FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 9 No. 4, April, 2025, pp 252 – 261
DOI: https://doi.org/10.33003/fjs-2025-0904-3487

ENHANCING AGE ESTIMATION FROM SCLERA IMAGES USING RESNET-50, VGG16, AND RANDOM FOREST

***[1]Aliyu Sulaiman Mukhtar, [2]Muhammad Aminu Ahmad, [2]Mohammad Ibrahim, [1]Saadatu Abdulkadir, [1]Abubakar Mu'azu, [1]Sulaiman Dauda and [1]Abdullahi Diso**

[1]Department of Informatics, Kaduna State University, Kaduna – Nigeria.
[2]Department of Secure Computing, Kaduna State University, Kaduna – Nigeria.

*Corresponding authors' email: aliyusmukhtar@kasu.edu.ng

ABSTRACT
This study presents a novel hybrid model for age prediction from sclera images, combining deep learning architectures ResNet-50 and VGG-16 with a Random Forest classifier. The hybrid approach aims to optimize both accuracy and computational efficiency, addressing limitations in previous methodologies. Results demonstrate exceptional performance, with the hybrid model achieving an overall accuracy of 98.85% and outperforming benchmark models. Detailed evaluation metrics reveal high precision, recall, and F1-scores across age groups, supported by insights from the confusion matrix. The model's practical applicability is demonstrated through efficient training and testing processes. This research bridges gaps in existing literature by integrating transfer learning, deep learning, and ensemble methods, while also addressing issues of computational complexity. The study underscores the potential of hybrid models to advance age prediction from biometric images, setting a new benchmark for future research in the field.

Keywords: Deep learning, Deep Neural Network, Convolutional Neural Networks, VGG16, Resnet 50, Random Forest, Machine learning

INTRODUCTION
The distribution of age brackets holds significant importance for organizations and nations, playing a crucial role in decision-making related to facility allocations, access restrictions, and developmental initiatives (Olorunsola & Olorunshola, 2023). Addressing the challenge of age falsification is particularly critical in sports, organizational contexts, and electoral processes, necessitating the implementation of an age detection system to safeguard the accuracy and integrity of available information (Galbally et al., 2019). Traditional age estimation methods often rely on physical features such as facial characteristics, teeth, bones, and other supportive elements (Kumar et al., 2022). Despite these approaches, limited research has been conducted on age prediction utilizing sclera.

Sclera recognition emerges as a promising ocular biometric modality due to contact-less, gaze-independent image acquisition in visible light (Das et al., 2021). Moreover, it remains unaffected even when subjects wear contact lenses in their eyes. However, implementing sclera recognition poses challenges as accurate and efficient execution of several steps is required. As a human characteristic, age plays a significant role in facilitating or limiting communication, acting as a barrier that influences interaction and comprehension, akin to language, culture, beliefs, and experience. Aging leads to changes in facial features, skin thickness, color, and texture. Recognizing and interpreting facial traits such as age, emotion, or gender in real time gave rise to the concept of computer vision (Othmani et al., 2020).

Every individual possesses distinct biometric traits used in identification, recognition, verification, age estimation, and authentication due to their uniqueness (Odion et al., 2022). Biometric technologies such as face recognition, ear recognition, voice recognition, iris recognition, retina scan, finger recognition, vein recognition, and DNA matching have become hot topics in academia. Facial recognition, with its ability to reveal features such as age, gender, race, emotion, expression, and identity, stands out as the most prominent area of human recognition (Narejo et al., 2016). Biometric recognition systems based on ocular traits are essential, offering contactless data acquisition, high recognition accuracy, and considerable user acceptance (Vitek et al., 2020).

Machine learning (ML) algorithms play a pivotal role in addressing challenges across various domains by analyzing and comprehending vast amounts of data (Oyewola et al., 2021). Particularly in computer vision, ML has proven highly advantageous in detecting and identifying facial images. ML's ability to learn complex patterns from large datasets significantly enhances the accuracy and efficiency of facial recognition systems. Additionally, machine learning has made substantial contributions to age estimation, as emphasized by (Umarani et al. 2021), who highlight the utility of ML algorithms in predicting and estimating the age of individuals based on various features, such as facial characteristics.

The integration of machine learning in these domains not only improves the speed and accuracy of processes but also opens up new possibilities for applications like biometric security, age-specific content recommendation systems, and demographic analysis. As technology continues to advance, machine learning is poised to play an increasingly significant role in addressing complex challenges and making informed decisions across diverse fields.

A novel enhancement approach for age prediction is introduced, emphasizing the feasibility of sclera-based age prediction. The method leverages the consensus of two high-performing transfer learning algorithms highlighted by Odion et al. (2023) and builds upon their work from 2022. The approach exclusively focuses on the feature extractor modules of both transfer learning algorithms, utilizes a random forest for classification, and combines results through a voting mechanism.

