[11]Ranjeet Bidwe
[2]Gouransh Agrawal
[3]Unnati
[4]Akshay Sangwan
[5]Himanshu Kulhari
[6]Sashikala Mishra
[7]Simi Bajaj

Eye Gaze for Monitoring Attention Through Hybrid Ensemble Learning

Abstract: - One of the countless tasks that call attention to monitoring is necessary for comprising healthcare, education, transportation safety, and human-computer interaction. This research describes novel work done in attention monitoring by fusing a hybrid eye gaze model with deep learning to monitor a driver's attention level. The hybrid eye gaze model proposed is described and its results are produced in this paper. The proposed model uses an augmented dataset where data augmentation techniques like rotation, shifting, shearing, and flipping are applied together with adjustments like changing the fill mode in terms of zooming into the image and rescaling. These are all crucial aspects in reliable and consistent training of the model. Our model is built on modern pre-trained architectures which include VGG16, VGG19, InceptionV3, EfficientNetB0, EfficientNetB7, and InceptionResNetV2. To aid in capturing very minute attention dynamics, we modify these architectures and then incorporate more layers. Later, we used a model ensemble to increase the accuracy and efficiency of the model. Later, the XGBoost model is integrated with all other models used before in the hybrid model technique to obtain better accuracy and efficiency of the model. The model performance is adequately evaluated using various evaluation measures like accuracy, precision, recall, F1 Score, and support. These metrics provide a holistic understanding of the model's capability to detect and predict attention patterns in different contexts. After using the models, we could get the best accuracy from VGG19 and InceptionResNetV2, i.e., 84.6% and 83.6% respectively. VGG16 hybrid models recorded 82% in the accuracy test. With deep learning and pre-trained architectures, the Hybrid Eye Gaze Model shows a strong and flexible attention monitoring solution for varying types of applications.

Keywords: Attention Monitoring, Data Augmentation, VGG16, VGG19, InceptionV3, EfficientNetB0, EfficientNetB7, InceptionResNetV2, XGBoost.

I. INTRODUCTION

In today's fast-paced digital landscape, the capability to sense and track human attention is of foremost importance for a diverse set of applications and industries. These include applications as diverse as, improved end-user experiences, improved healthcare diagnostics for generating optimum changes in instructions, and safety in high-stakes environments these technologies the autonomous driving that depends on attention monitoring. This has led to the development of some sophisticated and highly advanced tools specifically designed for sensing and analyzing the delicate complexities of human attention.

The approach that stands out among the rest as one of the most successful and promising techniques in attention detection is eye gaze tracking. Tracking what someone is looking at can tell a lot about a person’s priorities, interests, and places of interest in a given scenario. Applying this information will assist in generating more user-friendly

[1]Symbiosis Institute of Technology, Pune, Symbiosis International (Deemed University), Lavale, Pune, Maharashtra, India
[2]Symbiosis Institute of Technology, Pune, Symbiosis International (Deemed University), Lavale, Pune, Maharashtra, India
[3]Symbiosis Institute of Technology, Pune, Symbiosis International (Deemed University), Lavale, Pune, Maharashtra, India
[4]Symbiosis Institute of Technology, Pune, Symbiosis International (Deemed University), Lavale, Pune, Maharashtra, India
[5]Symbiosis Institute of Technology, Pune, Symbiosis International (Deemed University), Lavale, Pune, Maharashtra, India
[6]Symbiosis Institute of Technology, Pune, Symbiosis International (Deemed University), Lavale, Pune, Maharashtra, India
[7]Director of Academic Program & Deputy Associate Dean International Southeast Asia at Western Sydney University
ranjeetbidwe@hotmail.com, gouransh12345@gmail.com, unnatijha2001@gmail.com,
akshaysangwan8571@gmail.com, himanshukulhari28@gmail.com, sashikala.mishra@sitpune.edu.in,
k.bajaj@westernsydney.edu.au
Corresponding Author: ranjeetbidwe@hotmail.com

interfaces, detecting tired drivers, and personalized learning initiatives among many others. Conversely, real-world eye gaze tracking techniques may have their shortfalls, which may range from being sensitive to the environment, and calibration problems, as well as limited applicability under normal operating conditions.

The research proposes a new solution to these shortcomings by of the "Hybrid Eye Gaze Model for Attention Monitoring." The best properties of multiple eye-tracking strategies were merged in this model as also data augmentation was included to augment the robustness of the model it merges modern deep learning architectures for responsive and adaptable attention monitoring. The methodology uses pre-trained models like VGG16, VGG19, InceptionV3, EfficientNetB0, EfficientNetB7, and InceptionResNetV2. Extra layers are added to the proposed methodology to capture fine-grained attention patterns.

Different evaluation metrics such as F1-Score are used to analyze in detail aspects like accuracy, precision, recall, support, and performance of the hybrid eye gaze model. The evaluations shall present the abilities of the model and how they can redefine attention monitoring by offering an interpretable way flexible, dependable, and scalable way of improving human attention in different contexts.

II. Literature Review

This paper inferred significant hypotheses, evidence found, and evident gaps based on which Table 1 is drawn up from the systematic review of scholarly literature about the topic. These surveyed methods helped in motivating and implementing the proposed system in this paper.

