Efficient Real-Time Eye Gaze Tracking Detection for Human-Computer Integration Using Advanced Techniques

1st Laith H. Alzubaidi
The Islamic University
Najaf, Iraq
laith.h.alzubaidi@gmail.com

2nd Abbas Hameed Abdul Hussein
College of Pharmacy, Ahl Al Bayt University
Karbala, Iraq
Abdul.hussien@abu.edu.iq

3rd Mohammed Ayad Alkhafaji
National University of Science and Technology
Dhi Qar, Iraq
mohammed.alkhafaji@nust.edu.iq

4th N Shilpa
Department of ECE, School of Engineering
SR University
Warangal, Telangana, India
shilpa.ece6@gmial.com

5th Tejaswini N P
Department of information Science and Engineering
Nitte Meenakshi Institute of Technology
Bengaluru, India
tejaswini.np@nmit.ac.in

2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC) | 979-8-3503-1702-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICMNWC60182.2023.10436021

***Abstract*—The eye-gaze tracking for detecting different types of eye movements in a continuous stream of gaze data is limited, as it involves Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) features. These features employ individual detectors, each capable of detecting a single movement. However, in some applications, eye tracking may not occur due to issues such as overfitting and noise in image detection. These challenges are addressed in our method, which utilizes CNN and RNN. The goal is to design algorithms that can accurately differentiate between nuanced emotional states, thereby enhancing the precision of eye tracking. To achieve this objective, we employ the CNN State, which is trained on a dataset of human eye images captured by intelligent eyeglasses to obtain an eye state recognition model. This allows for efficient transfer learning, enabling the eye emotion tracking model to benefit from knowledge gained in other domains, even with limited labeled emotion data. It's worth noting that RNNs have a limited memory capacity, which can hinder their ability to effectively capture and retain information over extended periods. Nevertheless, our experimental results demonstrate high-level performance on the FER2013 dataset, achieving an accuracy of 0.97. This performance surpasses other existing models such as ResNet, K-Nearest Neighbors, and VGG19.**

Keyworks—Convolutional Neural Network, Deep Learning, Eye Tracking, Human Interaction and Recurrent Neural Network.

I. INTRODUCTION

Eye gaze tracking devices have become essential for efficient human-computer interaction applications. To evaluate vision information and process gaze tracking, various techniques have been widely applied, examining hidden processes and inspecting eye-tracking methodologies. Continuous observation of gaze routes for test subjects involves detecting the human eye [1]. The emotion of an individual plays a central role in human interaction, and detection methods are employed to access the image. However, the dullness of the image makes it challenging to exploit facial geometry, and temporal information is also essential for emotion recognition [2]. Different types of eye-tracking setups include categories such as free hand analysis of eye-tracking signals and head-free setups that may contain one or more eye trackers, typically mimicking the appearance of glasses. Head-boxed setups contain one or more remote eye trackers placed in a fixed location in front of the participant [3].

In eye-computer interaction, one method involves pupil detection and state recognition, but the challenge lies in occlusion of information and the inability to clearly detect pixels. The use of smart devices for real-time observation of eye movements has improved the performance in eye behaviour using deep learning methods [4]. Classification involves training with manually extracted features. Research in uncontrolled environments has focused more on deep learning-based techniques in eye detection. The utilization of deep convolutional neural networks (DCNN) and convolutional neural networks (CNN) for image analysis and object detection has provided more information [5]. The aim of this work is to address both of these limitations and fundamentally different approach for eye movement detection is proposed, involving learning a single detector end-to-end directly from raw gaze data to different types of eye movement on CNN specific type to use the RNN using image classification and appearance-based gaze [6]. The main contributions of this paper are below:

- The Convolutional Neural Network detects eye movements utilized some type such as fixations, saccades, and smooth pursuits simultaneously from a continuous sequence of gaze input data obtained through eye tracking.
- These methods are utilized to detect eye state recognition in small processor devices, supporting a drowsiness warning system without compromising accuracy.
- If eye tracking data is accompanied by information Recurrent Neural Networks (RNNs) can be integrated into multimodal models for a more holistic understanding of emotions in human-computer interaction.