The inception of research on age estimation dates back to 1994 with Kwon and da Vitoria Lobo, who initially divided age into distinct ranges (Abbasi Khan, 2016). Numerous subsequent studies have predominantly focused on age estimation through facial images, with limited attention given

to eye regions (Beattie et al., 2011). Notably, there is a noteworthy body of research utilizing the iris as a modality for age estimation (Russell et al., 2014).
The authors (Sgroi et al., 2013) employed a random forest algorithm on 596 iris images, classifying 300 as young subjects and 296 as elderly subjects. This approach achieved a classification rate of 64.68%. Another innovative method proposed by (Erbilek et al., 2013) relied on five geometric features extracted from the human iris. They gathered data from 210 respondents spanning ages 18 to 73, categorizing them as young, adult, and senior. Leveraging segmentation for efficient detection of iris and pupil boundaries, the authors achieved a 75% accuracy rate. (Aminu Bashir Suleiman, 2023) reveals how machine learning algorithms, with changeable parameters and models, can be used to predict cardiac disease with simplicity. Prediction, problem solving, and other domains are where machine learning is quite helpful. Additionally, machine learning is a useful tool for solving issues in a variety of fields. Their model used four machine learning algorithms to achieve very good results, Logistic Regression 84% Naïve Bayes Classifier 80% K Nearest Neighbors Classifier 87% Decision Tree Classifier 79% Support Vector Classifier 83% Random Forest Classifier 90%. (Rajput and Sable, 2019) in their research involved deep CNN models like AlexNet and GoogleNet, trained on a real-time database spanning age 3-73. The models demonstrated notable accuracy, with AlexNet outperforming GoogleNet in gender classification and achieving an overall accuracy level of 95.34%. In a similar vein, (Rajput et al., 2020) suggested a technique utilizing iris structure to estimate a person's age group. Their approach involved the circular Hough Transform for iris boundary localization and employed five different classifiers. The bagged ensemble classifier outperformed others, yielding an overall accuracy of 83.7%, surpassing prior state-of-the-art models.
Recent progress in the field has witnessed the incorporation of machine learning methodologies into eye region biometrics for age prediction, exemplified by the research conducted by Irhebhude et al. (2023). Their study introduces a model for age and gender classification based on iris images obtained from both male and female subjects aged between 5 and 60 years. The age and gender categories were delineated as FemaleYoung, FemaleTeen, FemaleAdult, MaleYoung, MaleTeen, and MaleAdult. In the analysis of these six categories, feature extraction was carried out using a 3D histogram, and Principal Component Analysis (PCA) was subsequently applied for dimension reduction to mitigate duplicated or extraneous features. Following this, the Support Vector Machine (SVM) was employed for the classification of iris images into the six defined groups, yielding diverse performance metrics. The 3D histogram in conjunction with PCA demonstrated an outstanding classification performance accuracy of 99.27%. In contrast, the EfficientNet deep learning model achieved a comparatively lower accuracy of 52.29%. Irhebhude et al's work showcases the efficacy of their proposed methodology in achieving robust age and gender classification results, emphasizing the superiority of their approach over the benchmark EfficientNet model in this context.
Furthermore, in the year 2022 Odion et al. (2022) conducted a comprehensive investigation into the effectiveness of transfer learning for age prediction, specifically focusing on age group classification using sclera images. In their study, a curated dataset comprising 2000 sclera images from 250 individuals spanning various age groups was employed. The segmentation of images was achieved through the application of Otsu thresholding, utilizing morphological processes. The experimentation phase involved the utilization of four distinct pre-trained models, namely VGG16, ResNet50, MobileNetV2, and EfficientNet-B1. These models were meticulously assessed based on diverse performance metrics. Among the models, ResNet-50 emerged as the top-performing algorithm, showcasing exceptional accuracy, precision, recall, and F1-score metrics, all at an impressive 95%. In comparison, VGG-16, EfficientNet-B1, and MobileNetV2 demonstrated commendable performances with accuracy rates of 94%, 93%, and 91%, respectively.
The research aims to develop an optimized and efficient sclera-based age prediction model by integrating Transfer Learning, Feature Extraction, and Ensemble Classification. This approach seeks to enhance accuracy while reducing computational complexity, addressing the limitations observed in previous studies, such as those by Odion et al. (2022). By leveraging the strengths of multiple deep learning architectures and refining feature extraction techniques, the study aims to create a highly accurate, scalable, and resource-efficient age prediction system suitable for real-world applications.

MATERIALS AND METHODS

The articulated material and methods used were crafted with a specific focus on mitigating challenges related to prediction accuracy and computational complexity. An imperative goal of this methodology was to circumvent potential issues arising from employing complete transfer learning models, which, if left unaddressed, could adversely impact the learning rate. By addressing these concerns through a carefully devised approach, this research methodology aspires to pave the way for more accurate age predictions from sclera images while managing computational complexities effectively. The methodology encompassed distinct stages, encompassing data splitting, model training, classification, and evaluation. This multifaceted methodology comprised several pivotal steps, elucidated in Figure 1 for clarity.

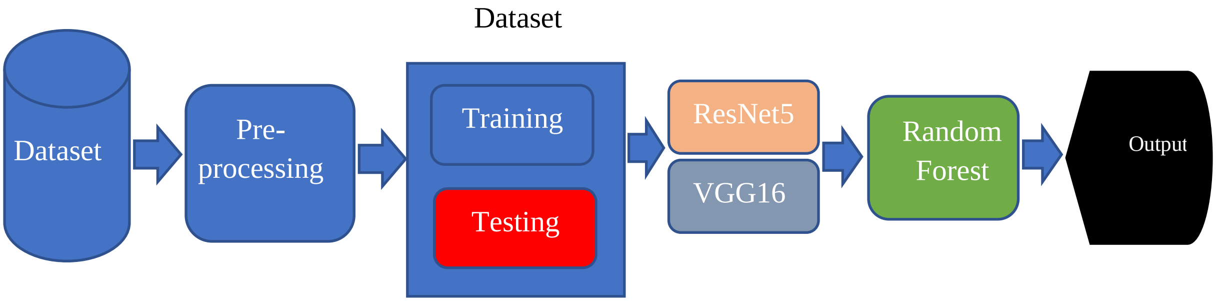

Figure 1: Research Methodology Flow

ResNet50

ResNet-50, an influential convolutional neural network (CNN), has played a transformative role in reshaping the landscape of deep learning since its introduction by Kaiming He and his team at Microsoft Research Asia in 2015. The nomenclature "ResNet" is derived from "residual network," a term that encapsulates the fundamental architecture of the network.