Table 1. Literature Review

References	Year	Algorithm Used	Summary
[1]	2023	Driver action recognition (DAR).	This paper proposes a novel hard attention network to tackle the immediate problem of distracted driving through driver action recognition in real-world driving scenes. Since the hard attention mechanism focuses on core data linked to the drivers, and leaves the irrelevant data, therefore it improves accuracy in identifying safe driving and distraction. The results are exciting: the accuracy rate of safe driving recognition could reach up to 95.83%, while that of distraction detection could be as high as 99.07%. Meanwhile, it outperforms soft attention-based models in terms of computing efficiency. This work has significant importance for increasing traffic safety.
[2]	2023	YOLOv5	The proposed approach in this study is intended to monitor students' attendance, attentiveness, and mood in a real-time classroom with the use of high-definition cameras, interfaces easy to comprehend, and behavior detection and face recognition based on YOLOv5.
[3]	2023	CNNs Residual Learning (ResNet) RNN	One of the key challenges in this work relates to the issue of abnormal data identification within monitoring data of structural measures, which has a major impact on condition assessment. For the categorization task, the researchers recommended the use of a Residual Attention Network (RAN) with residual learning and attention mechanisms to improve speed and accuracy. In their procedure, the researchers converted hourly segmented data into matrix form through mutual information correlation analysis. The RAN model is further validated by using datasets from both a cable-stayed bridge and an arch bridge. Excellent classification performance is obtained, as well

			as a generalisation, for the identification of most abnormalities. RAN is capable of outperforming past deep learning and preprocessing methods.
[4]	2023	Transfer Learning	This paper aims to predict autistic traits in children using eye gaze analysis. This paper briefs the procedure to perform eye gaze analysis and explains its benefits in the prediction of mental health disorders. Attentiveness may be further useful in predicting mental disorders. This paper presents a performance of transfer learning models for eye gaze analysis. InceptionV3 is best best-performing model with an accuracy of 88% on a dataset by Zenodo.
[5]	2022	Dense Residual Neural Network (DRN) Bi-directional CNN, LSTM.	Improvement of tool wear monitoring in modern-day production. A new multistep tool wear prediction framework has been proposed in this paper using the feature normalization and deep learning algorithms combined with an attention mechanism. In this context, the objective would be to forecast and monitor the multistep course of tool wear concurrently. For this, the system entails a bi-directional long short-term memory (BiLSTM) with a parallel convolutional neural network (CNN) for the monitoring of tool conditions.
[6]	2021	Hybrid of CNN and LSTM with Residuals ResNet	The present work proposes a deep learning model based on the Sequence to Sequence Model with Attention and Monotonicity Loss (SMA ML) that allows for simultaneous monitoring and prediction of the tool wear in the machining process. The data-driven techniques are found not to develop well the trend of tool wear degradation during continuous cutting, and the traditional tool wear monitoring involved laborious feature extraction and specialized knowledge.
[7]	2021	Autoregressive Integrated Moving Average Attention Mechanism for LSTM Network (ARIMA)	This work proposed a novel Sequence-to-Sequence Model with Attention and Monotonicity Loss (SMAML) under deep learning for the concurrent detection and prediction of tool wear in a machining process. Data-driven techniques have not captured the degradation trend of tool wear under continuous cutting, but traditional monitoring of tool wear has been highly labour intensive for feature extraction and asked for specialized knowledge. SMAML is an encoder-decoder architecture, with an integrated attention mechanism, as well as an additional loss function towards monotonicity within the same framework for sequence-to-sequence processing. These help to reduce the maintenance cost, enhance tool wear monitoring, and make early replacement decisions for the tools. It also brings out interpretability and clarity on tasks' relationships regarding the use of tools and sensor signals.
[8]	2020	CNN RNN	The research introduces the significance of developing the automated FER system in order to comprehend human emotion. It focuses on the applications of these systems of FER that are automated, with special emphasis on the interactions of humans and machines, within the context of the medical domain. There

			is a brief overview of historical FER research with reference to newer works present in the literature that have applied deep learning techniques and more specifically convolutional neural networks (CNN) to FER. It is focused on summarizing the recent deep learning methodologies and ideas in context of FER.
[9]	2020	Bidirectional LSTM Network using CNN (BiLSTM) Techniques for Attention Mechanism	The project has been planned for monitoring hydraulic systems at manufacturing places through deep learning. Data augmentation techniques are used are Jittering and Scaling, to cope with the lack of availability of the required data. The model, proposed in this work, further integrates an attention mechanism into bidirectional long short-term memory network (BiLSTM), convolutional neural network (CNN), and both for real-time monitoring. The results show that the model is effective in the monitoring of conditions of hydraulic systems that are required for industrial applications when data is poor.
[10]	2020	CNNs: Nonconvex Optimization Hybrid Approach	The project aims at developing a smartphone-based eye-tracking gadget. Calibration uses a geometric gaze estimation approach together with convolutional neural networks (CNNs) for purposes of feature extraction for reliable gaze tracking. During the calibration process, a nonconvex optimisation approach is used in the estimation of user-specific properties of the eye. Though the general motive is to make mobile eye-tracking a reach-for-all affair, it is coupled with general use by privacy concerns. The study assesses the system in a bid to minimize gaze estimate bias.
[11]	2020	Spatio-temporal CNN GRU cell RNN	The problem of the "speaking effect," a phenomena in which articulation of the voice during discussions affects the perception of the expression of the face, is addressed with two popular models of deep neural networks in this paper. These models are the spatio-temporal CNN and the GRU cell RNN. In the first instance, these models are only trained for the facial features, while in the latter, they are trained for the combination of the facial traits and the signs related to articulation of speech. On the other hand, the addition of articulatory related variables is shown to increase accuracy up to 12% in identifying emotion. Furthermore, models display higher accuracy as input consecutive frames increase.
[12]	2018	CDAE-CNN DVAE-CNN DAE-CNN	The classification of noisy picture using hybrid models was considered in the present research. It uses the denoising variational autoencoders (DVAE) together with the denoising autoencoders (DAE), while the last type of network convolutional denoising autoencoders (CDAE). These hybrid models present higher performances in noisy image classification to methods existing before, especially if they are developed with low noise. The DVAE-CNN model performed well with regular noise images, while the DVAE-CDAE-CNN model had even better performance with high-noise images. This hybrid approach powered by autoencoders and CNNs improved noisy image classification.