The paper is organized as follows: section 2 provides a literature review that summarizes Eye Tracking detection. Section 3 introduces the proposed method utilized CNN and RNN. Section 4 discuss the result and comparative analysis. Section 5 discuss the conclusion.

II. LITERATURE

Yusra Khalid Bhatti *et al*. [7] presented a Convolutional Neural Network (CNN) for facial expression feature representation generated using the 2D-CNN tool. The utilized feature model had layers that allowed the extraction of high-level, medium-level, and low-level features. The network employed an acyclic graph, arranging layers sequentially, similar to Alex Net. In this architecture, where the input dimension is high, a directed acyclic graph with multiple layers was processed to achieve efficient detection. The dataset explored realistic environments where emotions were expressed, taking into consideration dimensional values. Expressions were classified, but there was a lack of action in finding a unique solution, leading to overfitting and delays in computational time.

Akhilesh Kumar *et al*. [8] developed a K-Nearest Neighbors (KNN) algorithm, which was a simple and intuitive method. It was easy to implement and understand, making it accessible for applications in various fields, including eye emotion tracking. It could perform well with small datasets, which was advantageous in cases were collecting a large labelled dataset for training a complex model was not feasible. It was considered a lazy learner because it deferred the learning process until the prediction phase. This made it adaptable to changes in the dataset, allowing for dynamic updates without the need for retraining the entire model. The computation of distances between data points during the inference phase could be computationally intensive, especially as the dataset size increased. This could impact real-time processing and might be a limitation for human interactive applications.

M. Kalpana Chowdary et al. [9] developed a Convolutional Neural Network based on VGG19, which was designed with multiple convolutional layers, allowing it to automatically learn hierarchical features. This was beneficial for eye emotion tracking, as emotions often manifested through subtle spatial patterns that could be captured through these layers. The VGG19 pre-trained models on large datasets for general image classification tasks were used. This facilitated transfer learning, enabling the model to benefit from knowledge gained in other domains and improving its performance with limited labelled emotion data. VGG19 had a large number of parameters, making it computationally intensive, especially during training. This could be a limitation for real-time processing in interactive HCI applications and might require efficient hardware or model optimization.

Lei Zhao et al. [10] developed a Deep Convolutional Neural Network (DCNN) to detect eye recognition by updating the function that could automatically detect the eye region's location, and then, after cropping the image, the dataset could be fed into the network. The hidden layer, input layer, and output layers were divided, and the image was processed in between. This method reduced the size, and fully connected layers were applied similarly to the image, considering the dimensions of the input and output feature maps. Many expressions and images were labelled in the dataset, predicting the challenges of the eye state in the classified dataset. DCNNs might have struggled with occlusions, such as eyeglasses or partial obstructions of the eyes. Robustness to occlusions was essential for real-world applications but could be challenging to achieve consistently.

Kaviya P and Arumugaprakash T [11] presented an eye gaze tracking method utilizing Convolutional Neural Network (CNN) to explore and recognize facial expressions. The method was implemented for real-time processing, a crucial requirement for Human-Computer Interaction (HCI) applications. This enabled seamless integration of emotional feedback into interactive systems, enhancing the user experience. CNNs often required large labelled datasets for effective training. Obtaining diverse and representative datasets for various emotional expressions in different individuals could be challenging, potentially limiting the model's performance. CNNs were susceptible to overfitting, especially when the model was trained on a limited dataset. This could result in poor generalization to new, unseen data, impacting the model's ability to accurately track emotions.

III. PROBLEM STATEMENT

- The challenge lies in optimizing algorithms and processing pipelines to meet the stringent time constraints of interactive systems, ensuring the timely and seamless integration of emotional feedback into human-computer interactions.
- In human-resource interaction, predicting eye movement and rotation at the edges becomes challenging due to occlusions occurring during eye movement. This results in a lack of information and increases computational time.

