Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 2 (2025)

Attention Monitoring Using Eye Gaze with a Hybrid Ensemble Learning Approach

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

[1,2,3,4,5,6] 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.

[1]ranjeetbidwe@hotmail.com, [2]gouransh12345@gmail.com, [3]unnatijha2001@gmail.com,
[4]akshaysangwan8571@gmail.com, [5]himanshukulhari28@gmail.com,
[6]sashikala.mishra@sitpune.edu.in, [7]k.bajaj@westernsydney.edu.au

Corresponding Author: ranjeetbidwe@hotmail.com[1]

Article History:

Received: *03-08-2024*

Revised: *11-09-2024*

Accepted: *20-09-2024*

Abstract:
Healthcare, education, transportation safety, and human-computer interaction are just few of the many tasks that require monitoring. Monitoring is important for all of these tasks and more. This paper presents innovative work that has been done in the field of attention monitoring. The work involves combining a hybrid eye gaze model with deep learning in order to monitor the level of attention that a driver is paying. This paper provides a description of the hybrid eye gazing model that was suggested, as well as the results that were produced by the model. Data augmentation techniques such as rotation, shifting, shearing, and flipping are used to the proposed model, along with adjustments such as changing the fill mode in terms of zooming into the image and rescaling. The suggested model makes use of an augmented dataset. In order to ensure that the model is trained in a reliable and consistent manner, all of these aspects are essential. Modern pre-trained architectures, such as VGG16, VGG19, InceptionV3, EfficientNetB0, EfficientNetB7, and InceptionResNetV2, are the foundation upon which our model is constructed. These designs are modified, and then more layers are added, in order to facilitate the process of capturing very minute attention dynamics using them. The accuracy and effectiveness of the model were later improved by the utilization of a model ensemble. After some time has passed, the XGBoost model is amalgamated with all of the other models that were utilized previously in the hybrid model technique. This is done in order to achieve improved accuracy and efficiency of the model. Many different assessment measures, such as accuracy, precision, recall, F1 Score, and support, are utilized in order to conduct an adequate evaluation of the performance of the model. Through the use of these indicators, a comprehensive comprehension of the model's capacity to identify and forecast attention patterns in a variety of locations is achieved. We were able to obtain the highest level of accuracy from VGG19 and InceptionResNetV2 after utilizing the models, which was 84.6% and 83.6% respectively. A score of 82% was achieved by the VGG16 hybrid models during the accuracy test. The Hybrid Eye Gaze Model is a powerful and adaptable attention monitoring system that can be utilized for a wide range of applications. It utilizes deep learning and pre-trained architectures.

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

1. Introduction

The capability to perceive and track human attention is of utmost relevance for a wide variety of applications and sectors in the current digital landscape, which is characterized by a rapid speed. The autonomous driving technology that relies on attention monitoring is one example of these technologies. These technologies have a wide range of applications, including improved end-user experiences, improved healthcare diagnostics for creating optimal adjustments in instructions, and safety in high-stakes scenarios. These developments have resulted in the creation of a number of very sophisticated and cutting-edge technologies that have been developed expressly for the purpose of sensing and analyzing the delicate nuances of human attention.

Eye gaze tracking is considered to be one of the most successful and promising techniques in the field of attention detection. It stands out among the other approaches that are available. It is possible to learn a great deal about a person's priorities, interests, and areas of interest in a particular setting by monitoring what they are looking at. The application of this information will help in the generation of interfaces that are more user-friendly, the detection of drivers who are weary, and the implementation of tailored learning initiatives, amongst many other things. On the other hand, methodologies that are used for eye gaze tracking in the real world may have some limitations. These limitations may include sensitivity to the surrounding environment, difficulties with calibration, and a restricted ability to be applied in typical operating situations.

The findings of this study provide a novel approach to addressing these deficiencies through the utilization of the Hybrid Eye Gaze Model for Attention Monitoring. The best characteristics of a number of different eye-tracking algorithms were combined in this model, and data augmentation was also incorporated in order to increase the robustness of the model. Additionally, this model incorporates contemporary deep learning architectures in order to monitor attention in a manner that is both responsive and adaptive. The methodology makes use of models that have already been trained, such as VGG16, VGG19, InceptionV3, EfficientNetB0, EfficientNetB7, and InceptionResNetV2 [9] [16] [18] [22]. Additional layers are added to the methodology that has been established in order to capture attention patterns that are more fine-grained.

Various evaluation metrics, such as the F1-score, are utilized in order to conduct an in-depth analysis of the hybrid eye gaze model, which includes the examination of features such as accuracy, precision, recall, support, and performance. In the assessments, the capabilities of the model will be presented, along with how they might revolutionize attention monitoring by providing an interpretable, flexible, trustworthy, and scalable method of enhancing human attention in a variety of settings.

2. Literature Review

The comprehensive examination of scholarly literature on the subject was used to set up Table 1, which is based on the significant hypotheses, evidence found, and noticeable gaps that were deduced from this paper. It was possible to motivate and put into action the system that was described in this research with the assistance of these surveyed approaches.

