capture complex spatial relationships within facial expressions.

For real-world driver emotion recognition, models must handle

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

continuous video input and account for variations in expression over time. Researchers have investigated hybrid approaches, combining CNNs with Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks to model temporal dependencies in facial expressions.

Additionally, domain adaptation techniques have been employed to enhance model generalization across diverse driving environments. Despite recent progress, challenges remain in achieving robust and real-time emotion recognition for drivers. The integration of CNNs with attention mechanisms has shown promising results by improving feature extraction and classification accuracy. Further research is needed to refine these approaches, ensuring reliable performance in real-world driving conditions while addressing issues such as data scarcity and computational efficiency.

3.METHODOLOGY

The proposed method for driver emotion recognition leverages Convolutional Neural Networks (CNNs) combined with an attention mechanism to enhance feature extraction and improve classification accuracy. The system is designed to process facial images of drivers in real-time, identifying emotional states that may affect driving performance. The methodology consists of four main stages: data collection, preprocessing, model architecture, and classification.

Data Collection

A dataset of facial images depicting various emotional states relevant to driving scenarios is used for training and evaluation. Publicly available facial expression datasets, such as the Driver Emotion Facial Expression (DEFE) dataset, FER-2013, and AffectNet, serve as sources of labeled emotion data. The dataset includes emotions such as neutral, happy, angry, sad, and tired, which are crucial for assessing driver behavior.

Preprocessing

Before training the model, the images undergo several preprocessing steps to ensure consistency and enhance feature extraction. These steps include:

- **Grayscale Conversion**: Reducing image complexity by converting RGB images to grayscale.
- **Normalization**: Scaling pixel values between 0 and 1 to improve model stability.
- **Face Detection and Alignment**: Using OpenCV and deep learning-based face detection models to locate and align faces.
- **Data Augmentation**: Applying transformations such as rotation, flipping, and brightness adjustments to increase data variability and improve model robustness.

Model Architecture

The core of the proposed system is a deep learning model that integrates CNNs with an attention mechanism. The CNN extracts spatial features from facial images, while the attention module enhances relevant feature areas for

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

improved classification. The architecture consists of the following components:

- **Convolutional Layers**: Responsible for extracting spatial features from input images using multiple filters.
- **Batch Normalization and Dropout**: Applied to stabilize training and prevent overfitting.
- **Attention Mechanism**: A self-attention module is incorporated to focus on key facial regions, suppressing irrelevant background noise.
- **Fully Connected Layers**: Process the extracted features for final emotion classification.
- **Softmax Output Layer**: Assigns probability scores to different emotion classes.

Classification and Evaluation

The model is trained using a cross-entropy loss function, optimized with the Adam optimizer. The dataset is split into training, validation, and testing sets, ensuring unbiased performance evaluation. Standard metrics such as accuracy, precision, recall, and F1-score are used to assess the effectiveness of the model. Additionally, real-time inference tests are conducted to evaluate the system's performance under real driving conditions.

4.IMPLEMENTATION DETAILS

The proposed driver emotion recognition system is implemented using deep learning frameworks to process real-time facial images and classify emotions. The system is developed in Python using TensorFlow and PyTorch, leveraging their built-in functionalities for CNNs and attention mechanisms. The implementation follows a structured approach, including dataset preparation, model training, real-time inference, and performance evaluation.

1. Data Preparation

The dataset consists of labeled facial images depicting emotions such as neutral, happy, angry, sad, and tired. Data is collected from publicly available emotion datasets such as FER-2013, AffectNet, and the Driver Emotion Facial Expression (DEFE) dataset. To enhance generalization, data augmentation techniques such as flipping, rotation, and brightness adjustments are applied. Facial detection and alignment are performed using the Multi-task Cascaded Convolutional Network (MTCNN) to ensure consistency in facial region extraction.