IV. OBJECTIVE

- The objective is to adjust the angle of the image through rotation and scaling to better predict eye movement and address issues such as the appearance of wrinkles around the eyes, which are not clearly predicted at the edges.
- Specifically, the challenge involves handling occlusions during pursuit, making it complex to predict actions accurately and reducing the risk of overfitting. The labeling data is not aligned, further contributing to the computational time challenges.

V. PROPOSED METHODOLOGY

In this paper, the proposed methodology involves end-to-end training for eye movement detection, utilizing the Convolutional Neural Network (CNN) method. Eye tracking takes into consideration both horizontal and vertical screen coordinates, sampling images frequently to capture the eye gaze. A Recurrent Neural Network is combined to predict eye tracking. The FER2013 dataset is considered, comprising 35,567 grayscale images at a resolution of 48×48. These images are classified for emotions and occlusions to handle human interaction. Fig. 1 represented the expression of the eye tracing open and close eye.

Fig. 1. Sample image of the Eye Tracking for human interaction

A. Dataset

In this paper, the FER2013 dataset was utilized for data representation, considering 35,345 images with a dimension of 49×49 resolution. The data was classified into three classes: Fixation, Saccades, and Pursuit, concentrating on human-computer interaction. Pursuit was complex to access as it involved continuous concentration on an object, and the paper also focused on detecting emotion variations in the images using the FER2013 dataset. Table I represents the emotions in faces [12].

TABLE. I. NUMBER OF THE IMAGE TO EACH EMOTION OF THE DATASET FER2013

Emotion	Happy	Anger	Disgust	Sad	Surprise	Fear
Images	8989	4653	548	6014	4001	5142

