Article

Deep Learning-Based Eye Blink Detection in Video Frames: Performance Analysis of Various Models

Fahad Ata , Kubilay Ayturan , Fırat Hardalaç , Uğurhan Kutbay *

Posted Date: 8 August 2024

doi: 10.20944/preprints202408.0597.v1

Keywords: eye blink detection; deep learning; human computer interaction (HCI); image recognition

Article

Deep Learning-Based Eye Blink Detection in Video Frames: Performance Analysis of Various Models

Fahad Ata, Kubilay Ayturan, Fırat Hardalaç and Uğurhan Kutbay *

Department of Electrical and Electronics Engineering, Gazi University, Ankara 06570, Turkey
* Correspondence: ukutbay@gazi.edu.tr (U.K.)

Abstract: The human eye is a vital source of collecting information worldwide via daily observation. The computer science area recognizes it in research as a human-computer interaction. Human blink is an integral part of this observation as a research topic. Several studies have been conducted on eye blinking and left-right eye movements. Previous studies conducted on the eye blink consider the availability of hardware devices high in the budget, high Precision, and low light exposure, unlike our study, which uses low-budget devices as simple as a webcam and more versatile usable techniques, incredibly both dim and high light to name a few. This study offers the observational approach via ResNet101v2, VGG-19, and Convolution Neural Network (CNN) architectures. Outperforming both VGG-19 and CNN in this study, ResNet101v2 achieves an impressive accuracy of 98.2%. In contemporary times, deep learning in AI is taking an advanced form. This study attempts to provide a new insight into the real-world implementation of the eye blink.

Keywords: eye blink detection; deep learning; human computer interaction (HCI); image recognition

1. Introduction

The eye plays a vital role in computer science, like human-computer interaction. One of the helpful nonverbal behavioral clues for resolving various issues is the blinking of the eyes. Several eye-blinking applications exist, such as identifying driver fatigue and typing applications [1]. Most of the data we gather about the world is obtained through our eyes. Thus, the most vital sense in our system is vision. Several image processing and computer vision modes detect and monitor eye movements, eye tracking, and blinking [2]. Eye tracking offers findings that are being used more and more frequently in neurology, psychology, and ophthalmology to diagnose patients [3]. That is why it is essential to have a system that will enable precise eye tracking and provide information about the eye's direction and the motions that resulted in creating that path. The process of locating an eye in a video frame and using that information to determine the direction of the look is called eye Movement. It is essential in research and development, like product design and cognitive science. An eye-tracking system combines several devices with networked software to measure eye movements, counters, and positions, linking the results to the same eye in a series of images taken one after the other over time. Due to their versatility, eye trackers are becoming increasingly common in almost every industry [4]. Machine learning methods are employed to enhance eye-movement tracking techniques, Precision, and durability in estimating the position of the cornea in the eye and pupil's center. A gaze estimating algorithm independent of the subject's and cell phones' relative movements then uses the qualities [5]. The adaptive real-time eye blink detection system offers a non-intrusive and cost-effective solution for detecting eye blinks only with a webcam. However, there are several factors on which the methods will not rely, like sensitivity to the environment, user-specific challenges, real-time processing requirements, and also limited robustness. Even though many experiments on gaze engagement have been published, the efficiency of these approaches in intelligent interaction under real-world conditions still needs to improve [6,7]. This paper presents an approach to eye blink detection by applying RestNet101V2, VGG-19, and convolution Neural

Network(CNN) Architectures. The CNN architecture is known for its process in image-related tasks in which convolutional layers are used for feature extraction.VGG-19, known for its flexibility, and RestNet101V2, incorporating residual blocks to address vanishing gradient issues, offer alternative approaches for eye Blink detection. The training, validation, and testing results provide insight into each architecture to select the most suitable architecture for eye blink detection. This paper is organized as follows: section 2 covers literature Review. Section 3 describes the methodology used in this paper. The section contains results and discussion. The section presents a conclusion.

2. Literature Review

This section explores all the eye blink detection systems and their applications [8]. They developed a system to identify the eyes and face for driver tracking. The recurrent neural network-based system is presented in this approach. That network is trained using a synthetic event-based dataset called Neuromorphic-HELEN, which is simulated and has precise bounding box annotations. Furthermore, a technique is suggested to identify and evaluate the eye blinks of drivers by utilizing the excessive temporal resolution of event cameras. Blinking behavior gives more information about how tired or sleepy a driver is. The authors claim they have shown promising results. However, this system needs to be improved when drivers wear glasses [9]—proposed an eye blink detection system based on a camera to identify the driver's drowsiness during drawing. In this approach, the Dataset is collected from driving simulator experiments with the help of a remote camera. The eyelids' velocity is calculated using a threshold determined by a k-means clustering algorithm. According to the author, his system shows promising results regarding waking and drowsiness. However, this system is not reliable when the driver wears the glasses. Similarly, [10] developed a volunteer eye blinking detection system, and real-time data is handled with the help of an intelligent computer vision detector. The convolutional neural network and support vector machine are trained on different datasets, and performance is evaluated using different datasets such as ZJU, Cew, and ABD datasets. The 97% accuracy and 92% F1 score are obtained on the eye blink and ABD dataset. [11] introduce a system based on a convolutional neural network for eye blink detection. That system might be utilized to identify the driver's drowsiness level. From the results, it can be concluded that the system's performance is outstanding using the threshold algorithm. However, there is a problem in detection when the eyes are moved up and down [12], and a motion vector analysis-based eye blink detection system has been proposed. Such motion vectors are obtained with the help of the Gunnar–Farneback tracker and analyzed by utilizing the state machine. The motion vector in the proposed approach is normalized with the help of intraocular distance. The motion vectors are normalized to gain the invariance in the eye region size. In this suggested approach [13], a metaverse system was proposed whose purpose was to add any wordless data. The middle step of this approach was to perform a task in which the data about closed and open eyes was carried out with the help of MediaPipe. The results obtained from the extraction process showed a remarkable performance. The results showed an increase of 86.70% in the number of proper identifications in the circumstances of wearing glasses. The results obtained were better than those of the previous studies within the same domain. Moreover, the data collected through the outcomes proved enough for the virtual character to move. However, due to the limited number of experimental scenarios, verifying whether the system can be utilized for general purposes is essential. In addition to the limited experiment situation, integrating the proposed approach with a Metaverse system is essential. This integration will allow the researchers to test the actual functionality of the suggested technique.