Table 1. Literature Review

References	Year	Algorithm Used	Summary
[1]	2023	Driver action recognition (DAR).	In order to address the pressing issue of distracted driving, this research suggests a novel hard attention network that recognizes driver actions in actual driving scenarios. The hard attention mechanism increases accuracy in distinguishing safe driving and distraction because it concentrates on important information related to the drivers and ignores unimportant information. The findings are remarkable: the accuracy rate for distraction detection may be as high as 99.07%, while the accuracy rate for safe driving recognition may reach up to 95.83%. In the meanwhile, its computational efficiency is higher than that of soft attention-based models. This work will make a big difference in traffic safety.
[2]	2023	YOLOv5	This study's method uses high-definition cameras, user-friendly interfaces, behavior detection, and face recognition technology powered by YOLOv5 to track students' attendance, attentiveness, and mood in real-time throughout class sessions.
[3]	2023	CNNs Residual Learning (ResNet) RNN	Finding aberrant data in structural measure monitoring data is a major difficulty in this study since it has a big impact on condition assessment. In order to improve both speed and accuracy in the categorization task, the researchers suggested employing a Residual Attention Network (RAN) incorporating residual learning and attention mechanisms. Their method involved using mutual information correlation analysis to transform hourly segmented data into matrix form. Using datasets from an arch bridge and a cable-stayed bridge, the RAN model was further verified and showed good classification performance and generalization for detecting most abnormalities. The RAN performed better than earlier preprocessing and deep learning techniques.
[4]	2023	Transfer Learning	Using eye gaze analysis, the purpose of this work is to make a prediction about autistic features in children. This article provides a concise overview of the process of doing eye gaze analysis and elucidates

			the advantages of using this technique for the diagnosis of mental health conditions. It is possible that attentiveness can also be effective in the prediction of mental problems. The purpose of this research is to give a performance evaluation of transfer learning models to analyze eye gazing. The best-performing model is InceptionV3, which achieved an accuracy of 88% when applied to a dataset developed by Zenodo.
[5]	2022	Dense Residual Neural Network (DRN) Bi-directional CNN, LSTM.	Combining feature normalization, deep learning algorithms, and an attention mechanism, this study provides an enhanced framework for monitoring tool wear in current production. This framework is intended to be used in modern production. While concurrently forecasting and monitoring the course of tool wear over multiple steps, the goal is to achieve this. In order to efficiently monitor tool conditions throughout the process, the system makes use of a bi-directional long short-term memory (BiLSTM) network in conjunction with a parallel convolutional neural network (CNN).
[6]	2021	Hybrid of CNN and LSTM with Residuals ResNet	The current study suggests a deep learning model that enables the simultaneous monitoring and prediction of tool wear during the machining process. This model is based on the Sequence to Sequence Model with Attention and Monotonicity Loss (SMA ML). The trend of tool wear degradation during continuous cutting is not well developed by data-driven methodologies, and traditional tool wear monitoring required specialist knowledge and arduous feature extraction.
[7]	2021	Autoregressive Integrated Moving Average Attention Mechanism for LSTM Network (ARIMA)	In order to detect and predict tool wear in a machining process simultaneously, this work developed a unique Sequence-to-Sequence Model with Attention and Monotonicity Loss (SMAML) under deep learning. While traditional tool wear monitoring has required specialized knowledge and has required a high labor-intensive feature extraction process, data-driven solutions have not been able to capture the degradation trend of tool wear under continuous cutting. Within the same framework for sequence-to-sequence processing, SMAML is an

			encoder-decoder architecture that incorporates an additional loss function towards monotonicity and an integrated attention mechanism. These assist in lowering maintenance costs, improving tool wear monitoring, and determining whether to replace tools early on. It also highlights the interpretability and clarity of the linkages between tasks in terms of tool use and sensor signal interpretation.
[8]	2020	CNN RNN	This study emphasizes how crucial it is to create automated facial expression recognition (FER) systems in order to comprehend human emotions more fully. It highlights the uses of automated FER systems, especially in the realm of medicine and human-machine interactions. The article references more recent publications that use deep learning techniques, notably convolutional neural networks (CNN), for FER, and gives a brief summary of historical FER research. The main goal is to provide an overview of the most recent deep learning ideas and techniques within the framework of FER.
[9]	2020	Bidirectional LSTM Network using CNN (BiLSTM) Techniques for Attention Mechanism	Through the use of deep learning, the project is intended to monitor hydraulic systems that are located in manufacturing facilities. Jittering and scaling are two examples of data augmentation techniques that are utilized in order to address the issue of insufficient availability of the necessary data. For the purpose of real-time monitoring, the model that is suggested in this study incorporates an attention mechanism into a bidirectional long short-term memory network (BiLSTM), a convolutional neural network (CNN), and both of these networks. The findings demonstrate that the model is efficient in the monitoring of the conditions of hydraulic systems, which are necessary for industrial applications in situations when there is a lack of complete data.
[10]	2020	CNNs: Nonconvex Optimization Hybrid Approach	The project intends to produce an eye-tracking device based on a smartphone. In order to reliably track gaze, calibration use convolutional neural networks (CNNs) in conjunction with a geometric gaze estimation approach to extract features. The estimation of user-specific properties of the eye is