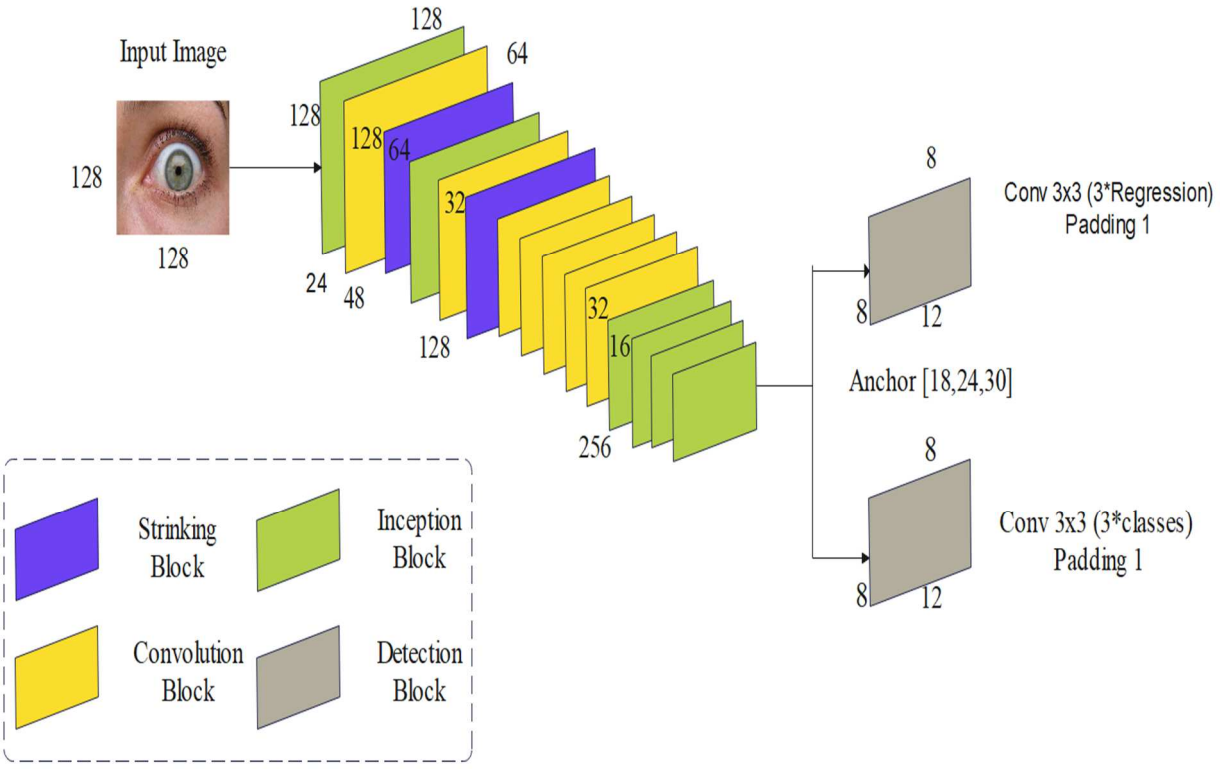

Fig. 2. The architecture of the CNN block diagram for Eye tracking

These feature channels were fed into a detailed view of the CNN, which employed 1D convolutional layers for eye movement classification, as explained above the diagram. The CNN considered four one-dimensional convolutional layers with rectified linear units (ReLu) in the first layer and sigmoid in the last layer. Each convolutional layer was followed by batch normalization and an average pooling layer with specific size values. Rotation and resizing of the image were utilized to ensure the detection of all emotions. A dropout layer with a rate of 0.4 was added before the first dense layer to prevent overfitting. Finally, the exploration in the image used the cross-entropy metric. As observed, the accuracy increment is directly proportional to the number of epochs. However, the accuracy remains constant after the 81st epoch, indicating that the accuracy does not increase significantly after that period.

To initialize the filter values and partial values, Gaussian distribution in the pooling layer, and output of the layer input equation (1) are considered in the convolutional layers.

B. Convolutional Neural Network

The research utilized the Convolutional Neural Network (CNN) in the field of pattern recognition to construct a learning model. The constructed CNN model was employed to detect blinks in the subjects. In this method, the CNN model inputted images of open and closed eyes with a resolution of 64 × 64, and under convolutional operations, it extracted feature values from the input values. The system then built a recognition model based on these feature values and applied it. The convolutional layers combined different local structures to present more useful features in a region. The experiment adjusted the input image of the convolutional layers to make the brightness more diversified and adaptable, as shown in Fig. 2, the architecture of the eye tracking block diagram [13].

$$C_{nv} = \frac{(K_v + 2P_a - F_v)}{S_{tr}} + 1 \qquad (1)$$

Where K_v is the dimension of the input file, P_a represents the padding of the dimension of the filter, and the stride is denoted by S_{tr}. The layer is a measurement of the neural network between convolution layers. To manage overfitting and reduce the scale of the network, two features are utilized: average pooling and full pooling. The objective of max-pooling values is to reduce the dimension of the layer to help with computational cost. In this technique, the network is down sampled. The formula considers for the architecture of the pooling layer, as shown in equation (2), is as follows.

$$P_{ol} = \frac{(I_d - F_v)}{S_{tr}} + 1 \qquad (2)$$

Where I_d is the dimension of the input to the pooling layer, S_{tr} represents the stride, and F_v represents the dimension of the filter. Since we cannot directly perform max-pooling, it is

divided into two parts. The average pooling layer is used to minimize overfitting and reduce the model's size. This layer is performed with dimensions of $1 \times 1 \times d$ to reduce the size of the single channel. The activation function is utilized on the input-single to convert it to the next layer's input for the neurons. In ReLU, not all neurons are activated at the same time. Negative values are set to zero, as explained in equations (3) and (4). The hyper-parameter a_i is used for this purpose, and adjusting its values results in a better eye tracking outcome. This method utilizes high-speed processing, reducing the training time [14].

$$f(x) = \begin{cases} 0 \; for \; x < 0 \\ x \; for \; x \geq 0 \end{cases} \tag{3}$$

$$f(x_i) = \begin{cases} a_i x_i \; for \; x < 0 \\ x_i \; for \; x \geq 0 \end{cases} \tag{4}$$

In the Convolutional Neural Network, the loss function plays a significant role as it measures the values for the strength of the model, increasing and reducing the risk of the module function. This type of NN is named convolutional because it performs the convolution operation with the inputs to produce useful information. The convolution layers adjust some learned parameters to extract the most useful information from the given data. The filter searches for a particular feature in the image, looking for a specific pattern in the whole image using just one filter.