Furthermore, the study suggested by [14] shed light on a live method for recognizing ASL gestures. ASL, American Sign Language, utilizes advanced machine learning and computer vision techniques to recognize tasks. The described method utilizes the Convolutional Neural Network and Mediapipe library, which are used for ASL sorting gestures and feature extraction. When this proposed method was tested, it showed remarkable accuracy. The results showed an accuracy of 99.95% in identifying all ASL alphabets. This remarkable accuracy proved that the study presented can be utilized in communication devices by people with different hearing problems. The suggested method can also be utilized for other sign languages, including hand gestures. This approach could improve the lives of individuals who have difficulty hearing. The presented research is about Mediapipe and CNN and their effective utilization in recognizing live sign language, thus resulting in advancements in machine learning and computer vision. The methodology described by [15] presents eye-tracking techniques. The presented study is conducted using a business-orientated webcam, which is used to identify the iris of the eye. A component of eye tracking study is detecting the iris's center. The suggested study utilizes the technique of observing the center of the iris under visible light instead of utilizing any particular eye-tracking device. Later on, observing the iris's center under the influence of visible light is utilized as input for a consumer-grade webcam for eye tracking, which is inexpensive compared to other webcams. With the use of the suggested eye-tracking technique approach, the movements of the eye are easily trackable. The presented eye monitoring techniques provide us with information about our visual patterns, like counting the blinks of the eyes, defining neglect, and the way a pupil of the eye responds to various surprises. The presented approach discussed eye monitoring technologies. Although eye monitoring is standard, its application and investigation can be comprehensive and complicated. The suggested research of [16] discussed a basic app created based on the Mediapipe framework. The manuals presented in this study are also referred to as technical instruction documents or user guides for particular systems designed to help individuals. The purpose of manuals is to reduce individuals' discomfort by assisting them with the technical issues they experience in their usage. The stated discomfort is reduced by providing the user manual that includes step-by-step instructions for managing a specific system that helps the user recognize, comprehend, and resolve those technical issues. During the experiment, a live picture was captured using Kinect. Later, a range of hand gesture data was trained, and every hand motion was recognized. After that, the identified hand motion data was delivered to the system of the app's user manual. The individual can store the data in the application's user manual according to the identified hand motions. To make the application's user manual more convenient to utilize and transform the manual from a manual to a user-friendly one, the presented study suggested including hand motion detection utilizing MediaPipe in the suggested app. Similarly, the approach presented by [17] describes a novel approach to correctly identifying eye blinks. The advantage of this study is that the individual does not require any background limitations, nor is the user required to wear any mark or sensor. Moreover, manual startup is not necessary for the suggested method. Also, the described approach works perfectly fine with offline and online scenarios. The proposed method takes the whole video and automatically checks the eye as open or closed. In addition, the suggested method is also assessed on the individuals who wear glasses. The results obtained from this assessment demonstrated the usefulness of the approach. The suggested solution is straightforward to set up and utilize. It just needs one inexpensive webcam and PC, which is entirely gentle. In this paper, a robust methodology and benchmark dataset are presented, described by [18], for the placement of significant reference points, which helps improve the accuracy of iris extraction. The Dataset of iris reference points was utilized to carry out numerous learning sessions for the VGG19, ResNet50, VGG16, and MobileNetV2. Meanwhile, ImageNet weights are utilized for model initialization. Measurements of the model loss, model capacity, and Mean Absolute Error (MAE) are taken to assess and verify the suggested model. The outcomes obtained from the measurements for specific factors show that the suggested model operates better than alternative approaches. If the comparison is taken out for the suggested method towards the other models, the suggested method shows the MAE values of 0.33 for ResNet50, 0.60 for MobileNetV2, 0.34 for VGG19, and 0.35 for VGG16. These models show an average reduced reaction time of 75% and a reduced size of 60%. The photos of the eye were gathered, and the suggested technique was utilized to label them. The Dataset described in this study is now openly accessible for academic use. This study's result is a more compact model that provides real-time detection. This suggested model provides the precise identification of iris reference points. In addition, the model is utilized to collect iris reference point annotations supplied to the model. In this suggested paper, [19] described a technique that is utilized for blink detection. The suggested technique utilizes 3D landmarks for identification. The suggested study will first discuss the previous studies of blink datasets. After discussing previous studies, the accuracy of the suggested work will be evaluated.

Then, the results achieved after applying the described study will be discussed. Furthermore, the presented methodology demonstrates the eye blink detection technique's applicability in the non-physical form of stress identification. This suggested methodology will facilitate reducing stress in the professional healthcare domain.

Table 1. Systematic Literature Review.

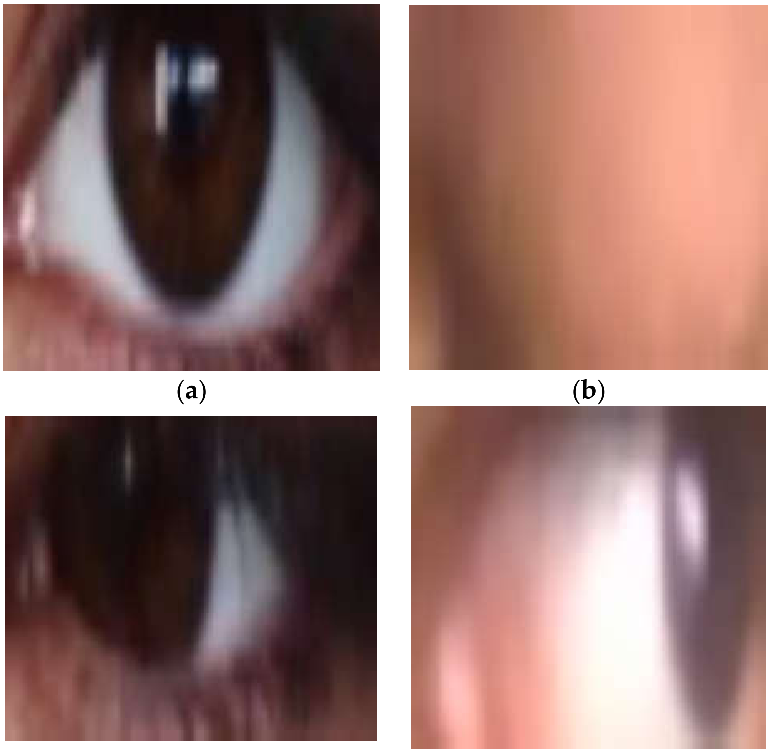
Author	Dataset	Techniques	Findings
[8]	Event-Based Dataset	Recurrent Neural Network	85% and 97% Precision and recall were achieved, respectively.
[9]	The Dataset is collected from simulator experiments	K-means Clustering	N/A
[10]	Autonomous Blink Dataset	Convolutional neural networks and SVM are utilized.	A 97% accuracy and 92% F1 score were obtained.
[11]	Training and testing datasets were collected while training.	The convolutional Neural Network model is utilized.	90% Accuracy is achieved.
[12]	Researcher's night Dataset is utilized	State Machine is utilized for vector analysis.	86% and 80% Precision and recall were obtained, respectively.
[13]	N/A	MediaPipe Technique is utilized	86% outcome is achieved
[14]	ASL Dataset is utilized	Convolutional Neural Network is utilized.	99% accuracy is achieved
[15]	N/A	MediaPipe and OpenCV technology are utilized	N/A
[16]	A custom dataset of hand gestures is created.	MediaPipe framework is utilized.	95% accuracy is achieved.
[17]	The custom dataset is utilized	The AdaBoost algorithm is utilized	98% accuracy is achieved
[18]	A custom dataset from videos is created.	ResNet50, MobileNet, Vgg16 and Vgg19 is utilized.	Maximum 90% accuracy is achieved.

3. Proposed Methodology

In this study, we propose a new detector for eye blink or eye movement detection. Eye blink detection applications are increasing daily due to their crucial role in detecting the driver's drowsiness while driving and for disabled people who can blink their eyes. Eye blink and eye movement detection are more challenging as compared to eye location. The proposed investigation provides a system that can detect the eye blink or eye movement (left or right). With eye blink detection, the proposed method can count the blink of eyes.