[13]	2017	CNN	In this research, the challenging problem of facial expression identification in computer vision is undertaken by using a convolutional neural network (CNN). The architecture was tuned in the research using the Visual Geometry Group (VGG) model and benchmarked on various public databases, and it resulted in an improvement over their results for facial emotion analysis. The work highlights how well CNNs interpret facial expressions and how this could improve human-machine communication.
[14]	2006	Perona-Malik model Total Variation Minimization (TVM) Motion by Mean Curvature (MMC)	This work presents a hybrid model for image restoration that treats an image corrupted by both impulses and Gaussians through a unified model by incorporating the combined Perona-Malik, Total Variation Minimization (TVM) and Motion by Mean Curvature (MMC) models. The impulse noise reduction was further introduced into the research with minimal dissipation difference schemes, Essential Nondissipative (ENoD) to preserve the picture edges. The present research considers the classification of noisy picture using hybrid models. For denoising, the proposed model and numerical approaches can cater to both grayscale and color images, with the technique of chromaticity-brightness decomposition as the main for color images.
[15]	2005	Support Vector Machines Linear Discriminant Analysis Adaboost	In one of the studies, there is offered a comparative study with a few of the machine learning methods involved in face emotion recognition. It has developed AdaBoost and Support Vector Machines to offer great accuracy through feature selection and classification. The device being real-time allows accurate recognition of facial action units as well as basic emotional emotions. Data used was from the Cohn and Kanade's DFAT-504 dataset. In this way, the research done demonstrates the importance of real-time processing for practical applications.

III. TECHNIQUES USED

A. VGG16

Define The VGG16 architectural design of a convolutional neural network (CNNs) is renowned for its high effectiveness and simplicity. It comprises three fully linked and thirteen weight convolutional layers. The tiny 3x3 convolutional layers facilitate capturing of small details concerning information in the network through their small filters. The implemented architecture diagram is as shown in Figure 1.

Application: VGG16 is applicable to purposes such as image classification applications. It has been applied in applications such as object recognition and scene understanding. Due to the customization, it can undergo for extracting attention-based aspects of eyegaze data in the context of the problem, it is appropriate for comprehending gaze patterns and attention shifts.

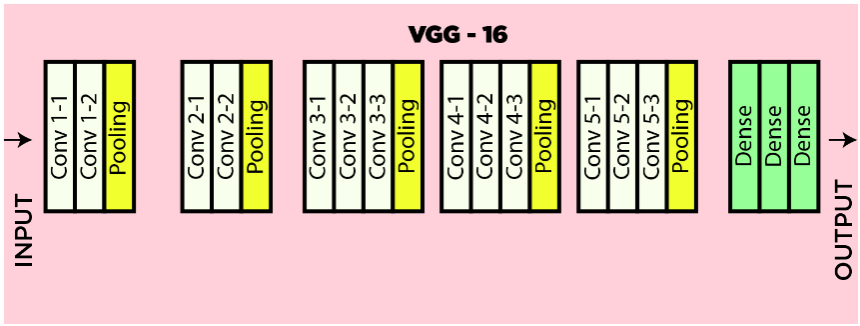

Figure 1. VGG16 Architecture Diagram

B. VGG19

It is deeper, with another 19 more layers extending VGG16. And it is the simplicity and uniformity of its design, and as mentioned before continuing with 3x3 convolutional filters. The architecture diagram is depicted in Figure 2.

Applications: Similar scenarios are employed using VGG19 and VGG16 largely in the image classification applications. When it comes to attention monitoring, its increased depth can be used to highlight a more detailed pattern of maybe improved analysis over underlying data eye gazes for studying the attention patterns.

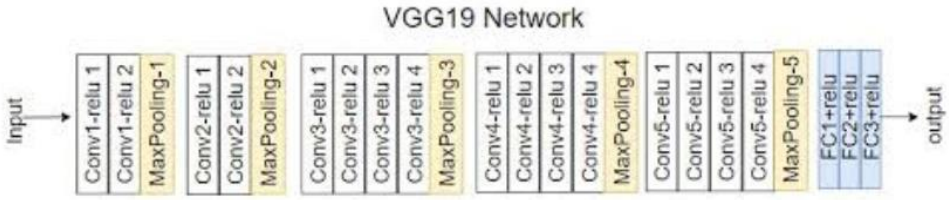

Figure 2. VGG19 Architecture Diagram

C. Inception V3

InceptionV3 is one of the members of a family of architectures called Inception. Like other Inception models, it makes use of conv layers with multiple filter sizes (1x1, 3x3, and 5x5) as well as pooling layers. It further exploits batch normalization and residual connections to make training stable. The architecture after this is illustrated in the Figure 3 that follows.

Application: With several computer vision tasks that are pertinent to the architecture InceptionV3 such as object detection and image categorization. In this regard of monitoring attention, it can, therefore, highlight the patterns of attention for various regions in a visual field at high levels of precision and recall values thus capturing good multi-scale attention dynamics.