			done by a nonconvex optimisation approach during the calibration phase. Privacy problems accompany the widespread usage of mobile eye tracking, despite the overarching goal of making it accessible to all. The study evaluates the system with the goal of reducing gaze estimate bias.
[11]	2020	Spatio-temporal CNN GRU cell RNN	In this paper, two widely used deep neural network models are used to address the issue of the "speaking effect," a phenomenon in which articulation of the voice during discussions impacts the perception of the face's expression. The spatio-temporal CNN and the GRU cell RNN are these models. In the first case, these models are trained solely on facial features; in the second case, they are trained on both facial traits and indicators associated with speech articulation. However, it has been demonstrated that include articulatory-related variables can improve accuracy in determining emotion by up to 12%. Additionally, models show increased accuracy with increasing numbers of consecutive input frames.
[12]	2018	CDAE-CNN DVAE-CNN DAE-CNN	The current study took hybrid model categorization of noisy images into consideration. In contrast to the previous type of network convolutional denoising autoencoders (CDAE), it employs denoising variational autoencoders (DVAE) in conjunction with denoising autoencoders (DAE). Compared to earlier methods, these hybrid models perform better in noisy image classification, particularly if they were built with minimal noise. When it came to photos with regular noise, the DVAE-CNN model did well; but, when it came to images with excessive noise, the DVAE-CDAE-CNN model performed even better. This hybrid method enhanced the classification of noisy images by using CNNs and autoencoders.
[13]	2017	CNN	This study investigates the challenging task of facial emotion recognition in computer vision using a convolutional neural network (CNN). After making adjustments to the architecture using the Visual Geometry Group (VGG) model and testing it on many public datasets, the architecture showed improved performance for face expression analysis.

			The study highlights how well CNNs interpret emotions on faces and how this could lead to better human-machine communication.
[14]	2006	Perona-Malik model Total Variation Minimization (TVM) Motion by Mean Curvature (MMC)	The hybrid model for image restoration presented in this work treats an image corrupted by both Gaussians and impulses through a unified model by incorporating the combined Perona-Malik, Total Variation Minimization (TVM), and Motion by Mean Curvature (MMC) models. The present research considers the classification of noisy pictures using hybrid models. The suggested framework and mathematical methods can cater to both grayscale and color images for denoising, with the technique of chromaticity-brightness break down being the main for color images.
[15]	2005	Support Vector Machines Linear Discriminant Analysis Adaboost	A comparison of a few machine learning techniques used in face emotion recognition is provided in one of the research. It has created AdaBoost and Support Vector Machines to choose and classify features with high accuracy. Because the technology operates in real-time, it can accurately recognize both basic emotional feelings and facial action units. The DFAT-504 dataset by Cohn and Kanade provided the data that was used. The study's findings thus highlight how crucial real-time processing is for real-world uses.

3. TECHNIQUES USED

A. VGG16

The convolutional neural network (CNN) architecture known as VGG16 is generally acknowledged for its ease of use and high level of efficiency. Convolutional layers number thirteen, while completely connected layers make up the remaining three levels. The network is able to collect minute details in the data on account of the small 3x3 convolutional filters. Presented in Figure 1 is an illustration of the architecture that was implemented.

Application: The VGG16 algorithm can be utilized for a variety of purposes, including picture classification applications. Object identification and scene comprehension are two examples of applications that have previously utilized this technology. As a result of the fact that it may be customized for the purpose of extracting attention-based features of eyegaze data within the context of the problem, it is suitable for understanding gaze patterns and attention changes.

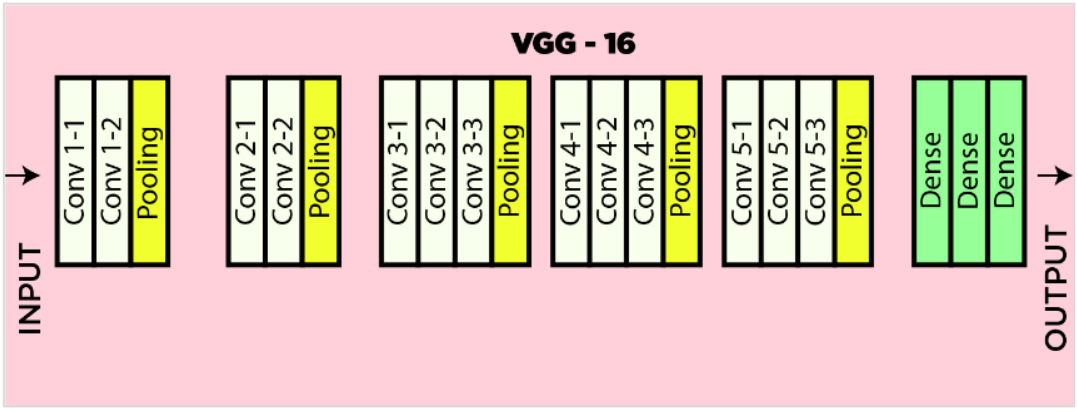

Figure 1. VGG16 Architecture Diagram

B. VGG19

The depth of it is greater, with an additional 19 layers that extend beyond VGG16. Additionally, it is the straightforwardness and consistency of its design, as was discussed earlier, as well as the continuation of the 3x3 convolutional filter configuration. Figure 2 presents a diagrammatic representation of the architecture.

Applications: In the majority of picture classification applications, VGG19 and VGG16 are utilized to handle instances that are very similar to one another. Whenever it comes to attention monitoring, its greater depth may be utilized to emphasize a more detailed pattern of possibly superior analysis of underlying data eye gazes for the purpose of researching attention patterns.

Figure 2. VGG19 Architecture Diagram

C. Inception V3

InceptionV3 is a member of the entire family of architectures that are collectively referred to as Inception. In the same way that previous Inception models do, it employs pooling layers in addition to convolutional layers that have several filter sizes (1x1, 3x3, and 5x5). In addition, it makes use of batch normalization and residual connections in order to ensure that training is consistent. After this, the architecture is depicted in Figure 3, which is presented in the next article.