3.1. Data Gathering:

The first stage in gathering data for investigations in the suggested system is data gathering. The Dataset of images is extracted from the video [30]. The video is randomly selected from the internet. After extracting frames from the videos, the data is categorized into four kinds: open eye, closed eye, left eye, and right eye. Two thousand one hundred sixty-two images are used for training, 322 for testing, and 722 for validations.


(**a**) (**b**)

(**c**) (**d**)

Figure 1. Sample Images **(a)** Open Eye, **(b)** Closed Eye, **(c)** Left), **(d)** Right.

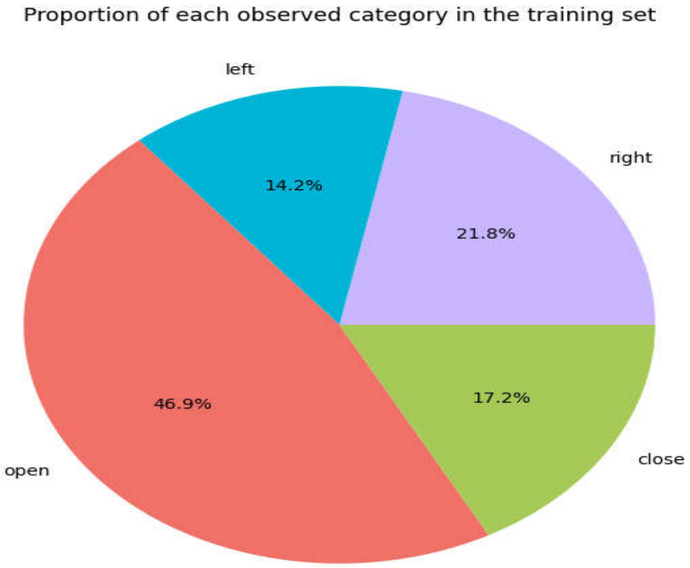


Figure 2. Dataset proportion for each category in the training set.

3.2. Image Preprocessing

Preprocessing is considered a procedure of transformation that is applied to data to make it worthwhile for various computer vision techniques. The preprocessing method transforms the unclean and noisy data into a pure data set. With the help of image processing techniques, the analysis can be performed efficiently by understanding the image's content. The primary purpose of image processing is to detect the interest points inside the image. These interest points are utilized for feature extraction. Preprocessing aims to optimize the image data by eliminating disruptive distortions and enhancing elements essential for additional processing. The proposed methodology applies the data preprocessing technique to the input images before giving the data to computer vision models. The image's noise is removed, and the image is enhanced for feature extraction.

3.3. Convolutional Neural Network

CNN has demonstrated impressive accomplishments. CNN, whose complete form is a convolutional Neural Network). Deep learning is regarded as one of the most significant neural networks. The convolutional neural network, based on computer vision, allows people to accomplish tasks previously thought to be unattainable, like intelligent medical care, face recognition, and autonomous vehicles [20,21]. A convolutional neural network is considered a feedforward neural network that can extract features from data with the help of convolution structures. The features were extracted manually in the traditional feature extraction procedure, but in convolutional neural networks, there is no need to extract the features manually. CNN's architecture takes inspiration from visual perception [22]. CNN kernels are unique receptors that are capable of reacting to multiple features. An artificial neuron is equivalent to a biological neuron. Activation functions resemble the process by which neuronal electric signals are only passed on to the following neuron when they are above a specific threshold. CNN is preferable to general artificial neural networks in many ways: every neuron is now only coupled to a limited selection of neurons from the preceding layer rather than all. That reduces parameters and improves overall convergence. Even more parameter reduction is possible when a set of interconnections carry identical weights. A pooling layer uses image local correlation to down-sample an image and minimize the quantity of data while maintaining meaningful information. Eliminating insignificant features also helps to lower the total range of parameters. Because of these three attractive features, CNN is becoming one of the foremost representative models in the deep learning industry. CNN has three layers [23]. The convolutional layer is the very beginning layer of the CNN. Meanwhile, the pooling layer comes in second place inside CNN. Then, moving forward comes a fully connected layer, the third layer of CNN. A CNN framework is created whenever those layers are combined.

3.3.1. Convolutional Layer

The main objective of the convolutional layer is to find the output of neurons linked with the input neuron's local region. This output is calculated by taking the scaler product among local regions of inputs and their weights. The function of ReLu is to employ the activation function like the sigmoid function on the output, which is generated by the preceding layer.

3.3.2. Pooling Layer

The pooling layer would subsequently conduct out-down sampling across the input's spatial dimensions, thus lowering the activation's parameter count.

3.3.3. Fully Connected Layers

The next layer of the CNN is the total layers. The work of these layers is to do the tasks that are present in conventional ANNs. In addition, these layers also try to generate class values via the activations. The generated class values will then be utilized for classification. Moreover, ReLu can be utilized among these layers to increase efficiency. By employing down sampling and convolution approaches, CNNs may provide class scores for regression and classification by transforming the initial input layer by layer via this straightforward transformation process.

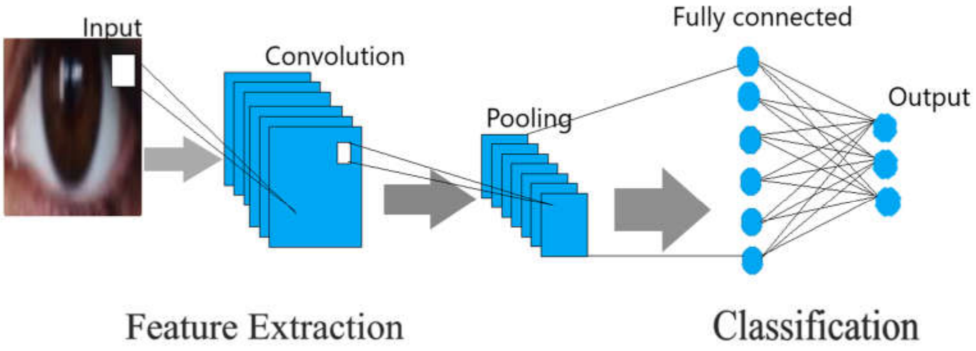


Figure 3. CNN Architecture.

3.4. VGG-19

The VGG network stands for visual geometry group, and it is also known as a deep neural network [24], which consists of multilayered operations. VGGNet uses the specific Dataset of ImageNet. The core base of VGGNet is the CNN model. The main advantage of VGG-19 is its flexibility. This network includes 3 × 3 convolutional layers positioned on top to rise with complexity level. This makes the use of VGG-19 beneficial. Max pooling layers were employed as a controller in VGG-19 to lower the volume capacity. Two fully connected layers employed Four thousand ninety-six neurons [24]. VGGNet DNN input data was derived from the vessel-segmented images. Convolutional layers were utilized for feature extraction during the training stage, and several of the convolutional layers' linked max pooling layers were employed to lower the dimensionality of the features. The features from the image are extracted with the help of the first convolutional layer, which consists of 64 kernels with a 3 x 3 filter size. The feature vector is generated with the help of fully connected layers. To reduce dimensionality and select features from the visual data to enhance classification outcomes, the obtained feature vector underwent further processing through SVD and PCA [25]. Decreasing the highly dimensional data utilizing PCA and SDA is a significant challenge. Compared to previous reduction approaches, PCA and SVD are quicker and more statistically stable, making them more valuable. In the end, the classification is performed with the help of the softmax activation function using 10-cross validation.