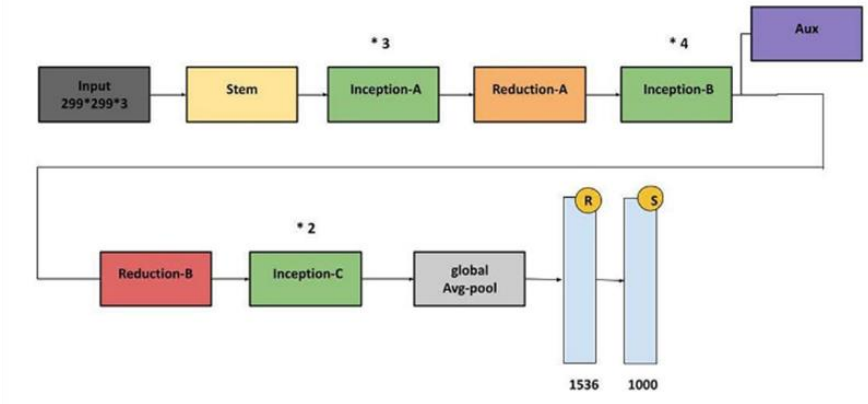

Figure 3. InceptionV3 Architecture Diagram

D. EfficientNetB0 and EfficientNetB7

The model family unites balancing in depth, width, and resolution through compound scaling. Most of the time, the base model is referred to as EfficientNetB0 while one of the larger variants is named EfficientNetB7 with enhanced depth and resolution. The architecture diagrams for EfficientNetB0 and EfficientNetB7 are shown in Figures 4 and 5, respectively.

Applications: This is only one of the applications where the better performance of EfficientNet models can be seen in terms of object detection and image classification. As EfficientNetB0 is a lightweight model, it is suitable to be used for real-time applications whereas EfficientNetB7 can also be applied when attention maps are complex because of the high-resolution eye gaze data.

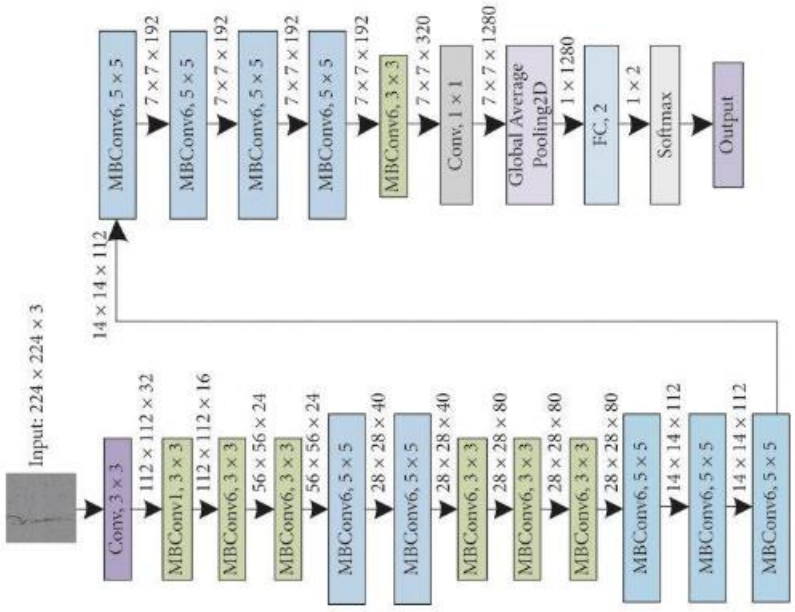

Figure 4. EfficientNetB0 Architecture Diagram

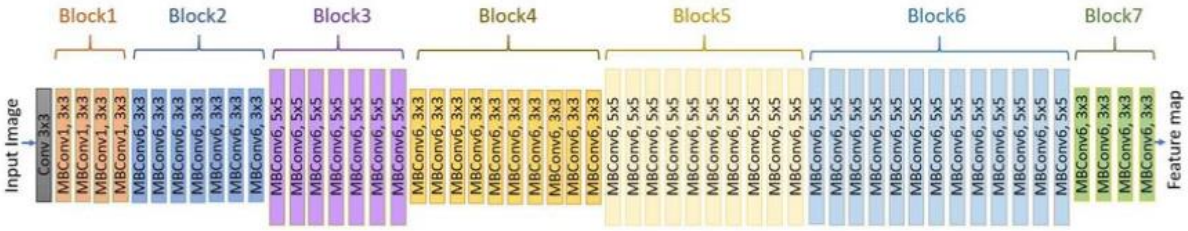

Figure 5. EfficientNetB7 Architecture Diagram

E. InceptionResNetV2

The inception architecture in the InceptionResNetV2 hybrid has additional residual connections in the inception model to facilitate easier training of very deep networks. Also, batch normalization, many convolutional layers with different filter diameters and residual blocks are used. The architectural diagram is as shown below in Figure 6.

Application: InceptionResNetV2 applications include a broad set of computer vision tasks, including object recognition and image classification. It combines the ideas from Inception and ResNet to capture sudden and prolonged changes in gaze attention within the context of attention monitoring. Being fine-tuned and incorporated with their distinctive structures and properties in the "Eye Gaze for Monitoring Attention Through Hybrid Ensemble Learning", these models can become powerful and flexible tools for the registration and analysis of the dynamics of attention in various contexts.