C. Recurrent Neural Network

In the eye tracking utilized another method is Recurrent Neural Network (RNN) inspired for the natural language of the processing in the raw data. To predicated the image robust the classified the action the input image signals are normalization of the dimensional to aggregate the layer. To fed into the layer utilized the LSTM model it gives best compare to another image some of the state it facing lack of the information. There are used to small size of the model parameters is load to in the network. Then layer extracted the predicated values of the direction to success rate is 0.6 and 0.45. there are three 1D convolution matrices for the binary dataset with the kernel size of $[25 \times 1], [76 \times 1]$, and $[126 \times 1]$. It represent for the kernel size to be the times of the data sample rate is 250Hz. In the second will be calculated in the matrices and also design the designed for features down sampling in the max-pooling is selected to the kernel size of the layer. Fig. 3. Shown as schematic of the CNN relevant RNN [15].

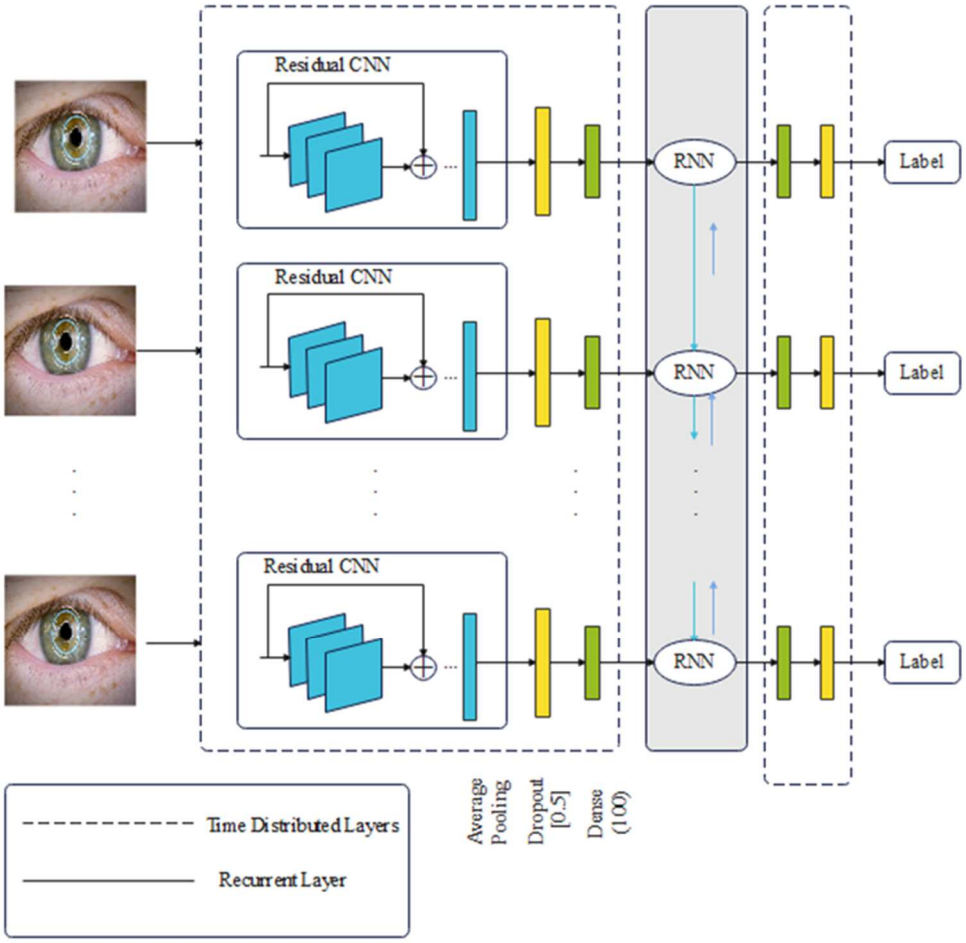

Fig. 3. Schematic of the proposed CNN-RNNs for relevance detection

In eye tracking, another method utilized is Recurrent Neural Network (RNN), inspired by natural language processing for raw data. To predict robust images and classify actions, the input image signals are normalized dimensionally to aggregate the layer. The LSTM model is then utilized in the layer, providing better results compared to other image processing methods, although it may face a lack of information in some states. The use of a small size for model