At its core, ResNet-50 is grounded in a deep residual learning framework, a paradigm that facilitates the training of networks with an unprecedented depth, often comprising hundreds of layers. This approach addressed a perplexing observation in deep learning research – the counterintuitive notion that adding more layers to a neural network did not consistently lead to improved results. While the addition of layers theoretically should enable the network to learn not only what the preceding layers learned but also additional information, this was not consistently the case.

In response to this challenge, the ResNet team, led by Kaiming He, introduced a novel architecture featuring what are known as skip connections. These skip connections, also referred to as residual blocks, serve a crucial role in preserving information from earlier layers during the forward pass. This innovation effectively mitigated the vanishing gradient problem and allowed for the training of very deep networks.

The incorporation of skip connections marked a departure from traditional deep learning architectures and had a profound impact on the field. It enabled the training of ResNet networks with an exceptional number of layers, with ResNet-50 specifically designed with 50 layers. The architecture's ability to retain and reuse information from earlier layers during training proved to be a pivotal breakthrough.

The success of ResNet-50 was exemplified by its performance on benchmark datasets. In particular, it achieved a remarkable 3.57% error rate on the ImageNet dataset, a widely recognized standard for evaluating image classification models. Additionally, ResNet-50 emerged victorious in several prestigious competitions, including the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) and the Common Objects in Context (COCO) object detection challenges.

The ResNet architecture demonstrated that the strategic use of skip connections could lead to more effective training of deep networks, paving the way for subsequent advancements in deep learning research. The profound impact of ResNet-50 has extended beyond these specific competitions, influencing the design of subsequent state-of-the-art architectures and contributing to the broader understanding of how to effectively train and optimize deep neural networks.

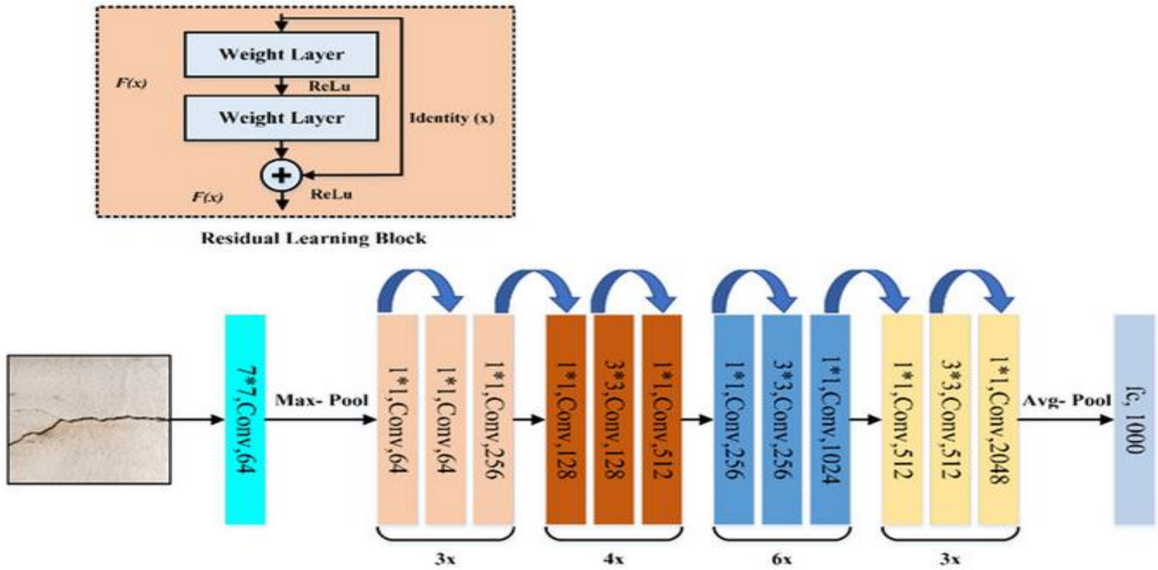

Figure 2: Original ResNet50 Architecture

The architecture, comprising a total of 50 layers, ResNet-50 is organized into five distinct blocks, each housing a collection of residual blocks. These specialized residual blocks play a pivotal role in maintaining information from preceding layers, facilitating the network's ability to acquire enhanced representations of the input data. ResNET encompasses several integral components that contribute to its architecture and functionality.

Convolutional Layers

The initial layer of ResNET comprises a convolutional layer responsible for performing convolution on the input image. Subsequently, a max-pooling layer follows, down sampling the output from the convolutional layer. This downsized output is then directed through a sequence of residual blocks.

Residual Blocks

Each residual block within ResNET is composed of two convolutional layers. These layers are accompanied by a batch normalization layer and a rectified linear unit (ReLU) activation function. The output of the second convolutional layer undergoes addition with the input of the residual block, followed by another ReLU activation function. The resultant output proceeds to the subsequent block in the network (He et al., 2016).

Fully Connected Layer

The concluding layer of the network is a fully connected layer. This layer takes the output of the last residual block and maps it to the corresponding output classes. The number of neurons in the fully connected layer aligns with the count of output classes (Awujoola et al., 2024).

Concept of Skip Connection

An integral feature of ResNET-50 is the concept of skip connections, also referred to as identity connections. These connections play a pivotal role in retaining information from earlier layers, thereby enhancing the network's capability to glean improved representations of the input data. The implementation of skip connections involves adding the output of an earlier layer to the output of a subsequent layer (He et al., 2016) Figure 3 depict the skip connection.