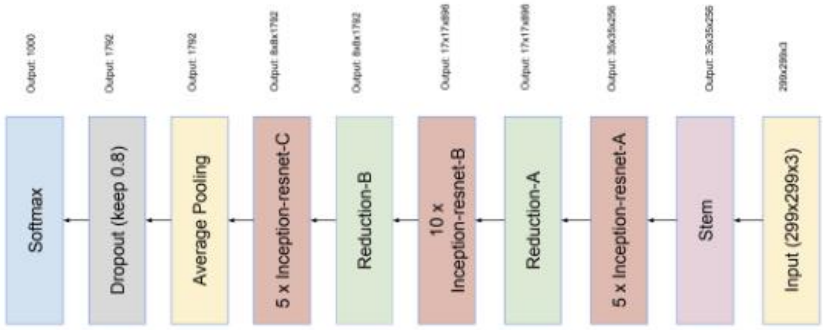

Figure 6. InceptionResNetV2 Architecture Diagram

F. XGBoost

The XGBoost Cross Breed model is a combination of a few powerful brain network architectures like VGG16, VGG19, InceptionV3, InceptionResNetV2 and others. This cross-breeding method, which provides an especially serviceable and viable model to perform highly on all ranges of AI tasks, takes the best and most useful features from numerous other techniques. prebuilt models are coupled with the inclination supporting computation XGBoost to uplift the capability of dealing with difficult highlight extraction, complex case handling and exact requirements of the model. These models are consequently combined in such a way that a new group sstrategy is built which utilizes the different representations of elements learned by each component design.The XGBoost hybrid model generalizes well for various data domains such as object detection and picture classification due to the pre-trained neural networks- learning transfer capabilities. The XGBoost Hybrid technique works to improving both models' accuracy and robustness after incorporating them into a single, integrated framework under which the two models operate in tandem within a flexible and efficient solution that is capable of flourishing in diverse contemporary machined learning application contexts.

IV. Experimental Framework

A. Dataset

Every image in the Kaggle dataset version has 2 class labels out of autistic and non-autistic for its 2936 JPEG images of size 224 * 224 pixel. The files were renamed to help organization purposes. These programs are for both autistic and non-autistic kids. It is also purposed for younger children. Photographs hold faces images for both boys and girls. Data is comprised of three folders within the dataset, which are train, valid, and testing. Each folder contains sections for those with people autism as well as no-autism. So, this particular dataset is chosen as it is openly available without any permission required even to access it. In addition, processing images are much less complex in comparison to videos.

B. Methodology

Creating the "Eye Gaze for Monitoring Attention Through Hybrid Ensemble Learning" involves a number of steps to create, refine and test the model. First, on next six pre-trained models, we performed the procedures as follows: VGG16, Vgg19, InceptionV3, EfficientNetB0, EfficientNetB7, and InceptionResNetV2, in order to increase accuracy and other evaluation metrics [16]. On the work with this done, we have yielded all the evaluation indicators. Next, we combined this XGBoost with these six pre-trained models in a hybrid model approach seeking the similar methods as outlined below. We managed to get all of the evaluation metrics using this method as shown above. Your generalized project methodology for this kind of work is thus given below:

- **Data Collection**: Record eye-movements referring to right sources of eye-gazing information or by making controlled experiments. All data collection should be done responsibly – by adherence to informed consent and privacy laws, if it is on humans.Do not mix complete spellings and abbreviations of units: “Wb/m2” or “webers per square meter”, not “webers/m2”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.

- **Preprocessing Data**: Missing value treatment, noise reduction and improvement on the consistency of the data that will be presented should be done for cleaning and preprocessing the eye gaze data. The pre-processing of the data might also require data preparation through calibration and alignment [17].
- **Data Augmentation**: Perform operations on data augmentation such as rotation, rescaling, shifting, shearing, flipping and zooming to give more varieties within the dataset. Augmentation makes the model more rugged in withstanding and adapting to random variations that occur naturally, thus increasing its generalization capability.
- **Model Choice**: Select a pre-trained deep learning model as the basis of the hybrid model. Popular pre-trained deep learning models include VGG16, VGG19, InceptionV3, EfficientNetB0, EfficientNetB7, and InceptionResNetV2 as previously mentioned [18].
- **Model Architecture**: After the selection of the pre-trained models, the addition of additional layers to trace attention in your hybrid model would be in order. This architecture may comprise convolutional layers, recurrent layers, and fully linked layers for aiming at capturing features of attention accordingly.
- **Training Setup:** This may involve building the hybrid model with additional layers containing attention tracking, which can be fused with the selected pre-trained models. The architecture may contain convolutional, recurrent as well as fully connected layers to properly represent the attention features.
- **Fine-Tuning:** A hybrid model can be created by fusing pre-trained models of interest with additional attention-tracking layers. The foremost elements offered consideration in this design encompass the convolutional, recurrent as well as fully connected layers to fit in the attention aspects.
- **Model Training:** Train your hybrid model with the preprocessed and improved dataset. Monitor your training process metrics like loss and accuracy. Try implementing learning rate schedules or early stopping techniques to improve the effectiveness of your training.
- **Evaluation Metrics**: Evaluate the model using different evaluation metrics that include accuracy, precision, recall, F1-Score, as well as support once it is trained [19]. These evaluation metrics will indeed help to have an indepth understanding of the model capacity to recognize and predict on attention patterns.
- **Model Testing**: Testing the trained hybrid model on new or untested data to access its general capacity. This stage is crucial in order to make sure that the model generality would be good if it would face practical data.
- **Optimization and Fine-Tuning**: Modify the model based on the results of testing and conclusions from evaluation to improve it. Modify some of its parameters to make its performance even better.
- **Deployment:** Deploy the model to fit applications such as user interfaces, medical monitoring systems, instruction platforms, or anything that needs at least similar performance criteria fitted earlier.

C. System Architecture

The architecture diagram that has been used during the model training and testing stages is represented below in Figure 7. After first splitting data into train and test datasets, beforehand, the train dataset is given to pre-trained model which takes out the features of the dataset. So, such traits are now directly passed on to the XGBoost model, boosting this hybrid for robustness and efficiency.