3.5. ResNet101V2

ResNet101V2 is considered a residual network and an extended version of the ResNet framework [26]. ResNet101v2 architecture includes high-performance multiple layers [27]. The primary distinction between version (V2) and version (V1) is that version (V2) applies batch normalization before every weight layer. ResNet is a convolutional neural network that can solve the vanishing gradient issue with the help of residual blocks [28]. Residual block is considered a critical point in ResNet101v2. Shortcut links in certain blocks let the network drop layers throughout backward and forward propagation. This architecture enables practical training of profound networks by avoiding the vanishing gradient issue. ResNet101v2 also includes enhanced activation methods and batch normalization, which help accelerate learning convergence. ResNet101v2 provides a reliable option for image identification jobs because of its careful design, which allows it to collect complex information inside images. The structure of ResNet101v2 was distinguished through its deep stack of residual blocks. Two 3x3 convolutional layers, encircled through ReLU and Batch Normalization activation operations, comprise every residual block [29]. The remaining links guarantee that gradients flow smoothly, which enables the structure to train efficiently, particularly as it gets deeper. There are four primary components in the design, and each one has several auxiliary blocks. ResNet101v2's ability to learn intricate hierarchical features using images is facilitated by its depth, making it an excellent tool for pattern recognition across various datasets.

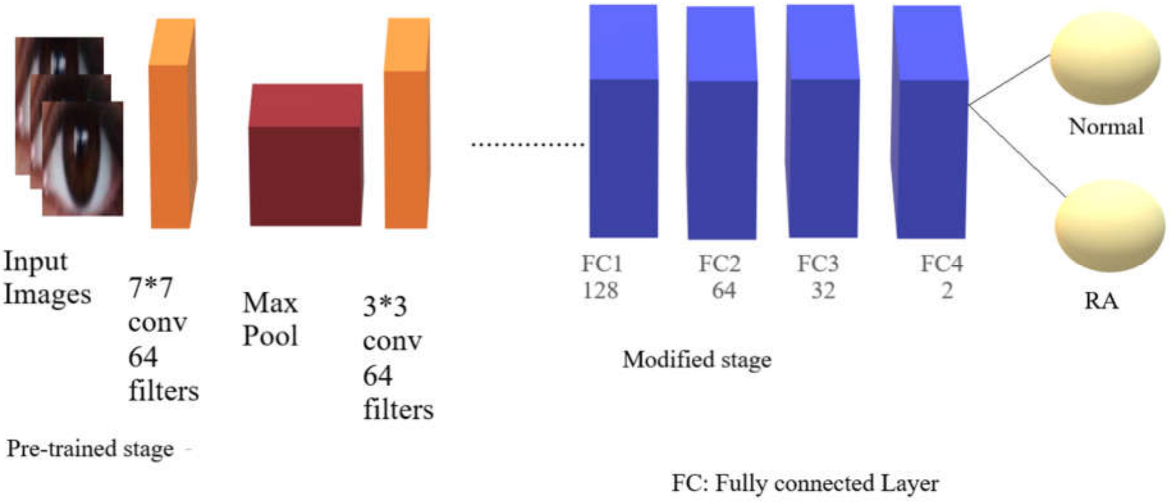


Figure 4. RESTNET101V2 Architecture.

4. Results

That section explains the experimental findings of the suggested methodology. The Dataset of images is extracted from the video to evaluate the performance of the proposed system [30]. The video is downloaded from the internet, and the frames of open eye, closed eye, left eye, and right eye are extracted from the video. The total number of frames is 3206, divided into training, testing, and validations.

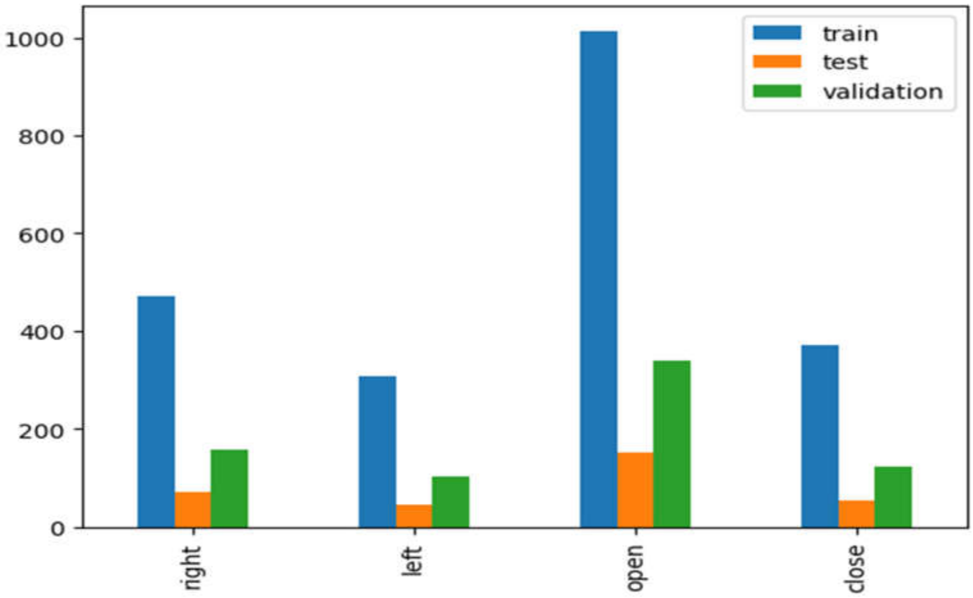


Figure 5. Dataset division.

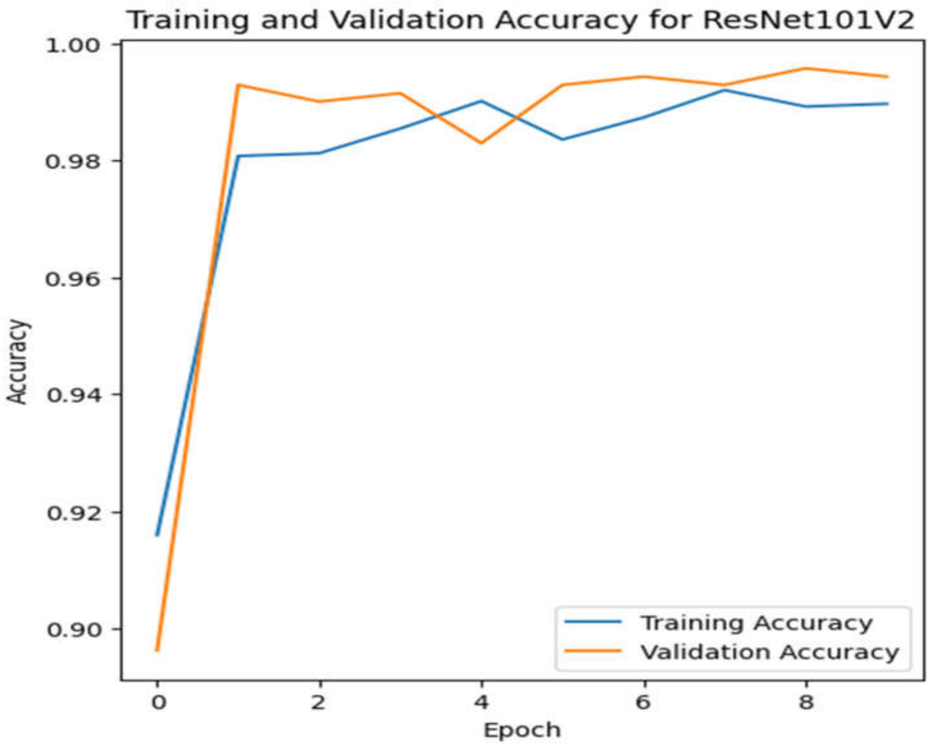


Figure 6. Validation as well as training accuracy for ResNet101v2.