Application: With a number of computer vision tasks that are relevant to the architecture InceptionV3, such as the detection of objects and the classification of images. As a result, it is able to emphasize the patterns of attention for distinct regions in a visual field at high levels of precision and recall values, thereby capturing good multi-scale attention dynamics. This is in relation to the monitoring of attention.

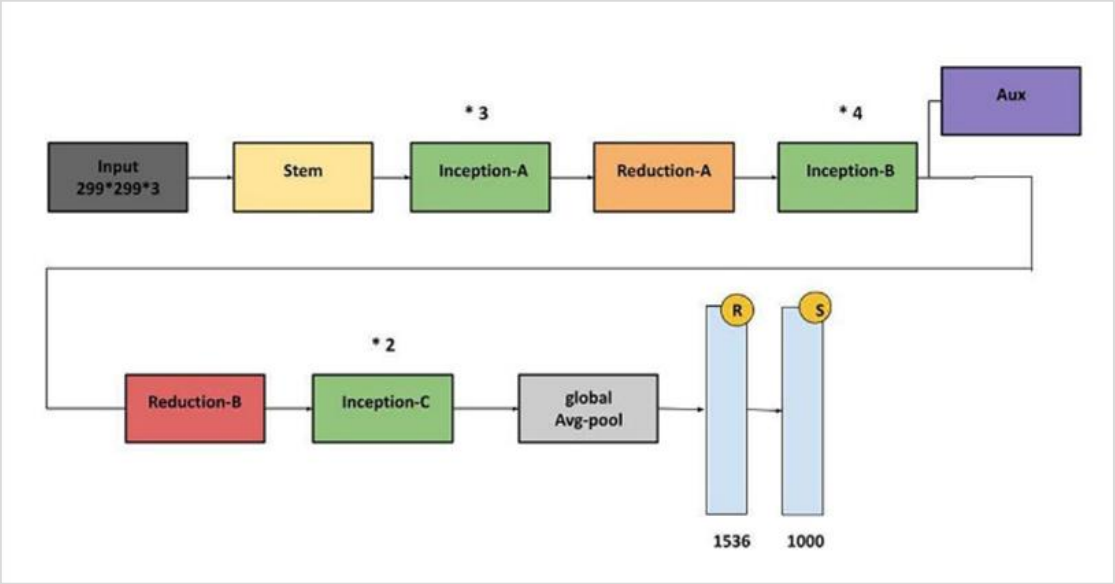

Figure 3. InceptionV3 Architecture Diagram

D. EfficientNetB0 and EfficientNetB7

By using compound scaling, the model family is able to establish a balance in terms of depth, width, and resolution. EfficientNetB0 is the name that is generally used to refer to the base model, whereas EfficientNetB7, which is one of the larger variants, provides enhanced depth and resolution. As can be seen in Figures 4 and 5, respectively, the architecture diagrams for EfficientNetB0 and EfficientNetB7 are presented.

Applications: This is just one scenario where object detection and image classification demonstrate the superior performance of EfficientNet models. Since EfficientNetB0 is a lightweight model, it may be utilized in real-time applications; nevertheless, due to its high-resolution eye gaze data, EfficientNetB7 can also be employed in sophisticated attention map applications.