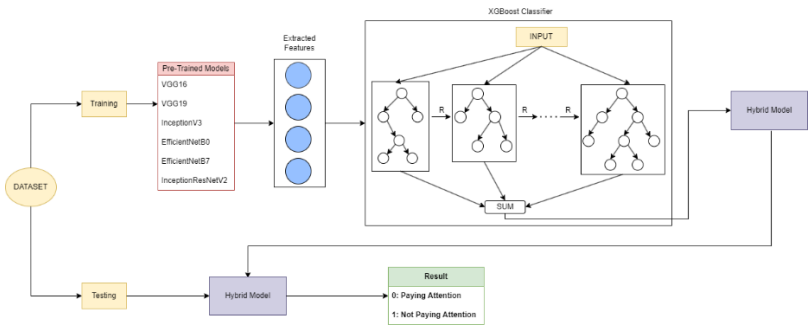

Figure 7. System Architecture

D. Performance Metrics

Some of the major performance metrics to gauge the effectiveness for "Eye Gaze for Monitoring Attention Through Hybrid Ensemble Learning" include:

- Accuracy:

$$Accuracy = TP + TN/Total\ Predictions \quad (1)$$

- Precision:

$$Precision = TP/TP + FP \quad (2)$$

- Recall (Sensitivity):

$$Recall = TP/TP + FN \quad (3)$$

- F1-Score:

$$F1 - Score = 2*(precision * recall)/precision + recall \quad (4)$$

- Mean Absolute Error (MAE):

$$MAE = \sum(predicated - actual) / n \quad (5)$$

- Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{(\sum (predicted\text{-}actual)^2)/n} \quad (6)$$

V. RESULT

This paper has empirically looked into the performance of six selected pre-trained models (VGG16, VGG19, InceptionV3, EfficientNetB0, EfficientNetB7, and InceptionResNetV2) in attention tracking. The achieved testing and training accuracies are summarized in Table 2.

Table 2. Accuracy Table for Pre-Trained Model

Pre-Trained Model	Training Accuracy	Testing Accuracy
VGG16	94.00%	82.30%
VGG19	94.30%	84.60%
InceptionV3	83.80%	77.90%
EfficientNetB0	86.30%	81.60%
EfficientNetB7	82.10%	80.30%
InceptionResNetV2	87.80%	83.60%

A. Performance Evaluation of Pre-trained Models

- Accuracy and Loss Plots: The accuracy and plot graphs for the best pre-trained models are depicted below in Figure 8 and Figure 9.

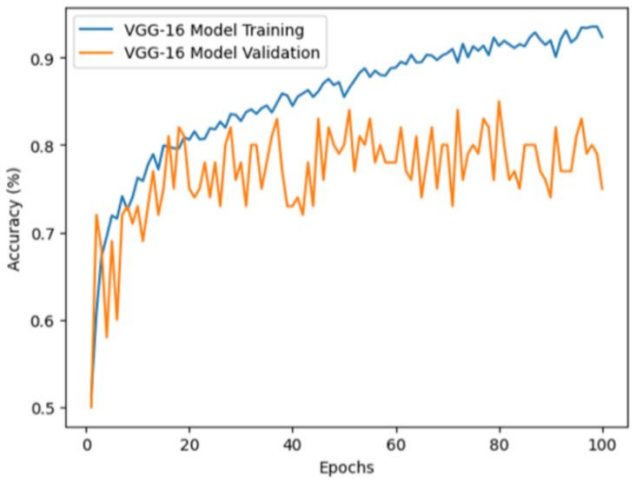

Figure 8. VGG16 Accuracy Plot

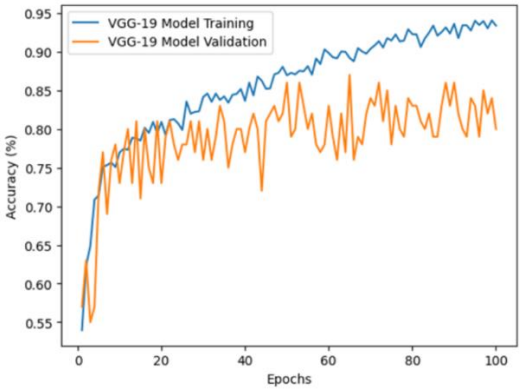

Figure 9. VGG19 Accuracy Plot

- Confusion Matrices: Below in Figure 10 and Figure 11 are the confusion matrices for the best pre-trained models.

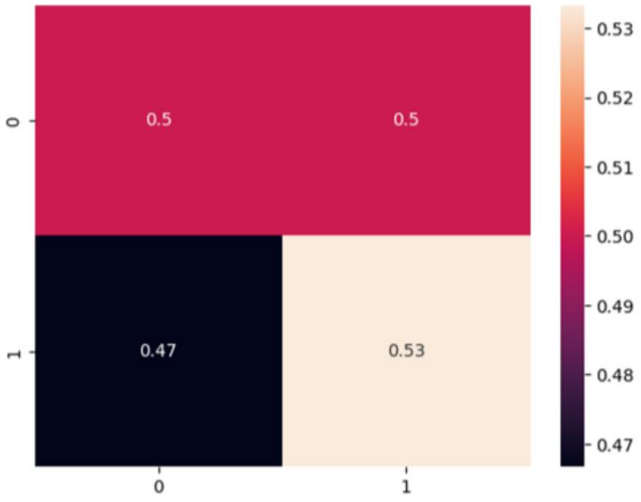

Figure 10. VGG16 Confusion Matrix

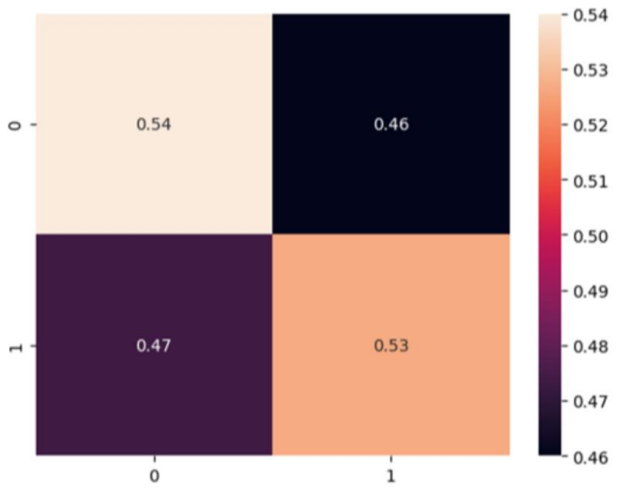

Figure 11. VGG19 Confusion Matrix

- Evaluation Results: In the following, we show all evaluation results, including precision, recall, f1-score, and support for pre-trained models in Table 3.