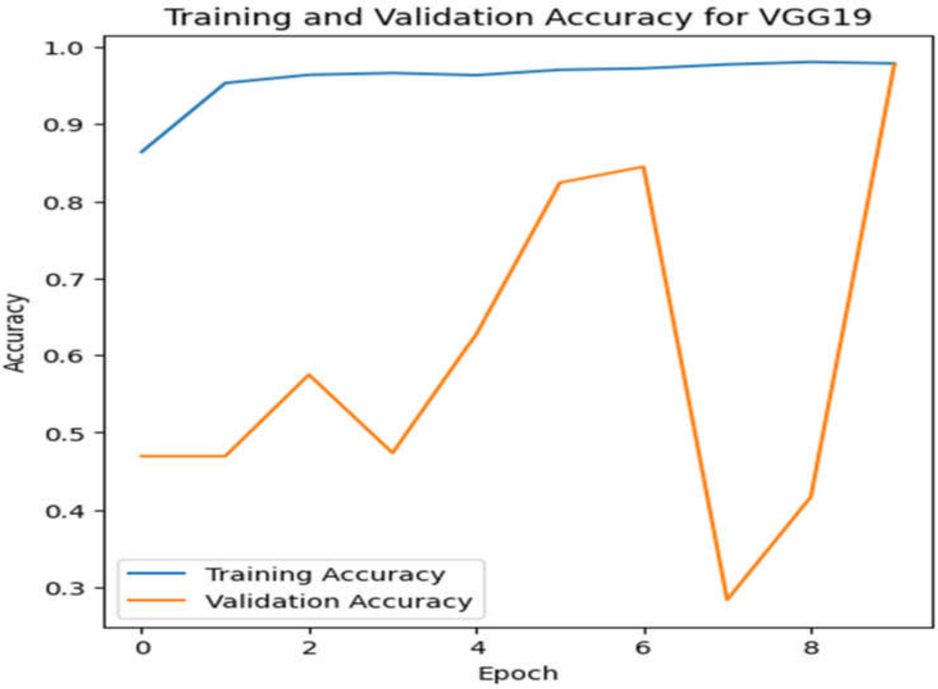


Figure 7. Validation as well as training accuracy for VGG19.

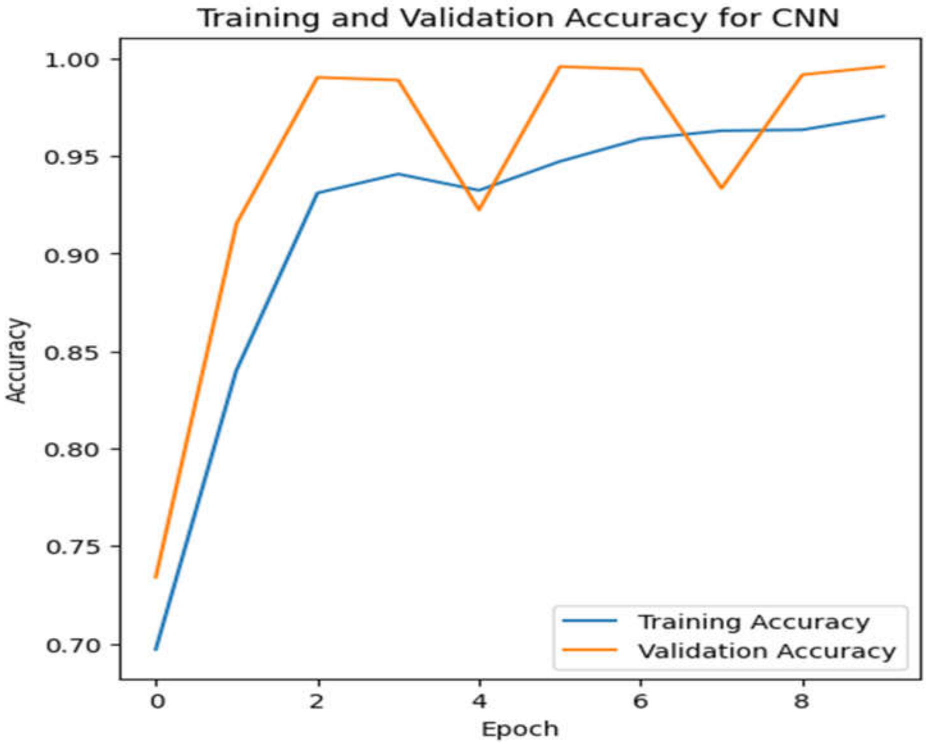


Figure 8. Validation as well as training accuracy for CNN.

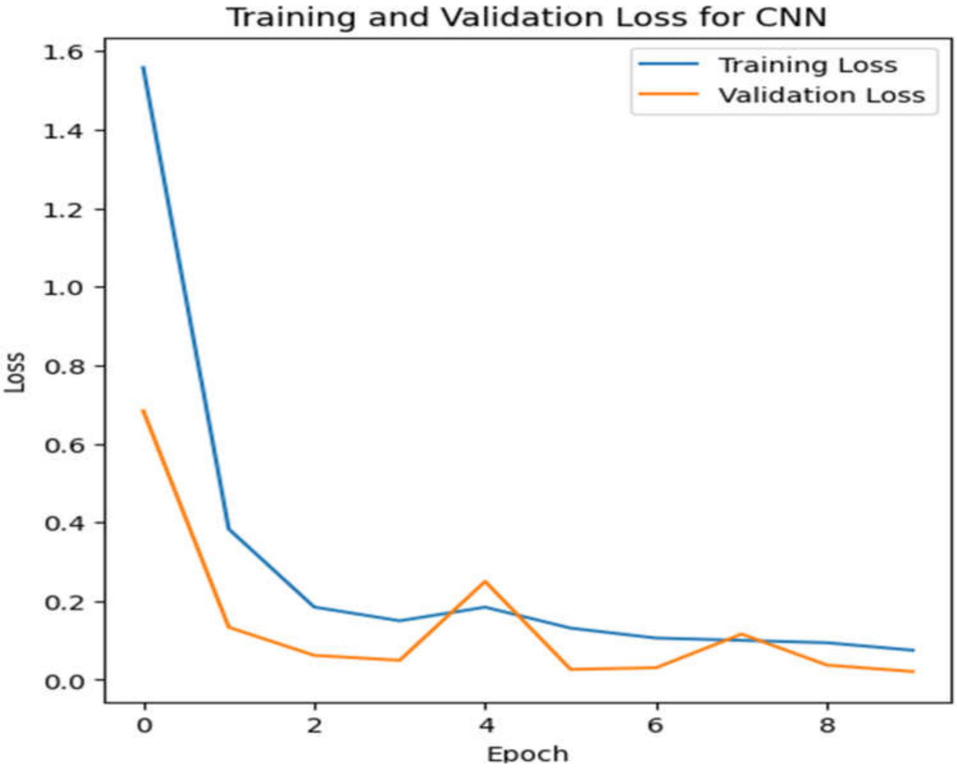


Figure 9. Validation as well as training loss for CNN.