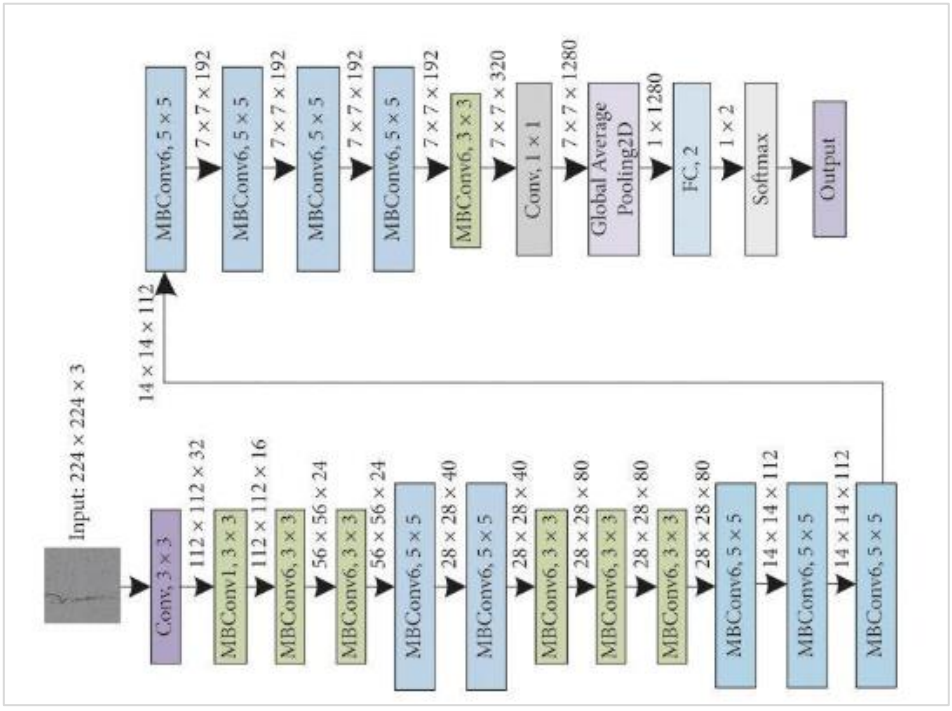

Figure 4. EfficientNetB0 Architecture Diagram

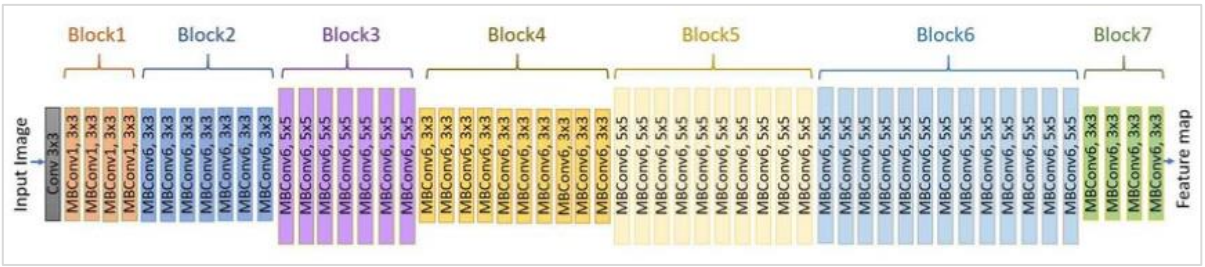

Figure 5. EfficientNetB7 Architecture Diagram

E. InceptionResNetV2

To make training very deep networks easier, the InceptionResNetV2 hybrid's inception design includes extra residual connections in the inception model. In addition, a number of convolutional layers with various filter sizes and residual blocks are employed, along with batch normalization. The architectural schematic is displayed in Figure 6 below.

Application: Beginning Applications for ResNetV2 cover a wide range of computer vision problems, such as picture categorization and object recognition. In the context of attention monitoring, it integrates concepts from ResNet and Inception to record abrupt and extended shifts in gaze focus. After being refined and integrated with their unique characteristics and structures in the "Eye Gaze for Monitoring Attention Through Hybrid Ensemble Learning," these models have the potential to develop into strong and adaptable instruments for the registration and examination of attention dynamics in a variety of settings.

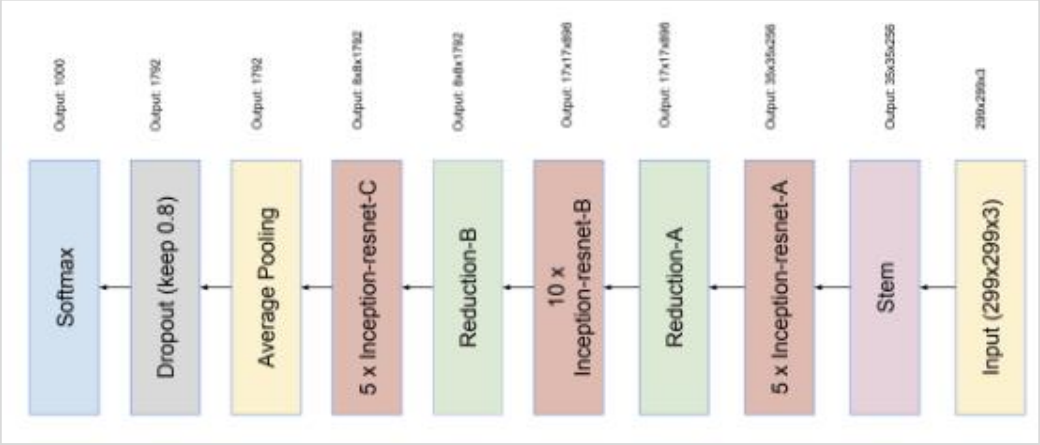

Figure 6. InceptionResNetV2 Architecture Diagram

F. XGBoost

A few potent brain network architectures, including VGG16, VGG19, InceptionV3, InceptionResNetV2, and others, are combined to create the XGBoost Cross Breed model. This cross-breeding method takes the best and most valuable aspects from several different techniques to build an extremely feasible and usable model to perform highly on all ranges of AI jobs. Prebuilt models are combined with the tendency supporting XGBoost computation to improve the model's capacity to handle complex cases, challenging highlight extraction, and precise requirements. As a result, these models are integrated in a way that creates a new group strategy that makes use of the many element representations that each component design has learned. The XGBoost hybrid model's pre-trained neural networks' learning transfer characteristics allow it to generalize effectively for a variety of data domains, including object detection and picture categorization. By combining the two models into a single, integrated framework, the XGBoost [21] Hybrid technique works to improve both models' accuracy and robustness. This allows the two models to work together in a flexible and effective solution that can thrive in a variety of modern machine learning application contexts.

4. EXPERIMENTAL FRAMEWORK

A. Dataset

The Kaggle version of the dataset has 2,936 JPEG images with a resolution of 224x224 pixels, and two class labels applied to each image: autistic and non-autistic. Renaming the files helped them become more organized. Targeting younger age groups, the dataset contains photos of children who are autistic as well as those who are not. It has pictures of boys' and girls' faces on it. Three folders containing sections for people with and without autism comprise the train, validation, and test folders containing the data. This dataset was chosen because processing images is more simpler than processing movies, and it is freely available to the public without requiring permission.