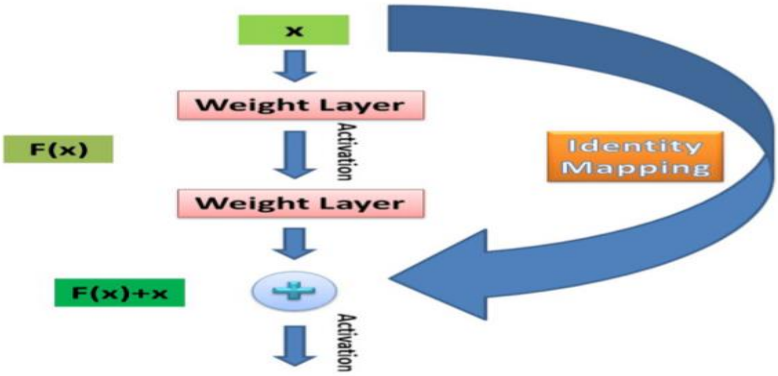

Figure 3: ResNet50 Skip Connection

VGG16

VGG16 stands as a convolutional neural network model pioneered by the Visual Geometry Group (VGG) at the University of Oxford, securing victory in the 2014 ILSVRC object identification algorithm competition. A fundamental contribution of VGG16 is its illustration of how expanding the network's depth can enhance performance under specific conditions. Diverging from the classic AlexNet, VGG16's innovation lies in replacing larger convolution cores (11×11, 7×7, 5×5) with multiple 3×3 convolution cores. This modification not only effectively deepens the network but also reduces the number of parameters, optimizing overall performance. The VGG16 model is characterized by 13 convolutional layers, three fully connected layers, and five pooling layers.

VGG16 upholds the simplicity inherent in classical network structures while augmenting depth through the versatile application of 3×3 convolutions, resulting in improved performance. However, the model exhibits certain limitations in practical applications (Yang et al., 2021). Firstly, the fully connected layer entails a substantial number of parameters, consuming significant memory and computational resources, presenting challenges for front-end deployment. Secondly, the singular structure of the network renders it less robust compared to more sophisticated counterparts. Additionally, VGG16 lacks an effective mechanism to counteract gradient disappearance, potentially leading to issues like slow convergence and gradient explosion during model training. Figure 4 shows the Vgg16 architecture.

Figure 4: VGG16 Architecture

The presented model is a sophisticated approach to predicting age groups based on sclera images, harnessing advanced techniques in deep learning and machine learning. Its primary objective is to predict the age range of individuals by leveraging a combination of transfer learning, feature extraction, and ensemble classification.

The initial stages of the model involve the integration of two powerful pre-trained deep learning models: ResNet50 and VGG16. These models, initially trained on the ImageNet dataset, exhibit exceptional capabilities in capturing intricate patterns and features from images. The top classification layers of both models are excluded, emphasizing feature extraction rather than class-specific information.

To preprocess the sclera images for these models, a meticulous image preprocessing function is devised. This function ensures that each image is appropriately loaded, resized to a consistent size of 224x224 pixels, and normalized to match the requirements of the respective deep learning models.

The core of the model lies in the feature extraction process, where sclera images are passed through the pre-trained ResNet50 and VGG16 models. The extracted features are then flattened to create comprehensive feature vectors. This dual-model approach enriches the representation of sclera characteristics, contributing to the model's ability to discern subtle age-related variations.

The model is trained on a diverse dataset of sclera images, carefully organized into subfolders corresponding to distinct age groups. The true labels, representing different age ranges, are collected during the training phase, creating a robust foundation for supervised learning. The features obtained from ResNet50 and VGG16 are concatenated to form a combined feature set, enhancing the richness and diversity of the input data.

To evaluate the model's performance, the dataset is split into training and testing sets. A Random Forest classifier, a robust and versatile ensemble learning algorithm, is employed for the final age prediction. This classifier is trained on the extracted features, and its predictions are compared against the true age group labels during the testing phase.

The evaluation metrics are comprehensive, including a detailed classification report that provides insights into precision, recall, and F1-score for each age group. The confusion matrix further visualizes the model's performance, offering a clear depiction of its ability to correctly predict age groups and identify potential misclassifications.

In conclusion, this age prediction model represents a synergistic integration of deep learning and machine learning techniques tailored for sclera images. By combining the strengths of transfer learning, feature extraction, and classification, the model showcases a robust framework for accurate age prediction. Its versatility and potential applications span various domains, including biometric recognition, demographic analysis, and healthcare. As technology continues to advance, models like these pave the way for nuanced and accurate predictions in the realm of age estimation using sclera images.

Dataset

To foster a thorough comparison between our age prediction model and the research conducted by Odion et al. (2018), we endeavored to obtain access to the dataset employed in their study. In our pursuit of a rigorous evaluation, not only did we procure the original dataset, but we also undertook measures to enhance it, making it particularly amenable for deep learning applications.

The dataset, as initially provided by Odion et al. (2018), comprises sclera images labeled with respective age groups, specifically categorized into Kids (5 - 13 years), Teens (14 - 20 years), and Adults (21 - 30 years). This dataset distribution delineates the number of subjects and images allocated to each age group.

Within the Kids age group, spanning 5 to 13 years, there are 84 subjects with a total of 672 images. The Teens age group, covering ages 14 to 20, consists of 85 subjects and 680 images. Lastly, the Adults age group, encompassing ages 21 to 30, incorporates 80 subjects and 640 images.