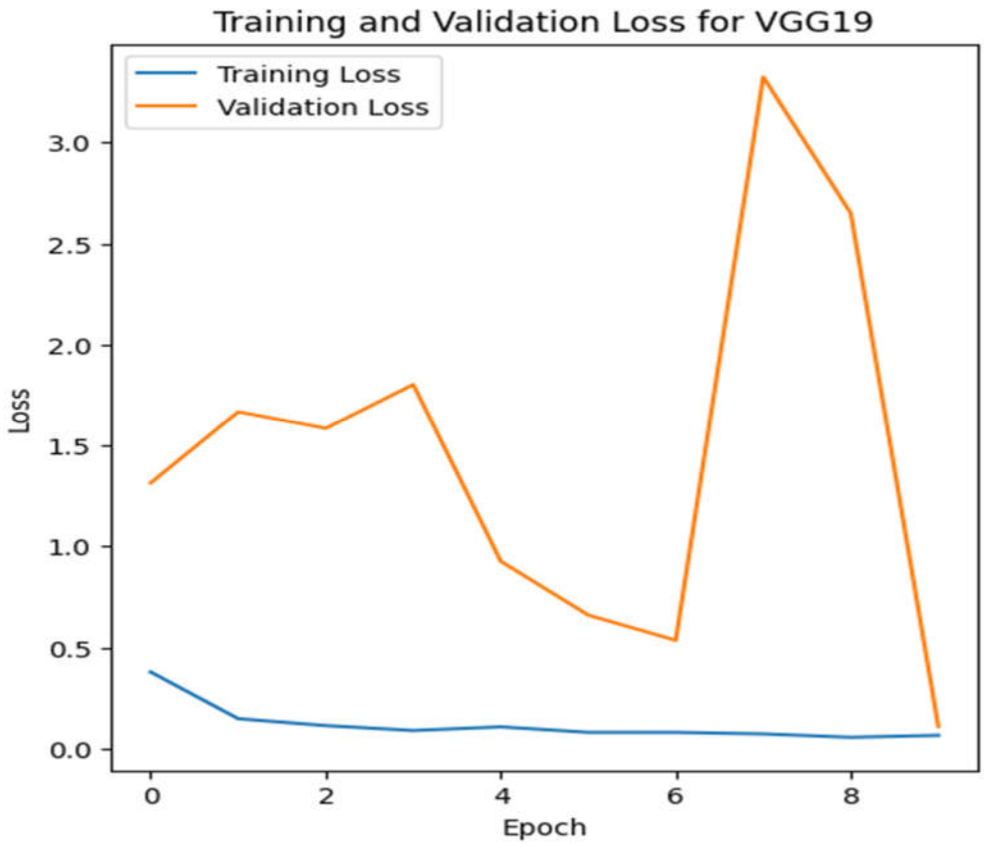


Figure 10. Validation as well as training loss for VGG19.

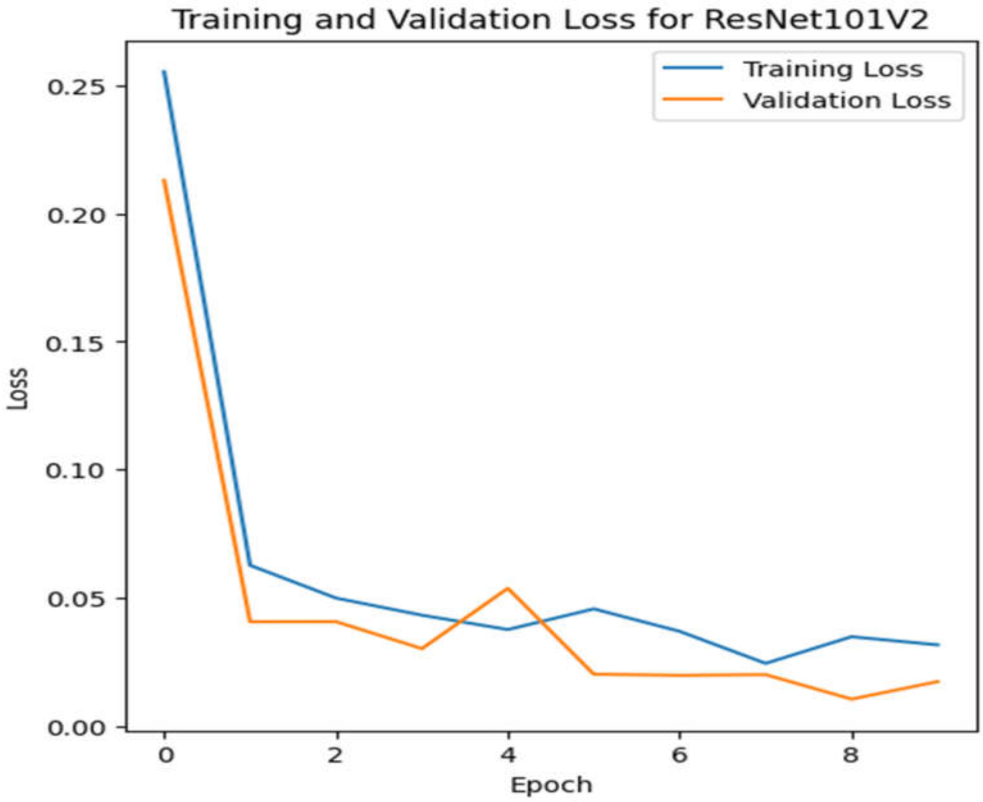


Figure 11. Validation as well as training loss for ResNet101v2.

4.1. Comparative Analysis

When comparing ResNet101v2, VGG16, and CNN, it can be concluded that ResNet101v2 gains a higher accuracy than VGG19 and CNN. The accuracy for ResNet101v2 is 98.2%, and the accuracy of VGG19 and CNN is 59% and 94%, respectively. The below Figure shows the accuracy of all the models.