B. Methodology

The process of developing the "Eye Gaze for Monitoring Attention Through Hybrid Ensemble Learning" may be broken down into several stages, including the creation, refinement, and testing of the model. To begin, we carried out the following methods on the subsequent six pre-trained models: VGG16, Vgg19, InceptionV3, EfficientNetB0, EfficientNetB7, and InceptionResNetV2 [16]. This was

done in order to improve the accuracy of the models as well as other assessment metrics. All of the evaluation indicators have been produced as a result of the work that has been completed. In the following step, we utilized a hybrid model technique by combining this XGBoost with these six pre-trained models in an effort to find methods that are comparable to those described below. Through the utilization of this strategy, as demonstrated above, we were able to acquire all of the evaluation metrics. Consequently, the following is a generic project approach that you can use for this kind of work:

- **Data Collection**: This can be accomplished by recording eye movements in reference to the appropriate sources of eye-gazing information or by conducting controlled trials. If the data collecting is going to be done on humans, then it should be done in a responsible manner, which means adhering to regulations regarding privacy and informed permission.Be sure not to confuse complete spellings with abbreviations of units: "Wb/m2" or "webers per square meter" should be used instead of "webers/m2." Whenever units are mentioned in the text, they should be spelled out: "... a few henries" rather than "... a few H."

- **Preprocessing Data**: Cleaning and preparing the eye gaze data should include missing value treatment, noise reduction, and improvement of the consistency of the data that will be displayed. Data alignment and calibration may also be necessary for the pre-processing of the data [17].

- **Data Augmentation**: To provide more variety to the dataset, perform operations on data augmentation such as rotation, rescaling, shifting, shearing, flipping, and zooming. By adding more features, the model becomes more resilient to naturally occurring random changes and is therefore better able to generalize.

- **Model Choice**: Decide on a pre-trained deep learning model to serve as the hybrid model's foundation. As was already mentioned [18], well-known pre-trained deep learning models are VGG16, VGG19, InceptionV3, EfficientNetB0, EfficientNetB7, and InceptionResNetV2.

- **Model Architecture**: Following the selection of the pre-trained models, it would be appropriate to incorporate additional layers into your hybrid model in order to track attention. It is possible that this architecture comprises convolutional layers, recurrent layers, and fully linked layers, all of which are designed to capture attention-grabbing characteristics in the appropriate manner.

- **Training Setup:** In order to accomplish this, it may be necessary to construct the hybrid model with additional layers that include attention tracking. These layers can then be combined with the pre-trained models that have been chosen. It is possible for the architecture to incorporate convolutional, recurrent, and fully connected layers in order to accurately capture the attention features.

- **Fine-Tuning:** By combining pre-trained models of interest with new attention-tracking layers, it is possible to generate a hybrid model. This architecture takes into consideration the convolutional, recurrent, and fully connected layers in order to accommodate the attention factors. These are the most important elements that are presented for consideration.

- **Model Training:** Your hybrid model should be trained using the dataset that has been preprocessed and enhanced. Ensure that you are monitoring the KPIs of your training process, such as loss and accuracy. In order to enhance the efficiency of your training, you might want to consider employing learning rate plans or early stopping approaches.

- **Evaluation Metrics**: Evaluation of the model should be performed using a variety of evaluation metrics, such as accuracy, precision, recall, F1-score, and support, once the model has been trained [19]. In point of fact, these evaluation metrics will be of great assistance in gaining a comprehensive comprehension of the model's ability to identify and forecast attention patterns when applied.

- **Model Testing**: Putting the trained hybrid model through its paces on fresh or inexperienced data in order to evaluate its overall capabilities. This stage is essential in order to guarantee that the generality of the model would be satisfactory in the event that it was confronted with practical data.

- **Optimization and Fine-Tuning**: In order to make the model better, you should modify it based on the findings of the tests and the conclusions drawn from the evaluation. Make adjustments to a few of its settings in order to improve its performance even further.

- **Deployment:** Deploy the model to accommodate applications such as user interfaces, medical monitoring systems, instruction platforms, or anything else that requires at least equivalent performance standards to be fitted early.

C. System Architecture

There is a representation of the architecture diagram that has been utilized throughout the stages of model training and testing in Figure 7, which can be found below. Following the initial step of dividing the data into train and test datasets, the train dataset is then provided to a pre-trained model, which then extracts the features from the dataset. Therefore, these characteristics are now being directly transferred to the XGBoost model, which enhances the hybrid in terms of its resilience and efficiency.