parameters reduces the load on the network. The layer extracts predicted values for the direction, with success rates of 0.6 and 0.45. Three 1D convolution matrices are used for the binary dataset with kernel sizes of $[25 \times 1], [76 \times 1]$, and $[126 \times 1]$.The kernel size represents the times of the data sample rate, which is 250Hz. The second matrix is calculated, and it is designed for features downsampling. Max-pooling is selected for the kernel size of the layer.

The RNN module is consider in the three inception modules for the pointed the eye ball reflection that convolution layer with the kernel size is 1×1, and (5) is shown as below;

$$y = \mathfrak{F}(x_0, \{W_i\}) + x \tag{5}$$

Where x_0 is the input vector, $\mathfrak{F}(x_0, \{W_i\})$ represents the residual mapping of the vector input, x is the output of the layer, and y is the summation of the RNN layer output and inception of the module. This effectively solves the problem of overfitting, ensuring consistency in the input and output dimensions of the modules. RNN has been implemented for attention level detection, typically using a series of data to classify shapes and eye hair. This helps RNN efficiently access the image and decreases the over time of the input, replacing the sensitivity to input of the activation of the hidden layer. The network tends to forget the overfitting problem, reducing the lack of information. The RNN regularizes the hidden layer, which is substituted by the memory block of the dataset, extending information and reducing the vanishing gradient. Then, the output layer is adapted to the units of the size.

Next, to reduce the issues of eye action and occlusion in the data and dropout the overfitting, methods are well-suited for processing sequential data. In the context of eye emotion tracking, RNNs can effectively capture temporal dependencies in eye movement sequences, allowing for a more comprehensive understanding of emotional expressions. They have the ability to retain the memory of past inputs, which is valuable for analyzing the evolution of emotions over time. This feature is beneficial in tracking the dynamic nature of eye movements associated with different emotional states. However, RNNs are prone to vanishing and exploding gradient problems, especially in long sequences. This can impact the network's ability to capture dependencies in lengthy eye movement patterns. While RNNs can capture short-term dependencies effectively, they may struggle to model long-term dependencies in sequences, potentially limiting their ability to recognize subtle and prolonged changes in emotional states.

VI. Experimental Result

In this analysis, the two combined methods of CNN and RNN for eye gaze detection are based on detecting images. Fast motion blur in some action frames can distort the spatial content of instruments, making them even invisible to the human eye. The FER2013 dataset, with 3589 images, is utilized for testing denoising and processing. Simultaneously, low resolution processing techniques, such as rotation and scaling, are applied to faces that have already been occluded in the horizontal and vertical directions. CNNs are designed to capture spatial hierarchies in data, which is particularly beneficial for eye emotion tracking. Emotions are often expressed through subtle spatial patterns in the eyes, and CNNs can discern these intricate patterns, enhancing the precision of emotion recognition. Performance metrics are defined using the equations presented in the animal image context as (6), (7), (8), and (9).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

$$Precision = \frac{TP}{TP+FP} \tag{7}$$

$$Recall = \frac{TP}{TP+FP} \tag{8}$$

$$F1_{score} = \frac{2 \times Precision \times Sensitivity}{Precision+Sensitivity} \tag{9}$$

Where, TP, TN, FP and FN illustrate the True Positive, True Negative, False Positive, False Negatives respectively.