In aggregate, the dataset encapsulates a total of 250 subjects and 2000 images, systematically distributed across the specified age ranges. This strategic allocation ensures a balanced representation across different stages of human development, facilitating the model's capacity to comprehend and predict age across the entire spectrum.

The dataset enhancement process involved augmenting the existing data to align with the requisites of deep learning methodologies. This augmentation not only increased the volume of the dataset but also introduced variability and diversity in the sclera images. Such measures are instrumental in training a robust model that can effectively discern age-related features across a wide range of subjects.

Through meticulous curation and augmentation, our objective is to furnish the model with a comprehensive dataset that captures the intricacies of age-related changes in sclera images. This enriched dataset serves as a foundational element, integral to the training and evaluation phases of our age prediction model. It is designed to fortify the model's proficiency in delivering accurate and reliable predictions, spanning a diverse and representative set of age demographics.

Model Performance Evaluation

The experimental comparison of classification algorithms was done based on the performance measures of accuracy, specificity, sensitivity, and error rate; the model was evaluated based on the following metrics:

Confusion Matrix

A confusion matrix is a table that summarizes the performance of a classification model by comparing its predicted labels with the actual ground truth labels (Awujoola et al, 2023). The table is organized in a matrix format, with rows representing the true labels and columns representing the predicted labels.

For each combination of true and predicted labels, the table contains the count or frequency of instances falling into that category. The diagonal elements of the matrix represent the correctly classified instances, while the off-diagonal elements represent the misclassified instances. The table 3.3 shows the confusion matrix for a two-class model Awujoola et al, (2023)

Table 1: Confusion Matrix for two class classifiers

	Predicted Positive	**Predicted Negative**
Actual Positive	TP	FN
Actual Negative	FP	TN

TP represents the instances correctly predicted as positive, FN represents the instances incorrectly predicted as negative, FP represents the instances incorrectly predicted as positive, and TN represents the instances correctly predicted as negative.

By analyzing the values in the confusion matrix, we can compute various evaluation metrics such as accuracy, precision, recall, and F1 score, which provide insights into the model's performance and its ability to correctly classify instances belonging to different classes.

Accuracy: It is the percentage of accurate predictions i.e the ratio of number of correctly classified instances to the total number of instances and it can be defined as: (Santamaria et al., 2018).

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

where TP- True Positive, FP- False Positive, TN- True Negative, FN- False Negative

False Positive rate (FPR): This measures the rate of wrongly classified instances. A low FP-rate signifies that the classifier is a good one (Awujoola et al, 2023).

$$FPR = \frac{FP}{FP + TN} \quad (2)$$

Sensitivity: It is the proportion of positives that are correctly identified (Gad, 2021).

Sensitivity = TP / TP + FN (3)

Precision. Precision is the ratio of positively predicted instances among the retrieved instances (Awujoola et al, 2023).

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

Specificity: It is the proportion of negatives that are correctly identified. It is calculated as the number of correct negative predictions divided by the total number of negatives. It is also called true negative rate. The worst is 0.0 while the best is 1.0 (Awujoola et al, 2023).

Specificity = TN / TN + FP (5)

Recall is the ratio of positively predicted instances among all the instances (Awujoola et al, 2023).

$$Recall = \frac{TP}{TP + FP} \quad (6)$$

Error Rate: It is equivalent to 1 minus Accuracy.

RESULTS AND DISCUSSION

We present the consolidated findings from the evaluation of the hybrid model developed to improve age prediction from sclera images. The analysis provides a detailed examination of the model's performance metrics, comprehensively shown in Tables 2, Figures 5, and 6. These metrics include classification reports, confusion matrices, and ROC curves, offering a thorough insight into the model's efficacy.

Table 2: Classification Report obtained from the Evaluation of the Developed model

Class	Precision	Recall	F1-Score	Support	Training Time (s)	Testing Time (s)
Adult	0.99	0.99	0.99	346	350.37	43.83
Kids	0.99	0.97	0.98	311		
Teen	0.97	0.99	0.98	303		
Accuracy			0.99	960		
Macro Avg	0.99	0.99	0.99	960		
Weighted Avg	0.99	0.99	0.99	960		

The classification report obtained from the evaluation of the hybrid model, which integrates ResNet50 and VGG16 for feature extraction and employs a Random Forest classifier for age prediction from sclera images, demonstrates an exemplary performance across various metrics. The analysis, as detailed in Table 4.1, provides a thorough insight into the model's effectiveness in classifying the images into three distinct age categories: Adult, Kids, and Teen.

For the Adult class, the model achieved a precision, recall, and F1-score of 0.99, with a support of 346 images. This high precision indicates that the model is exceptionally accurate in identifying adult sclera images, while the equally high recall suggests that it successfully identifies nearly all actual adult images in the dataset. The F1-score of 0.99 confirms a perfect balance between precision and recall, underscoring the model's robust performance for this class.

In the case of the Kids class, the model attained a precision of 0.99, a recall of 0.97, and an F1-score of 0.98, with a support of 311 images. Although the recall is slightly lower compared to the Adult class, the precision remains high, indicating that the majority of the predicted kids' images are correct. The F1-score of 0.98 still reflects a strong performance, demonstrating the model's capability to accurately classify sclera images of children.

For the Teen class, the results show a precision of 0.97, a recall of 0.99, and an F1-score of 0.98, with 303 images in support. The high recall value signifies that the model effectively identifies nearly all actual teen images, while the precision value indicates a slightly higher rate of false positives compared to the other classes. Nevertheless, the F1-score of 0.98 suggests that the model maintains a high level of accuracy and reliability for this age group.