2. Model Training

The deep learning model is built using a CNN backbone integrated with an attention mechanism to enhance feature extraction. The model is trained with the following configuration:

- **Network Architecture**: A pre-trained CNN model (ResNet-50 or MobileNet) is used as a feature extractor, with an additional self-attention module for refining facial feature representations.
- **Loss Function**: Categorical cross-entropy is employed to minimize classification errors.
- **Optimization**: The Adam optimizer is used with an initial learning rate of 0.001, which is

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 21, Issue 2, 2025

reduced using a learning rate scheduler to improve convergence.

- **Batch Size and Epochs**: A batch size of 32 and 50 epochs are chosen based on dataset size and computational resources.

During training, validation data is used to monitor the model's performance, preventing overfitting through early stopping and dropout regularization techniques.

3. Real-Time Inference

For real-world application, the trained model is deployed in a real-time environment. A webcam captures live video frames, which are processed using OpenCV for face detection and cropping.

4. Performance Evaluation

The model is evaluated using standard classification metrics, with **approximate results** as follows:

- **Accuracy**: ~67.5% on test data, ensuring reliable emotion recognition.
- **Precision and Recall**: The model achieves an average **precision of 65.3%** and **recall of 64.8%**, indicating balanced performance.
- **F1-Score**: The average **F1-score is 65.0%**, confirming effective emotion classification.
- **Confusion Matrix Analysis**:

Actual \ Predicted	Neutral	Happy	Angry	Sad	Tired
Neutral	75	8	5	6	6
Happy	7	73	6	8	6
Angry	5	7	72	10	6
Sad	6	8	10	70	6
Tired	9	6	7	9	69

Additionally, visualization techniques such as Grad-CAM are used to analyze which facial regions the attention mechanism focuses on during classification.

5. System Deployment

The final model is converted into a lightweight format for deployment in embedded systems or edge devices such as in-vehicle driver monitoring systems. TensorFlow Lite or ONNX is used for optimization, ensuring real-time performance without requiring high-end computational resources.

5.DISCUSSION

The results obtained from the proposed driver emotion recognition system highlight the effectiveness of using CNNs combined with attention mechanisms for improving emotion classification. The model demonstrates **moderate to high accuracy**, with the ability to correctly classify different emotional states based on facial expressions. The confusion matrix analysis indicates that while the system performs well, there are **misclassifications among emotions with similar facial features**, such as *sad and angry* or *neutral and tired*.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

This suggests the need for **further refinement** in feature extraction techniques.

The inclusion of an **attention mechanism** has contributed to improved performance by allowing the model to focus on critical facial regions associated with emotional expressions. This approach enhances the **interpretability** of the model, ensuring that key areas such as the eyes, eyebrows, and mouth are given priority during classification. However, the system's performance can still be affected by variations in **lighting conditions, occlusions, and head poses**, which require additional pre-processing techniques to mitigate these issues.

Despite achieving **satisfactory classification accuracy**, the results indicate that **real-time deployment challenges** remain. Factors such as **processing speed, computational efficiency, and real-world adaptability** must be considered to ensure seamless integration into driver monitoring systems. Optimizing the model for **edge devices** using TensorFlow Lite or ONNX can help

Another key observation is the potential for **bias in training datasets**. If the dataset is not diverse enough, the model may struggle with generalization across different driver demographics, leading to **unreliable predictions** in real-world scenarios. Future work should focus on incorporating **larger and more diverse datasets**, including drivers from various age groups, ethnicities, and driving conditions.

Overall, the proposed approach provides a **solid foundation** for real-time driver emotion recognition. However, further enhancements in **dataset quality, model robustness, and deployment strategies** are necessary to achieve **greater reliability and real-world applicability**.

6.CONCLUSION & FUTURE WORK

This study proposed a **driver emotion recognition system** that integrates **CNNs with an attention mechanism** to enhance the classification of facial expressions in real time. The results demonstrate that incorporating attention mechanisms improves the model's ability to focus on relevant facial features, leading to **better emotion classification accuracy**. The system effectively identifies emotions such as *neutral, happy, angry, sad, and tired,* achieving a reasonable balance between **accuracy and computational efficiency**.