A. Quantitative and Qualitative Analysis

This paper presents a comprehensive quantitative and qualitative analysis of the proposed models, namely Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). CNNs often require large labelled datasets for effective training, and obtaining diverse and representative datasets for various emotional expressions in different individuals can be challenging, potentially limiting the model's performance. KNN is considered a lazy learner because it defers the learning process until the prediction phase. VGG19 is designed with multiple convolutional layers, allowing it to automatically learn hierarchical features. This facilitates transfer learning, enabling the model to benefit from knowledge gained in other domains and improving its performance with limited labelled emotion data. RNNs, however, have a limited memory capacity, which can hinder their ability to effectively capture and retain information over extended periods. Ensuring convergence and preventing issues such as overfitting or underfitting requires careful parameter tuning and regularization. Table II. Represent the performance metrics in the FER2013.

TABLE. II. The performance of detection using FER2013 dataset

Method	Accuracy	Precision	Recall	F1-score
ANN	0.93	0.76	0.94	0.82
ResNet	0.94	0.77	0.95	0.83
KNN	0.95	0.78	0.96	0.84
RCNN	0.96	0.79	0.97	0.85
CNN and RNN	0.97	0.80	0.99	0.86

B. Comparative Analysis

This method generates suitable parameter values for each test image and trained data. It utilizes labelled data, employing performance metrics such as Accuracy, Precision, Recall, and F1-score, as presented in Table III. The evaluation of classified eye tracking is compared with existing results from references [7], [8], and [9]. Eye motions are complex and can manifest in various ways. CNNs' ability to adapt and learn complex patterns makes them well-suited for recognizing diverse emotional expressions in the eyes, facilitating a more nuanced understanding of human emotion. While CNNs are proficient at learning features automatically, interpreting the learned features can be challenging. On the FER2013 dataset, it achieves an accuracy of 0.97, Precision of 0.80, Recall of 0.99, and F1-Score of 0.86. RNNs can be more challenging than training feedforward networks. Ensuring convergence

and preventing issues such as overfitting or underfitting requires careful parameter tuning and regularization.

TABLE. III. THE COMPARATIVE ANALYSIS OF PROPOSED METHOD USING FER2013 DATASET.

Author	Method	Accuracy	Precision	Recall	F1-Score
Yusra Khalid Bhatti et al. [7]	ResNet	0.82	0.79	0.87	0.85
Akhilesh Kumar et al. [8]	KNN	0.33	0.45	0.34	0.42
M.Kalpana Chowdary et al. [9]	VGG19	0.96	0.84	0.98	0.83
Proposed CNN & RNN Method		0.97	0.86	0.99	0.86

C. Discussion

In this section, limitations are discussed, and the proposed method involves training Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). CNNs are designed to capture spatial hierarchies in data, particularly beneficial for eye emotion tracking, as emotions are often expressed through subtle spatial patterns in the eyes. CNNs can discern these intricate patterns, enhancing the precision of emotion recognition. However, interpreting the learned features of CNNs can be challenging is sensitive to outliers in the dataset, and these outliers can disproportionately influence the classification results. This sensitivity may lead to inaccurate emotion tracking if outliers are present in the eye data. The storage requirements for VGG19 models are substantial, which may be a constraint in resource-limited environments, especially on devices with limited memory or bandwidth. Training RNNs can be more challenging than training feedforward networks, requiring careful consideration to ensure convergence and prevent issues such as overfitting.

VII. CONCLUSION

In this paper, a real-time algorithm is proposed for accurate centre of eye localization, in low-resolution images. The proposed methodology Convolutional Neural Networks (CNNs) at automatically learning hierarchical features from input data. In the context of eye emotion tracking, CNNs can effectively learn and extract discriminative features from raw eye images, improving the accuracy of emotion recognition. While CNNs are proficient at learning features automatically, interpreting the learned features can be challenging. RNNs are prone to vanishing and exploding gradient problems, particularly in long sequences. This can impact the ability of the network to capture dependencies in lengthy eye movement patterns. Experimental results demonstrate a detection accuracy of 0.97 surpassing other existing models such as KNN, ResNet and VGG19 on the FER2013 dataset. In the future, to enhance the CNN approach, we can improve it to completely remove all under-range noise in the generated eye part. Additionally, we can modify it to account for wrinkles, face tilts, and poses while identifying the state of closed eyes and blinking. The proposed algorithm's low computational complexity enables combining another noise reduction method and controlling overfitting with the specular reflection reduction method to improve efficiency, especially in the presence of strong specular highlights.