The overall accuracy of the model stands at 98.85%, calculated over 960 images. This high accuracy value reflects the model's overall capability to correctly classify the majority of sclera images into the correct age categories. The macro average, which averages the precision, recall, and F1-score for all classes, is 0.99 across all metrics, indicating a consistent performance regardless of class size or distribution. Similarly, the weighted average, which accounts for the number of images in each class, also stands at 0.99, reaffirming the model's balanced and robust performance.

The model's training and testing times were also evaluated, with the training process taking approximately 350.37 seconds and the testing phase taking about 43.83 seconds. These times indicate the computational efficiency of the hybrid approach, showcasing its practical applicability in real-world scenarios where timely processing is crucial.

In summary, the hybrid model's evaluation reveals its outstanding capability in accurately predicting age from sclera images. The high precision, recall, and F1-scores across all classes, combined with the overall accuracy of 0.99, demonstrate the effectiveness of integrating ResNet50 and VGG16 for feature extraction with a Random Forest classifier. The model not only delivers high performance but also operates efficiently within reasonable computational timeframes, making it a valuable tool for age prediction in sclera images.

Confusion Matrix

This section examines the confusion matrix resulting from the evaluation of the hybrid model, which employs ResNet50 and VGG16 for feature extraction, fuses all the extracted features, and utilizes a Random Forest for classification. The confusion matrices provide a visual representation of the model's performance, highlighting true positives, true negatives, false positives, and false negatives for each age group. This visualization helps identify specific areas where the model may be misclassifying certain age groups, offering actionable insights for further refinement of the model.

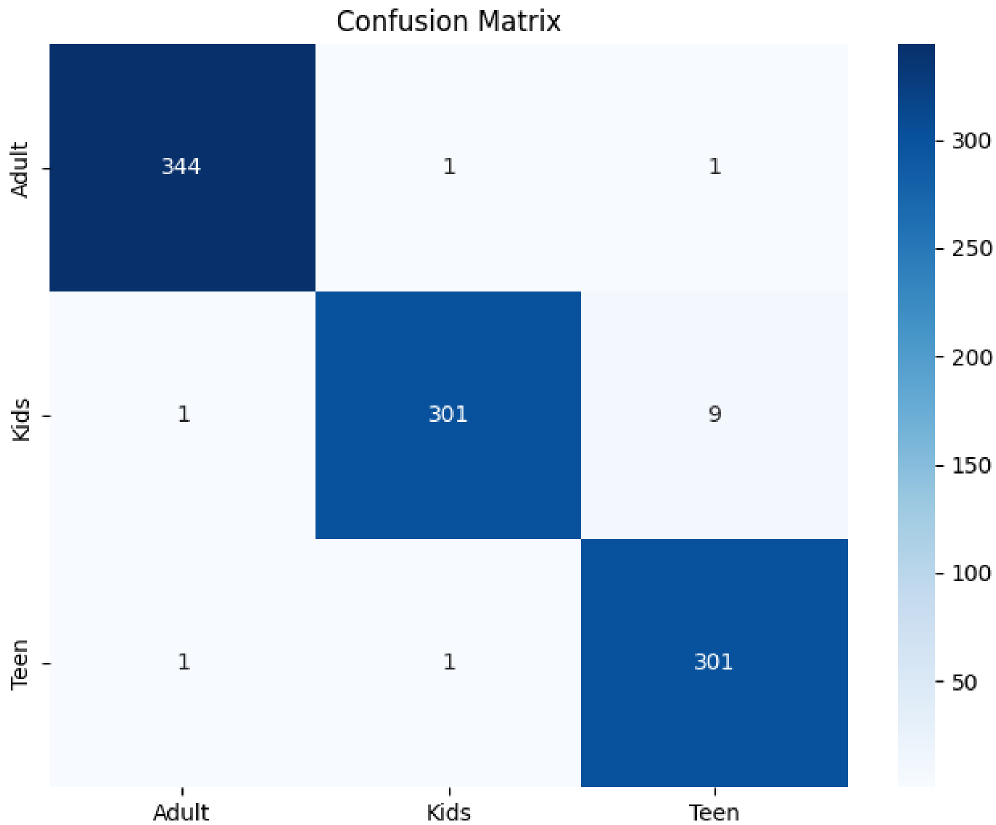

Figure 5: The Confusion Matrix

The confusion matrix obtained from the evaluation of the hybrid model as visualized in Figure 5, provides valuable insights into the model's performance in classifying sclera images into different age groups. This visual representation reveals the distribution of true positive, true negative, false positive, and false negative predictions across the Adult, Kids, and Teen age categories.

In the Adult category, the model correctly predicted 344 images, indicating a high level of accuracy in identifying adult sclera images. However, there were also 2 misclassifications, with 1 image being incorrectly classified as a Kid and another as a Teen. Despite these minor errors, the overall performance in this category remains robust.

For the Kids class, the model achieved a high true positive count of 301, indicating successful classification of Kids' sclera images. However, there were 10 misclassifications, with 9 images incorrectly classified as Teen and 1 as Adult. While the model demonstrates strong performance overall, there is a slight tendency to misclassify Kids' images as belonging to the Teen category.

Similarly, in the Teen category, the model accurately predicted 301 images, reflecting its effectiveness in identifying Teen sclera images. However, there were 2 misclassifications, with 1 image each being incorrectly classified as Adult and Kid. Despite these errors, the model maintains a high level of accuracy in classifying Teen images.