Table 3. Evaluation Results for Pre-Trained Models

Models	Precision		Recall		F1-Score		Support	
	0	**1**	**0**	**1**	**0**	**1**	**0**	**1**
VGG16	0.52	0.52	0.50	0.53	0.51	0.52	150	150
VGG19	0.53	0.53	0.54	0.53	0.54	0.53	150	150
InceptionV3	0.46	0.47	0.43	0.50	0.45	0.48	150	150
EfficientNetB0	0.50	0.00	1.00	0.00	0.67	0.00	150	150
EfficientNetB7	0.50	0.00	1.00	0.00	0.67	0.00	150	150
InceptionResNetV2	0.54	0.53	0.49	0.58	0.52	0.56	150	150

B. Hybrid Model Approach

Finally, all of the pre-trained models were hybridized with XGBoost. Surprisingly, every hybrid model showed high learning as high as 100% in training accuracy.

Comparing the testing accuracies of the models with the different display in Table 4 is displayed.

Table 4. Accuracy table for Hybrid Models

Hybrid Model	Training Accuracy	Testing Accuracy
VGG16 with XGBoost	100%	82.00%
VGG19 with XGBoost	100%	81.30%
InceptionV3 with XGBoost	100%	75.60%

EfficientNetB0 with XGBoost	100%	78.00%
EfficientNetB7 with XGBoost	100%	72.30%
InceptionResNetV2 with XGBoost	100%	77.60%

C. Performance Evaluation of Hybrid Models

- Confusion Matrices: Post these two graphs, the confusion matrices from the best performing hybrid models are given in Figure 12 and Figure 13.

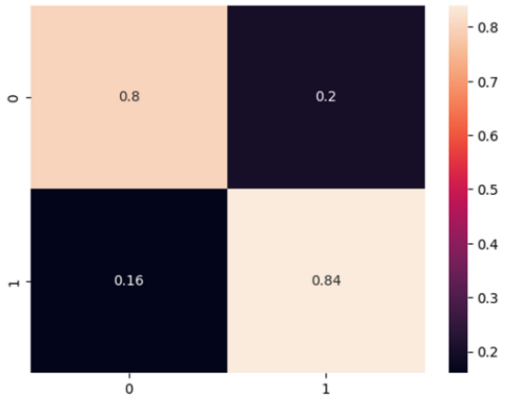

Figure 12. VGG16-XGBoost Confusion Matrix

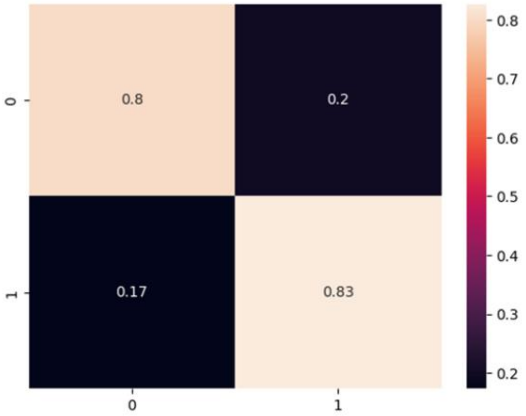

Figure 13. VGG19-XGBoost Confusion Matrix

- Evaluation Results: These are presented in Table 5 together with the remaining components, such as precision, recall, f1-score and support.

Table 5. Evaluation Results for Hybrid Models

Hybrid Models	**Precision**		**Recall**		**F1-Score**		**Support**	
	0	**1**	**0**	**1**	**0**	**1**	**0**	**1**
VGG16-XGBoost	0.52	0.52	0.50	0.53	0.51	0.52	150	150
VGG19-XGBoost	0.53	0.53	0.54	0.53	0.54	0.53	150	150

InceptionV3-XGBoost	0.46	0.47	0.43	0.50	0.45	0.48	150	150
EfficientNetB0-XGBoost	0.50	0.00	1.00	0.00	0.67	0.00	150	150
EfficientNetB7-XGBoost	0.50	0.00	1.00	0.00	0.67	0.00	150	150
InceptionResNetV2-XGBoost	0.54	0.53	0.49	0.58	0.52	0.56	150	150

Besides, the VGG16-XGBoost and VGG19-XGBoost models outperformed other hybrid models in the monitoring attention test. They conducted testing accuracies at least higher than other hybrid models, indicating that models perform better than other hybrid models [20].

VI. Discussion and Future Work

These findings of the pre-trained and hybrid model study were further confirmed through more investigation using additional assessment metrics, as well as through differences in the performance of individual pre-trained models. Generally, the work confirms the possibility of integration of XGBoost with pre-trained models for attention monitoring and points to the highest performance of models VGG16-XGBoost and VGG19-XGBoost.

Following are a few key points that need to be considered for proposing the future of this work.

- **Multi-modal data integration:** Integrating eye-gaze data with other data sources of a biometric nature, like heart rate, electroencephalography (EEG), and facial expressions, to produce a more comprehensive attention monitoring model. Multi-modal integration is way more telling in the sense of how a person is focused in their attention.