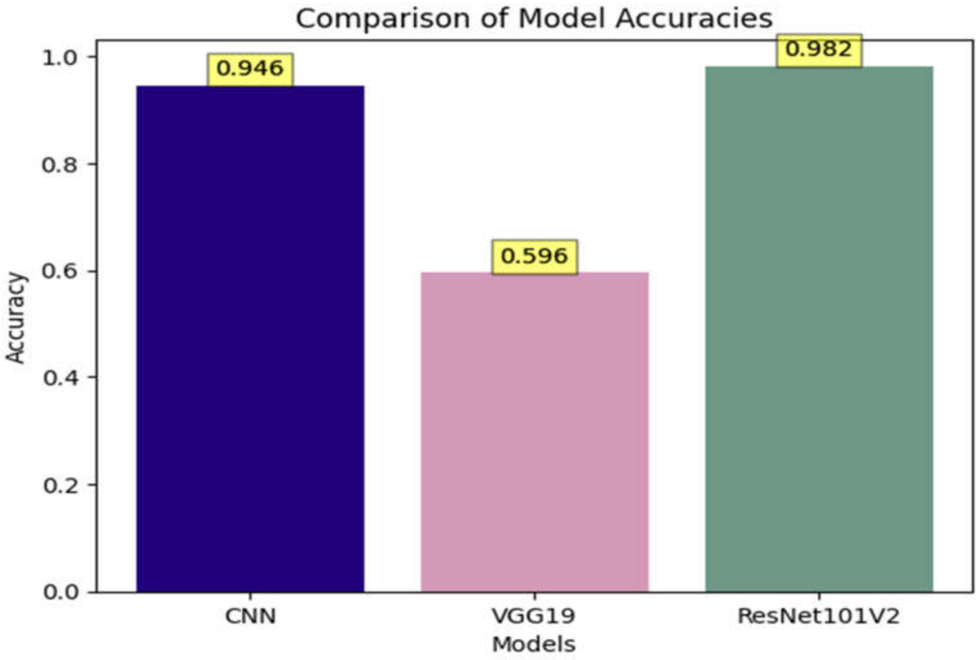
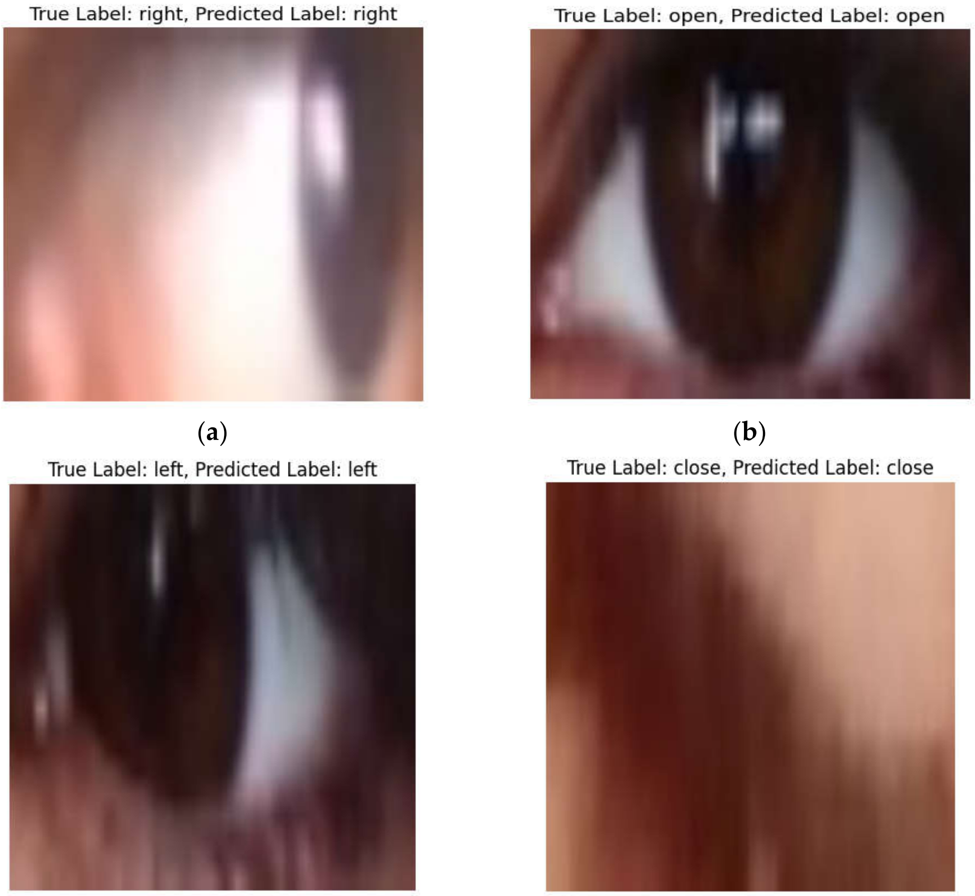


Figure 12. Comparison between CNN, VGG19 and ResNet101v2 accuracies.


True Label: right, Predicted Label: right
True Label: open, Predicted Label: open
True Label: left, Predicted Label: left
True Label: close, Predicted Label: close

(**a**) (**b**)

(**c**) (**d**)

Figure 13. RestNet101V2 Model Testing on dataset Result, **(a)** Right, **(b)** Open **(c)** Left), **(d)** Close.

4. Discussion and Conclusions

In this study three different models used for detecting eyes convolutional neural network(CNN), VGG-19, RestNet101V2 shows how these three models excel in detecting eyes blinks in video frames. Among the three models RestNet101V2 model overpowered bothVGG-10 and CNN, and achieved a remarkable accuracy of 98.2%. The major benefits of of using RestNet101V2 model is its strength which lies in its use of residual block, which helps in resolving the vanishing gradient problem- a common issue occur while training a deep neural networks. Using These features in the model helps in performing accurate and helps in growing the network into its deeper state. It makes it suitable for tasks like detecting eye blinks under various lightning conditions.

On the other hand, VGG-19 which is famous for its flexibility and higher accuracy in image classification, showed lower accuracy as compared to RestNet101V2. One of the reason for this lower accuracy in this study is because VGG-19 has simpler architecture which may not able to detect the nuances of eye blinks as effectively as RestNet101V2 residual connection. Also VGG-19 has large number of parameters which require more computational resources and can lead to over fitting if not properly managed. CNN model in the study shows the lowest accuracy as compared to other two models. It is because models like RestNet101V2 and VGG-19 have advantages of their advanced architecture like residual connections and deep layered for a task like eye blinks detection to get a higher accuracy.

This study also highlights an exciting aspect that the use of affordable equipment, such as simple webcam, to collect data. It makes the system more practical and accessible to the real world use, specifically in scenarios like where the cost-effective is major concern. The versatility of the system to work in various lighting conditions like low light, and bright light, making it suitable for working in various environment. Useful techniques like noise removal, and image enhancement plays a crucial role in pre-steps for preparing the data for effective feature extraction by models. These steps ensure that models can accurately detect the different categories open, closed, left, right eye state.

To conclude this paper, this article offers a comprehensive approach to eye blink detection utilizing the ResNet101v2, VGG-19, and Convolutional Neural Network (CNN) architectures. Results show how well these models perform on datasets taken from online motion pictures, thus classifying images into four categories: open, closed, left, and right eyes. With the help of canceling noise and enhancing features, the preprocessing step dramatically improves the models' performance and optimizes the input data for computer vision applications. Moreover, the CNN architecture provides a strong foundation for detecting eye blinks with its convolutional, pooling, and fully linked layers. Furthermore, ResNet101v2's residual blocks and VGG-19's adaptability clarify why both models perform better when analyzing complex images. Achieving quite an impressive accuracy of 98.2%, ResNet101v2 outperformed both VGG-19 and CNN in the comparative analysis. Despite its accuracy and convergence rate, ResNet101v2 can effectively handle problems like vanishing gradients, making it a reliable option for complex image recognition. The researcher is confident that there are practical ramifications, particularly for applications where precise eye blink detection is essential, such as human-computer interfaces or driver fatigue monitoring. Proper knowledge about the advantages and disadvantages will help the apprentices choose the best deep-learning architecture for their desired tasks, allowing them to make well-informed decisions. Furthermore, as computer vision develops, this work advances real-world implementations. It expands deep learning across various domains by offering insightful information about how CNN, VGG-19, and ResNet101v2 are used for eye blink detection.