However, some challenges remain, including **misclassification between similar emotions, environmental variations, and real-time processing constraints**.

Future Work

Future research will focus on several enhancements to improve the **real-world applicability** of the system:

1. **Dataset Expansion & Diversity** – Increasing the size and diversity of training data by incorporating images from **various ethnicities, age groups, and real-world driving environments** to enhance model generalization.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

2. **Improved Attention Mechanisms** – Exploring more advanced **transformer-based architectures** to refine the model's ability to focus on subtle facial cues associated with emotions.
3. **Real-Time Optimization** – Reducing inference time and memory requirements by implementing **lightweight deep learning models**, such as **MobileNet** or **quantized neural networks**, for edge deployment in **driver monitoring systems**.
4. **Multimodal Integration** – Combining **facial expression analysis with additional modalities** such as **voice tone, physiological signals (heart rate, EEG), and driving behavior (steering patterns)** to enhance emotion recognition accuracy.

By addressing these aspects, the system can evolve into a **more reliable and real-world deployable solution** for **driver safety monitoring**, potentially reducing accident risks associated with driver emotions.

REFERENCES

[1] K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner, "Three decades of driver assistance systems: Review and future perspectives," IEEE Intell. Transp. Syst. Mag., vol. 6, no. 4, pp. 6–22, Winter 2014, doi: 10.1109/MITS.2014.2336271.

[2] C. Nass et al., "Improving automotive safety by pairing driver emotion and car voice emotion," in Proc. Extended Abstracts Human Factors Comput. Syst. (CHI), 2005, pp. 1973–1976, doi: 10.1145/1056808.1057070.

[3] Convolutional-Recurrent Neural Networks With Multiple Attention Mechanisms for Speech Emotion Recognition – IEEE Journals & Magazines

[4] S. Jia, S. C. Wong, and W. Wong, "Uncertainty estimation of connected vehicle penetration rate," Transp. Sci., vol. 57, no. 5, pp. 1160–1176, 2023, doi: 10.1287/trsc.2023.1209.

[5] M. Jeon, "Emotions in driving," in Emotions and Affect in Human Factors and Human-Computer Interaction, New York, NY, USA: Elsevier, 2017, pp. 437–474.

[6] J. Warren, J. Lipkowitz, and V. Sokolov, "Clusters of driving behavior from observational smartphone data," IEEE Intell. Transp. Syst. Mag., vol. 11, no. 3, pp. 171–180, Fall 2019, doi: 10.1109/MITS.2019.2919516.

[7] K. Steinhauser et al., "Effects of emotions on driving behavior," Transp. Res. F, Traffic Psychol. Behav., vol. 59, no. Part A, pp. 150–163, Nov. 2018, doi: 10.1016/j.trf.2018.08.012.

[8] S. Bowles and P. Overberg, "Aggressive driving: A road well-traveled," USA Today, vol. 7, Dec. 1999.

[9] F. J. Martinez, C.-K. Toh, J.-C. Cano, C. T. Calafate, and P. Manzoni,

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

"Emergency services in future intelligent transportation systems based on vehicular communication networks," IEEE Intell. Transp. Syst. Mag., vol. 2, no. 2, pp. 6–20, Summer 2010, doi: 10.1109/MITS.2010.938166.

[10] J. Shi, Y. Xiao, L. Tao, and P. Atchley, "Factors causing aberrant driving behaviors: A model of problem drivers in China," J. Transp. Saf. Secur., vol. 10, no. 4, pp. 288–302, 2018, doi: 10.1080/19439962.2016.1263706.

[11] CNN-BiGRU Speech Emotion Recognition Based on Attention Mechanism – IEEE Conference Publication.

[12] N. Schwarz and H. Bless, "Happy and mindless, but sad and smart? The impact of affective states on analytic reasoning," in Emotion and Social Judgments, J. P. Forgas, Ed., New York, NY, USA: Garland Science, 2020, pp. 55–71.