REFERENCES

[1] Ou, W.L., Kuo, T.L., Chang, C.C. and Fan, C.P., 2021. Deep-learning-based pupil center detection and tracking technology for visible-light wearable gaze tracking devices. Applied Sciences, 11(2), p.851.

[2] Nayak, S., Nagesh, B., Routray, A. and Sarma, M., 2021. A Human–Computer Interaction framework for emotion recognition through time-series thermal video sequences. Computers & Electrical Engineering, 93, p.107280.

[3] Valtakari, N.V., Hooge, I.T., Viktorsson, C., Nyström, P., Falck-Ytter, T. and Hessels, R.S., 2021. Eye tracking in human interaction: Possibilities and limitations. Behavior Research Methods, pp.1-17.

[4] Xun, Z., Baoqing, H., Dian, L., Jingyuan, W., Chenchen, Y., Yu, W., Qiong, M., Henggang, X. and Hongxiang, K., 2023. Eye behavior recognition of eye–computer interaction. Multimedia Tools and Applications, pp.1-17.

[5] Manjanaik, N., B. D. Parameshachari, S. N. Hanumanthappa, and Reshma Banu. "Intra Frame Coding In Advanced Video Coding Standard (H. 264) to Obtain Consistent PSNR and Reduce Bit Rate for Diagonal Down Left Mode Using Gaussian Pulse." In IOP Conference Series: Materials Science and Engineering, vol. 225, no. 1, p. 012209. IOP Publishing, 2017.

[6] Nguyen, D.L., Putro, M.D. and Jo, K.H., 2021. Eye state recognizer using light-weight architecture for drowsiness warning. In Intelligent Information and Database Systems: 13th Asian Conference, ACIIDS 2021, Phuket, Thailand, April 7–10, 2021, Proceedings 13 (pp. 518-530). Springer International Publishing.

[7] Bhatti, Y.K., Jamil, A., Nida, N., Yousaf, M.H., Viriri, S. and Velastin, S.A., 2021. Facial expression recognition of instructor using deep features and extreme learning machine. Computational Intelligence and Neuroscience, 2021, pp.1-17.

[8] Kumar, A. and Kumar, A., 2021, December. Analysis of Machine Learning Algorithms for Facial Expression Recognition. In International Conference on Advanced Network Technologies and Intelligent Computing (pp. 730-750). Cham: Springer International Publishing.

[9] Chowdary, M.K., Nguyen, T.N. and Hemanth, D.J., 2021. Deep learning-based facial emotion recognition for human–computer interaction applications. Neural Computing and Applications, pp.1-18.

[10] Zhao, L., Wang, Z., Zhang, G., Qi, Y. and Wang, X., 2018. Eye state recognition based on deep integrated neural network and transfer learning. Multimedia Tools and Applications, 77, pp.19415-19438.

[11] Kaviya, P. and Arumugaprakash, T., 2020, June. Group facial emotion analysis system using convolutional neural network. In 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184) (pp. 643-647). IEEE.

[12] Dataset: https://www.kaggle.com/datasets/yusufkorayhasdemir/fer2013csv (access on November)

[13] Kiran, P., and B. D. Parameshachari. "Resource optimized selective image encryption of medical images using multiple chaotic systems." Microprocessors and Microsystems 91 (2022): 104546.

[14] Minhas, A.A., Jabbar, S., Farhan, M. and Najam ul Islam, M., 2022. A smart analysis of driver fatigue and drowsiness detection using convolutional neural networks. Multimedia Tools and Applications, 81(19), pp.26969-26986.

[15] Zhang, C., Kim, Y.K. and Eskandarian, A., 2021. EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification. Journal of Neural Engineering, 18(4), p.046014.

[16] Chakraborty, P., Ahmed, S., Yousuf, M.A., Azad, A., Alyami, S.A. and Moni, M.A., 2021. A human-robot interaction system calculating visual focus of human's attention level. IEEE Access, 9, pp.93409-93421.