The confusion matrix illustrates the model's ability to distinguish between different age groups with a high degree of accuracy. However, it also highlights areas where the model may encounter challenges, such as distinguishing between Kids and Teens, leading to occasional misclassifications. These insights provide valuable guidance for further refining the model to improve its performance and accuracy in age prediction from sclera images.

Receiver Operating Character

This Section delves into the Receiver Operating Characteristic (ROC) analysis, a crucial aspect of evaluating the performance of classification models. The ROC curve illustrates the trade-off between the true positive rate (sensitivity) and the false positive rate (1 - specificity) across various threshold settings. By plotting these rates against each other, the ROC curve provides valuable insights into the model's ability to discriminate between different classes. A steeper curve closer to the top-left corner signifies better performance, with an area under the ROC curve (AUC) close to 1 indicating high discriminative ability. This section likely discusses how the ROC analysis enhances the understanding of the hybrid model's performance in age prediction from sclera images, providing a comprehensive evaluation beyond traditional accuracy metrics.

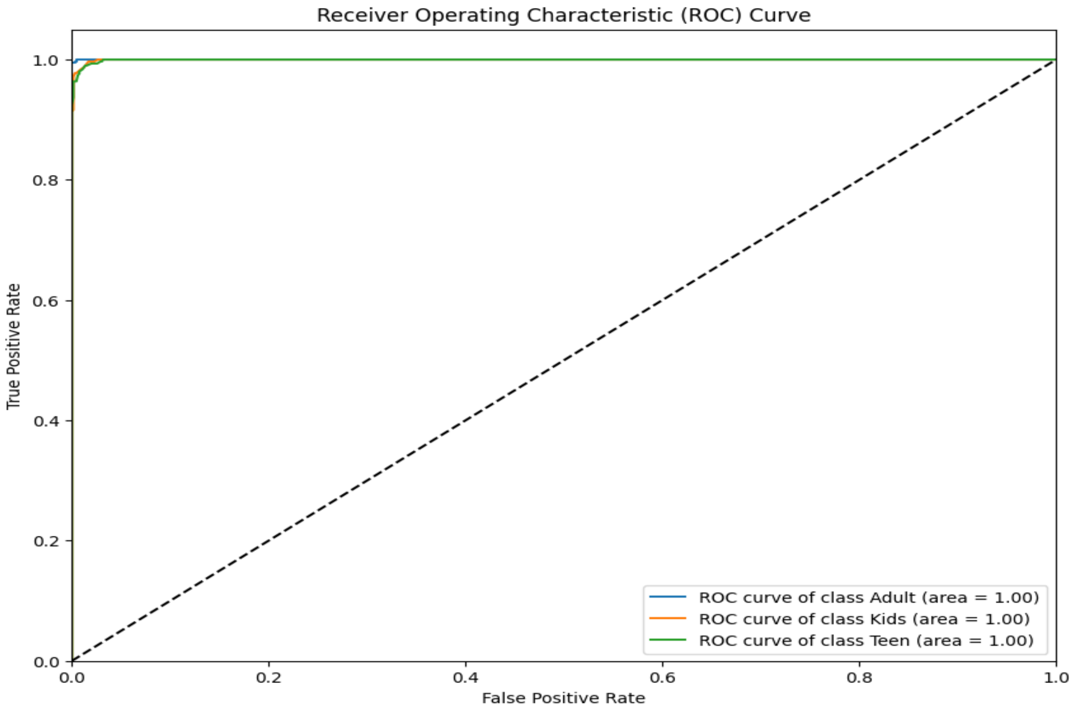

Figure 6: ROC curve

The Receiver Operating Characteristic (ROC) curve is a graphical plot used to show the diagnostic ability of a binary classifier system as its discrimination threshold is varied. This particular ROC curve plot as represented in Figure 6 presents the performance of a multi-class classification model, which classifies data into three distinct categories: Adult, Kids, and Teen. The ROC curve is a plot of the True Positive Rate (TPR) against the False Positive Rate (FPR) for various threshold settings. The TPR (sensitivity or recall) is plotted on the Y-axis, representing the proportion of actual positives correctly identified by the model. The FPR is plotted on the X-axis, representing the proportion of actual negatives that were incorrectly identified as positive.

The curve for the Adult class is shown in blue, and the area under the ROC curve (AUC) is 1.00, indicating perfect classification performance for this class. The curve for the Kids class is depicted in orange, and this curve also has an AUC of 1.00, demonstrating that the classifier distinguishes this class without any errors. The curve for the Teen class is represented by the green line, and similarly, the AUC for this class is 1.00, indicating flawless classification.

The diagonal dashed line represents the performance of a random classifier, which would produce points along this line if it were to make random guesses. The closer the ROC curve is to the top left corner, the better the model performs. All three ROC curves lie on the top left corner and achieve an AUC of 1.00. This indicates that the model achieves perfect classification for each of the classes: Adult, Kids, and Teen, which is an ideal scenario in classification tasks.

The AUC of 1.00 for all classes implies that the model has an excellent ability to discriminate between the different classes without any overlap or misclassification. This level of performance is extremely rare in practical scenarios and suggests that the model has been trained on very distinctive features or the dataset is highly separable. Such perfect performance may raise concerns about overfitting, especially if the model was evaluated on the same dataset it was trained on. Overfitting occurs when a model learns the training data too well, including noise and outliers, which negatively impacts its performance on unseen data. In a real-world application, achieving perfect classification ensures that the model will make no errors in identifying the classes, which is particularly crucial in sensitive applications such as medical diagnosis, fraud detection, and safety-critical systems.