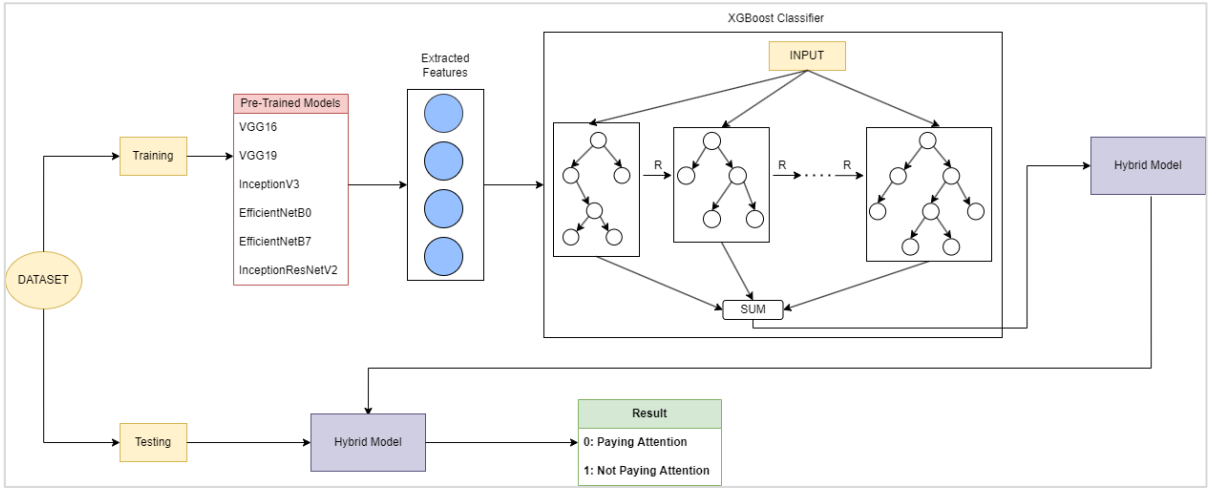

Figure 7. System Architecture

D. Performance Metrics

The following are some of the most important performance measures that may be used to evaluate the efficiency of the "Eye Gaze for Monitoring Attention Through Hybrid Ensemble Learning" system:

- Accuracy:

$$Accuracy = \frac{TP + TN}{TotalPredictions} \quad (1)$$

- Precision:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

- Recall (Sensitivity):

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

- F1-Score:

$$F1 - Score = \frac{2 * (precision * recall)}{precision + recall} \quad (4)$$

- Mean Absolute Error (MAE):

$$MAE = \frac{\sum (predicted - actual}{n} \quad (5)$$

- Root Mean Squared Error (RMSE):

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

5. RESULT

For the purpose of attention tracking, this study has conducted an empirical investigation of the performance of six different pre-trained models, namely VGG16, VGG19, InceptionV3, EfficientNetB0, EfficientNetB7, and InceptionResNetV2. Table 2 provides a summary of the accuracy levels that were reached in both testing and training.

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: Both Figure 8 and Figure 9 below illustrate the accuracy and plot graphs for the best pre-trained models. These graphs can be found below.

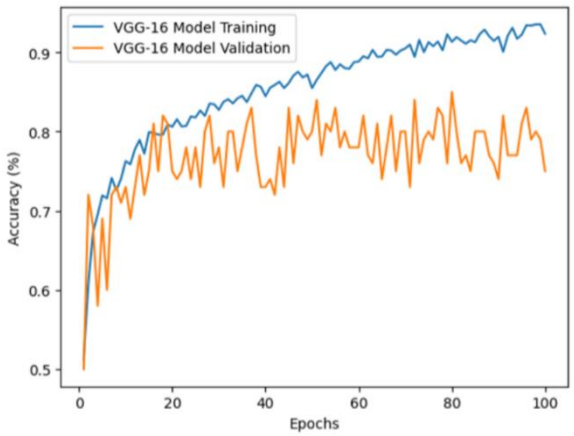

Figure 8. VGG16 Accuracy Plot

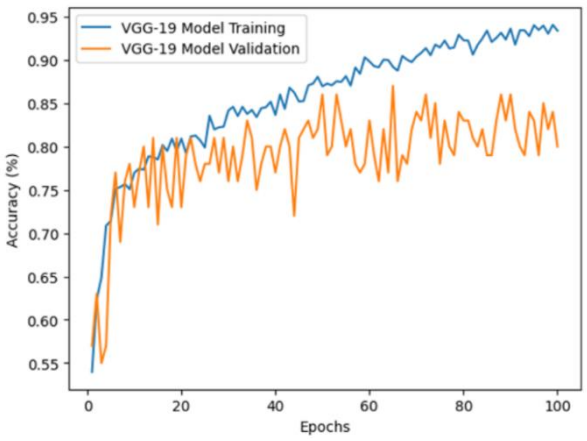

Figure 9. VGG19 Accuracy Plot

- Confusion Matrices: Additionally, the confusion matrices for the most effective pre-trained models are presented in Figures 10 and 11, respectively.

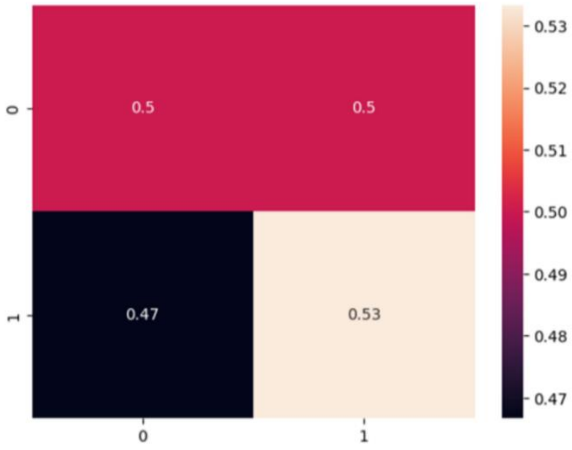

Figure 10. VGG16 Confusion Matrix

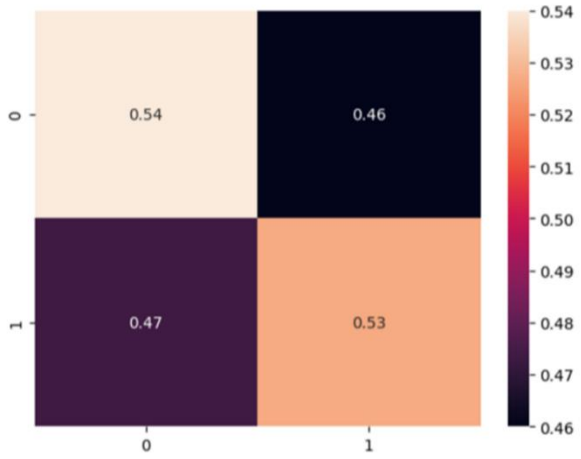

Figure 11. VGG19 Confusion Matrix

- Evaluation Results: All of the evaluation findings, including precision, recall, f1-score, and support for pre-trained models, are presented in the following table, which may be found in reference number 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

When everything was said and done, XGBoost was used to hybridize all of the pre-trained models. Unbelievably, each hybrid model demonstrated high learning, reaching as high as one hundred percent in terms of training accuracy.