Author Contributions: F.A.: methodology, software, data curation, validation, formal analysis, investigation, visualization, writing—original draft preparation, writing—review and editing; K.A.: conceptualization, validation, writing—review and editing, supervision, project administration; F.H.: validation, writing—review and editing, formal analysis. U.K.: writing—review and editing, formal analysis, supervision, project administration. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The authors used open access datasets available on the Internet at https://github.com/Fahadata002/Eye-Gaze (accessed on 22 July 2024)

Acknowledgments: The authors would like to thank the editors and anonymous reviewers for providing insightful suggestions and comments to improve the quality of research paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

3D	Three Dimensional
ANN	Artificial Neural Network
CEW	Closed Eyes in The Wild
CNN	Convolutional Neural Network
DNN	Deep Neural Network
PC	Personel Computer
PCA	Principal Component Analysis
ReLu	Rectified Linear Unit
SDA	Stacked Denoising Autoencoders
SVD	Singular Value Decomposition
VGG	Visual Geometry Group

References

1. Sugur, S. M., Hemadri, V. B., & Kulkarni, U. P. (2013). Drowsiness detection based on eye shape measurement. International Journal of Computer and Electronics Research, 2(2), 5.
2. Singh, M., & Kaur, G. (2012). Drowsy detection on eye blink duration using algorithm. International Journal of Emerging Technology and Advanced Engineering, 2(4), 363-365.
3. Van Hal, B. (2013). Real-Time stage 1 sleep detection and warning system using a low-cost EEG headset.
4. Sahayadhas, A., Sundaraj, K., & Murugappan, M. (2012). Detecting driver drowsiness based on sensors: a review. Sensors, 12(12), 16937-16953.
5. Ashtiani, B., & MacKenzie, I. S. (2010, March). BlinkWrite2: an improved text entry method using eye blinks. In Proceedings of the 2010 symposium on eye-tracking research & applications (pp. 339-345).
6. Barrett, M., Skovsgaard, H., & San Agustin, J. (2009, May). Performance evaluation of a low-cost gaze tracker for eye typing. In Proceedings of the 5th Conference on Communication by Gaze Interaction, COGAIN (Vol. 9, pp. 13-17).
7. Porta, M., Ravarelli, A., & Spagnoli, G. (2010, March). ceCursor, a contextual eye cursor for general pointing in windows environments. In Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (pp. 331-337).
8. Ryan, C., O'Sullivan, B., Elrasad, A., Cahill, A., Lemley, J., Kielty, P., ... & Perot, E. (2021). Real-time face & eye tracking and blink detection using event cameras. Neural Networks, 141, 87-97.
9. Baccour, M. H., Driewer, F., Kasneci, E., & Rosenstiel, W. (2019, June). Camera-based eye blink detection algorithm for assessing driver drowsiness. In 2019 IEEE Intelligent Vehicles Symposium (IV) (pp. 987-993). IEEE.
10. de Lima Medeiros, P. A., da Silva, G. V. S., dos Santos Fernandes, F. R., Sánchez-Gendriz, I., Lins, H. W. C., da Silva Barros, D. M., ... & de Medeiros Valentim, R. A. (2022). Efficient machine learning approach for volunteer eye-blink detection in real-time using webcam. Expert Systems with Applications, 188, 116073.
11. Jordan, A. A., Pegatoquet, A., Castagnetti, A., Raybaut, J., & Le Coz, P. (2020). Deep learning for eye blink detection implemented at the edge. IEEE Embedded Systems Letters, 13(3), 130-133.
12. Fogelton, A., & Benesova, W. (2016). Eye blink detection based on motion vectors analysis. Computer Vision and Image Understanding, 148, 23-33.
13. TODA, M., & TANAKA, H. (2023). Extraction of eye open/close state by mediapipe and its application to avatar expression. In International Symposium on Affective Science and Engineering ISASE2023 (pp. 1-4). Japan Society of Kansei Engineering.
14. Kumar, R., Bajpai, A., & Sinha, A. (2023). Mediapipe and cnns for real-time asl gesture recognition. arXiv preprint arXiv:2305.05296.
15. S. Doijad, A. Bhalerao, and S. Khan. (2022) "Impact Factor : 2.205 EYE IRISES TRACKING USING MEDIAPIPE AND OPENCV," Int. J. Progress. Res. Eng. Manag. Sci., vol. 02, pp. 319–322.

16. Harris, M., & Agoes, A. S. (2021, November). Applying hand gesture recognition for user guide application using MediaPipe. In 2nd International Seminar of Science and Applied Technology (ISSAT 2021) (pp. 101-108). Atlantis Press.
17. Galab, M. K., Abdalkader, H. M., & Zayed, H. H. (2014). Adaptive real time eye-blink detection system. International Journal of Computer Applications, 99(5), 29-36.
18. Adnan, M., Sardaraz, M., Tahir, M., Dar, M. N., Alduailij, M., & Alduailij, M. (2022). A robust framework for real-time iris landmarks detection using deep learning. Applied Sciences, 12(11), 5700.
19. Kraft, D., Hartmann, F., & Bieber, G. (2022, September). Camera-based blink detection using 3d-landmarks. In Proceedings of the 7th International Workshop on Sensor-Based Activity Recognition and Artificial Intelligence (pp. 1-7).
20. Sun, Y., Xue, B., Zhang, M., & Yen, G. G. (2019). Completely automated CNN architecture design based on blocks. IEEE transactions on neural networks and learning systems, 31(4), 1242-1254.
21. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, 8, 1-74.
22. Aydogdu, M. F., & Demirci, M. F. (2017, May). Age classification using an optimized CNN architecture. In Proceedings of the International Conference on Compute and Data Analysis (pp. 233-239).
23. Chavda, A., Dsouza, J., Badgujar, S., & Damani, A. (2021, April). Multi-stage CNN architecture for face mask detection. In 2021 6th International Conference for Convergence in Technology (i2ct) (pp. 1-8). IEEE.
24. Mascarenhas, S., & Agarwal, M. (2021, November). A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification. In 2021 International conference on disruptive technologies for multi-disciplinary research and applications (CENTCON) (Vol. 1, pp. 96-99). IEEE.
25. Sinha, S., & Le, D. V. (2021, November). Completely automated CNN architecture design based on VGG blocks for fingerprinting localisation. In 2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN) (pp. 1-8). IEEE.
26. Adar, T., Delice, E. K., & Delice, O. (2022, October). Detection of COVID-19 from a new dataset using MobileNetV2 and ResNet101V2 architectures. In 2022 Medical Technologies Congress (TIPTEKNO) (pp. 1-4). IEEE.
27. Ali, S., & Agrawal, J. (2024). Automated segmentation of brain tumour images using deep learning-based model VGG19 and ResNet 101. Multimedia Tools and Applications, 83(11), 33351-33370.
28. Asswin, C. R., KS, D. K., Dora, A., Ravi, V., Sowmya, V., Gopalakrishnan, E. A., & Soman, K. P. (2023). Transfer learning approach for pediatric pneumonia diagnosis using channel attention deep CNN architectures. Engineering Applications of Artificial Intelligence, 123, 106416.
29. Krishnamurthy, S., Srinivasan, K., Qaisar, S. M., Vincent, P. D. R., & Chang, C. Y. (2021). [Retracted] Evaluating Deep Neural Network Architectures with Transfer Learning for Pneumonitis Diagnosis. Computational and Mathematical Methods in Medicine, 2021(1), 8036304.
30. Eye-Gaze. Available online: https://github.com/Fahadata002/Eye-Gaze (accessed on 22 July 2024)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.