- **Real-Time Monitoring and Feedback:** Provide customers with real-time solutions to help them keep track of attention. Technologies of this kind are potentially useful in practice areas such as education, where feedback on the level of attention is customized and enhances learning opportunities.

- **Collaboration in Neuroscience**: Collaboration with neuroscientists will be encouraged for the enrichment of research work in neuroscience with a better understanding of the processes of the brain related to attention and their relationship with gaze patterns.

VII. Conclusion

The leading-edge multi-disciplinary approach termed "Eye Gaze for Monitoring Attention through Hybrid Ensemble Learning" summary leverages eye gaze, deep learning, and data augmentation for recognition, understanding, and prediction of patterns of attention within a wide range of environments. Thus, the potential of such a novel paradigm includes application domains with broad areas of influence as education, healthcare diagnostics, or the optimization of the interaction between the human and computer in our currently understood status of attention tracking.

This model makes use of pre-trained models like VGG16, VGG19, InceptionV3, EfficientNetB0, EfficientNetB7, and InceptionResNetV2, all of which allow the powerful characteristics of efficient network designs, multi-scale feature utilization, Rotation, shift, and rescaling augmentations are performed to provide the model with the capacity to generate more realistic kinds of objects with other representations that it was not trained on, hence making the model more robust and flexible so that it functions well in the real world.

References

[1] Jegham and others, “Deep learning-based hard spatial attention for driver in-vehicle action monitoring,” Expert Syst Appl, vol. 219, p. 119629, 2023.
[2] Z. Trabelsi and others, “Real-Time Attention Monitoring System for Classroom: A Deep Learning Approach for Student’s Behavior Recognition,” Big Data and Cognitive Computing, vol. 7, no. 1, p. 48, 2023.
[3] X. Lei and others, “Mutual information based anomaly detection of monitoring data with attention mechanism and residual learning,” Mech Syst Signal Process, vol. 182, p. 109607, 2023.

[4] R. V. Bidwe, S. Mishra, and S. Bajaj, “Performance evaluation of Transfer Learning models for ASD prediction using non-clinical analysis,” in Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing, New York, NY, USA: ACM, Aug. 2023, pp. 474–483. doi: 10.1145/3607947.3608050.
[5] M. Cheng and others, “Intelligent tool wear monitoring and multi-step prediction based on deep learning model,” J Manuf Syst, vol. 62, pp. 286–300, 2022.
[6] G. Wang and F. Zhang, “A sequence-to-sequence model with attention and monotonicity loss for tool wear monitoring and prediction,” IEEE Trans Instrum Meas, vol. 70, pp. 1–11, 2021.
[7] L. Li and others, “Monitoring and prediction of dust concentration in an open-pit mine using a deep-learning algorithm,” J Environ Health Sci Eng, vol. 19, pp. 401–414, 2021.
[8] S. A. R. ˙I. Meriem, A. Moussaoui, and A. Hadid, “Automated facial expression recognition using deep learning techniques: an overview,” International Journal of Informatics and Applied Mathematics, vol. 3, no. 1, pp. 39–53, 2020.
[9] K. Kim and J. Jeong, “Real-time monitoring for hydraulic states based on convolutional bidirectional LSTM with attention mechanism,” Sensors, vol. 20, no. 24, p. 7099, 2020.
[10] B. Brousseau, J. Rose, and M. Eizenman, “Hybrid eyetracking on a smartphone with CNN feature extraction and an infrared 3D model,” Sensors, vol. 20, no. 2, p. 543, 2020.
[11] S. Bursic and others, “Improving the accuracy of automatic facial expression recognition in speaking subjects with deep learning,” Applied Sciences, vol. 10, no. 11, p. 4002, 2020.
[12] S. S. Roy, M. Ahmed, and M. A. H. Akhand, “Noisy image classification using hybrid deep learning methods,” Journal of Information and Communication Technology, vol. 17, no. 2, pp. 233–269, 2018.
[13] Fathallah, L. Abdi, and A. Douik, “Facial expression recognition via deep learning,” in 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), 2017.
[14] S. Kim, “PDE-based image restoration: A hybrid model and color image denoising,” IEEE Transactions on Image Processing, vol. 15, no. 5, pp. 1163–1170, 2006.
[15] M. S. Bartlett and others, “Recognizing facial expression: machine learning and application to spontaneous behavior,” in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), 2005.
[16] R. V Bidwe and others, “Deep Learning Approaches for Video Compression: A Bibliometric Analysis,” Big Data and Cognitive Computing, vol. 6, no. 2, p. 44, Apr. 2022, doi: 10.3390/bdcc6020044.
[17] D. Mane, K. Shah, R. Solapure, R. Bidwe, and S. Shah, “Image-Based Plant Seedling Classification Using Ensemble Learning,” 2023, pp. 433–447. doi: 10.1007/978-981-19-2225-1_39.
[18] S. Nalwar et al., “EffResUNet: Encoder Decoder Architecture for Cloud-Type Segmentation,” Big Data and Cognitive Computing, vol. 6, no. 4, p. 150, Dec. 2022, doi: 10.3390/bdcc6040150.
[19] D. Mane, R. Bidwe, B. Zope, and N. Ranjan, “Traffic Density Classification for Multiclass Vehicles Using Customized Convolutional Neural Network for Smart City,” 2022, pp. 1015–1030. doi: 10.1007/978-981-19-2130-8_78.
[20] G. Agrawal, U. Jha, and R. Bidwe, “Automatic Facial Expression Recognition using Advanced Transfer Learning,” in Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing, New York, NY, USA: ACM, Aug. 2023, pp. 450–458. doi: 10.1145/3607947.3608047.