Table 4 presents a comparison of the testing accuracies of the models with the various displays that are available.

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: The confusion matrices from the hybrid models that performed the best are presented in Figure 12 and Figure 13, respectively, after these two graphs have been presented.

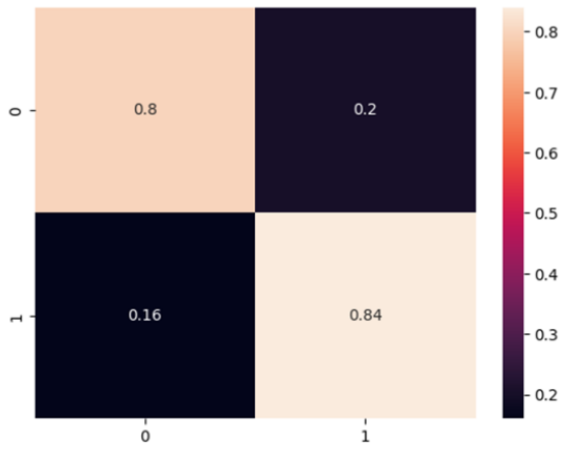

Figure 12. VGG16-XGBoost Confusion Matrix

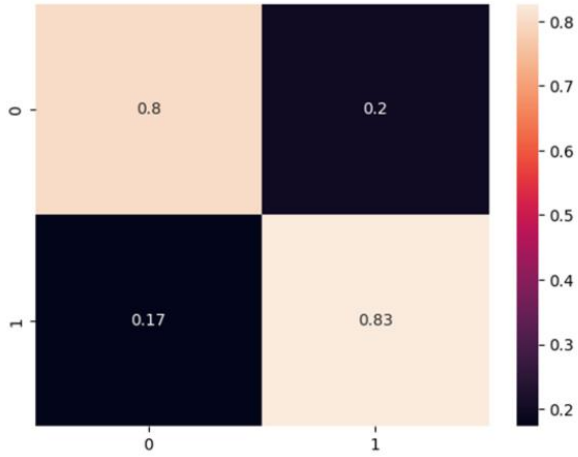

Figure 13. VGG19-XGBoost Confusion Matrix

- Evaluation Results: The remaining components, including precision, recall, f1-score, and support, are also included in Table 5, which also contains these additional components.

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

In addition, the VGG16-XGBoost and VGG19-XGBoost models fared better than other hybrid models when it came to the monitoring attention test. As a result of their testing, they achieved accuracy levels that were at least higher than those of other hybrid models, which indicates that the models operate better than other hybrid models [20].

6. Discussion and Future Work

These results from the pre-trained and hybrid model studies were further supported by variations in the performance of the individual pre-trained models and by additional research employing different assessment measures. Overall, the work indicates that VGG16-XGBoost and VGG19-XGBoost have the best performance and validates that XGBoost may be integrated with pre-trained models for attention monitoring.

A few important considerations for suggesting the direction of this endeavor are listed below.

- **Multi-modal data integration:** combining physiological data sources such as electroencephalography (EEG), heart rate, and facial expressions with eye-gaze data to provide a more complete attention monitoring model. In terms of how someone is paying attention, multi-modal integration is far more revealing.

- **Real-Time Monitoring and Feedback:** Give clients immediate solutions to assist them in managing their attention. These kinds of technologies have potential applications in fields like education, where personalized attentional feedback improves learning chances.

- **Collaboration in Neuroscience**: In order to improve neuroscience research by gaining a deeper comprehension of the brain's attention-related processes and how they connect to gaze patterns, collaboration with neuroscientists will be encouraged.

7. Conclusion

Eye Gaze for Monitoring Attention through Hybrid Ensemble Learning, a state-of-the-art multidisciplinary technique, combines deep learning, eye gaze monitoring, and data augmentation to identify, comprehend, and forecast attention patterns in a variety of contexts. Applications for this new paradigm could be found in a wide range of fields, including education, healthcare diagnostics, and improving human-computer interface for better attention tracking. The model makes use of pre-trained architectures that provide the benefits of multi-scale feature extraction and efficient network designs, such as VGG16, VGG19, InceptionV3, EfficientNetB0, EfficientNetB7, and InceptionResNetV2. By using data augmentation techniques like rotation, shifting, and rescaling, the model is able to provide more lifelike representations of items that it was not trained on, increasing the system's resilience and adaptability to real-world situations.

References

[1] I. 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] A. 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.

[21] Khetani, Vinit, et al. "Cross-domain analysis of ML and DL: evaluating their impact in diverse domains." International Journal of Intelligent Systems and Applications in Engineering 11.7s (2023): 253-262.

[22] Mane, D. (2024). Nonlinear Analysis in Skin Cancer Detection: Customized Convolutional Neural Networks Approach. Communications on Applied Nonlinear Analysis, 31(2s), 320-338