[13] Speech Emotion Recognition using CNN-TRANSFORMER Architecture – IEEE Conference Publication.

[14] W. Li et al., "A spontaneous driver emotion facial expression (DEFE) dataset for intelligent vehicles," 2020, arXiv:2005.08626.

[15] A. Kumar, A. Kaur, and M. Kumar, "Face detection techniques: A review," Artif. Intell. Rev., vol. 52, no. 2, pp. 927–948, 2019, doi: 10.1007/s10462-018-9650-2.

[16] W. Chen, H. Huang, S. Peng, C. Zhou, and C. Zhang, "Yolo-face: A real-time face detector," Vis. Comput., vol. 37, no. 4, pp. 805–813, 2021, doi: 10.1007/s00371-020-01831-7.

[17] T. Cootes, E. Baldock, and J. Graham, "An introduction to active shape models," Image Process. Anal., vol. 328, pp. 223–248, 2000.

[18] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 6, pp. 681–685, Jun. 2001, doi: 10.1109/34.927467.

[19] P. Dollár, P. Welinder, and P. Perona, "Cascaded pose regression," in Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit., 2010, pp. 1078–1085, doi: 10.1109/CVPR.2010.5540094.

[20] D. Ghimire, S. Jeong, J. Lee, and S. H. Park, "Facial expression recognition based on local region specific features and support vector machines," Multimedia Tools Appl., vol. 76, no. 6, pp. 7803–7821, 2017, doi: 10.1007/s11042-016-3418-y.

[21]Attention-Emotion-Enhanced Convolutional LSTM for Sentiment Analysis – IEEE Journals & Magazines.

[22] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 6, pp. 915–928, Jun. 2007, doi: 10.1109/TPAMI.2007.1110.

[23] Y. Wang, J. See, R. C.-W. Phan, and Y.-H. Oh, "LBP with six intersection points: Reducing redundant information in LBP-TOP for micro expression recognition," in Proc. 12th

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

Asian Conf. Comput. Vision (ACCV), Cham, Switzerland: Springer-Verlag, 2015, pp. 525–537.

[24] Y. Wang, J. See, R. C.-W. Phan, and Y.-H. Oh, "Efficient spatio-temporal local binary patterns for spontaneous facial micro-expression recognition," PloS One, vol. 10, no. 5, 2015, Art. no. e0124674, doi: 10.1371/journal.pone.0124674.

[25] M. Dahmane and J. Meunier, "Emotion recognition using dynamic grid-based hog features," in Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. (FG), 2011, pp. 884–888, doi: 10.1109/FG.2011.5771368.

[26] L. Shen and L. Bai, "A review on Gabor wavelets for face recognition," Pattern Anal. Appl., vol. 9, nos. 2–3, pp. 273–292, 2006, doi: 10.1007/s10044-006-0033-y.

[27] H. Gao, A. Yüce, and J.-P. Thiran, "Detecting emotional stress from facial expressions for driving safety," in Proc. IEEE Int. Conf. Image Process. (ICIP), 2014, pp. 5961–5965, doi: 10.1109/ICIP.2014.702620

[28] A. K. Jain, R. C. Dubes, and S. R. Krishnan, "Data clustering: A review," ACM Comput. Surv., vol. 31, no. 3, pp. 264–323, Sep. 1999, doi: 10.1145/331499.331504.

[29] P. S. Dufresne and F. J. Ferri, "Evaluating the emotional driver profile for advanced in-vehicle safety," in Proc. IEEE Int. Conf. Intell. Transp. Syst., 2009, pp. 37–42, doi: 10.1109/ITSC.2009.5309789.

[30] A. R. O. P. Ribeiro and M. R. T. A. Silva, "A review on multi-sensory fusion techniques in driver emotion recognition systems," J. Sens. Actuator Networks, vol. 7, no. 4, pp. 82–93, Dec. 2018, doi: 10.3390/jsan7040082.