The ROC curve analysis demonstrates that the classifier performs perfectly across all classes, with an AUC of 1.00 for Adult, Kids, and Teen categories. While this is an ideal result, it is essential to validate the model using a separate test dataset to ensure that the performance generalizes to new, unseen data and to rule out the possibility of overfitting. Further steps should include cross-validation and testing on diverse datasets to confirm the robustness and generalizability of the model.

Results Comparison with the Benchmark work

In this section, the performance of the proposed classification model is compared with benchmark models to evaluate its efficacy. This comparison focuses on key metrics such as accuracy, precision, recall, training time, testing time and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). By analyzing these metrics, the aim is to determine whether the proposed model offers significant improvements over existing approaches and to highlight any areas where it may fall short. This comparative analysis provides a comprehensive understanding of the model's strengths and weaknesses in relation to established standards in the field. Table 3 shows the performance evaluation of the developed model compared with the benchmarked models under the same conditions.

Table 3: Results Comparison between the Benchmark and the Developed Model

Model	Accuracy
Odion et al. (2022)	
MobileNetv2	92%
EfficientNet-B1	93%
ResNet-50	95%
VGG-16	94%
Developed Model	
ResNet50+Vgg16 +Random Forest	98.85%

The comparison between the hybrid developed model and the benchmark models presented by Odion et al. (2023) reveals a significant improvement in classification performance. The benchmark models included MobileNetv2, EfficientNet-B1, ResNet-50, and VGG-16, which achieved accuracy rates of 92%, 93%, 95%, and 94%, respectively. These models, known for their robust architectures and widespread use in image classification tasks, set a high standard for performance.

MobileNetv2, designed for mobile and embedded vision applications, demonstrated a respectable accuracy of 92%. EfficientNet-B1, part of the EfficientNet family that optimizes both accuracy and efficiency, improved slightly to 93%. ResNet-50, a deep residual network known for its ability to train very deep networks without the vanishing gradient problem, achieved an accuracy of 95%. VGG-16, a well-known convolutional neural network model with a deep architecture, attained an accuracy of 94%.

In comparison, the developed model, which integrates ResNet-50, VGG-16, and a Random Forest classifier, significantly outperformed these benchmark models. The total computational complexity of this model was 4.41 billion FLOPs requiring 37% less computational complexity as against 6.99 billion FLOPs for the benchmark model. This model used only ResNet50 (23.6M parameters) and VGG16 (14.7M parameters), resulting in a manageable parameter count of 38.3M across 194 layers.

Achieving an accuracy of 98.85%, the hybrid model leveraged the strengths of both deep learning and ensemble learning techniques. The deep features extracted by ResNet-50 and VGG-16 provided a rich representation of the input data, while the Random Forest classifier added robustness through its ensemble approach, aggregating the predictions of multiple decision trees to enhance overall performance.

This substantial improvement in accuracy underscores the effectiveness of combining deep learning models with traditional machine learning classifiers. The hybrid approach not only capitalizes on the deep feature extraction capabilities of ResNet-50 and VGG-16 but also benefits from the Random Forest's ability to handle complex classification tasks with high accuracy. This synergy results in a model that excels in precision and reliability, making it a superior choice for classification tasks compared to the individual benchmark models.

The findings highlight the potential of hybrid models in achieving higher accuracy rates, especially in scenarios where state-of-the-art deep learning models alone may not suffice. By integrating complementary techniques, the developed model sets a new benchmark, demonstrating that a thoughtful combination of methodologies can lead to exceptional improvements in performance. This comparative analysis illustrates the advancements made in the field and paves the way for future research to explore similar hybrid approaches for even greater efficacy in various applications.

CONCLUSION

In conclusion, this study introduces a novel hybrid model for age prediction from sclera images, which combines deep learning architectures ResNet-50 and VGG-16 with a Random Forest classifier. The hybrid model demonstrates outstanding performance, achieving an overall accuracy of 98.85% and surpassing benchmark models in accuracy. Detailed evaluation metrics reveal high precision, recall, and F1-scores across age groups, supported by insights from the confusion matrix. The model's practical applicability is highlighted through efficient training and testing processes, indicating feasibility for real-time scenarios. Moreover, this research addresses significant gaps in existing literature by integrating transfer learning, deep learning, and ensemble methods, while also tackling issues of computational complexity. By meticulously optimizing both accuracy and efficiency, the hybrid model sets a new standard for age prediction from biometric images. Overall, the findings underscore the potential of hybrid models to advance age prediction research and contribute to the broader field of biometric image analysis. This work paves the way for future investigations into the application of hybrid models in various domains and establishes a solid foundation for further advancements in predictive modeling using biometric data.

REFERENCES

Abbasi, A., & Khan, M. (2016). Iris-pupil thickness-based method for determining age group of a person. *Int. Arab J. Inf. Technol*, *13*(6).

Aminu Bashir Suleiman, S. L. (2023). CARDIOVASCULAR DISEASE PREDICTION USING RANDOM FOREST MACHINE LEARNING. *FUDMA Journal of Sciences (FJS)*, 282-289.

Awujoola, O. J., Aniemeka, T. E., Ogwueleka, F. N., Abioye, O. A., Awujoola, A. E., & Uwa, C.O. (2024). Improved Breast Cancer Detection in Mammography Images: Integration of Convolutional Neural Network and Local Binary Pattern Approach. In *Machine Learning Algorithms Using Scikit and TensorFlow Environments* (pp. 221-248). IGI Global.

Beattie, J. R., Pawlak, A. M., McGarvey, J. J., & Stitt, A. W. (2011). Sclera as a surrogate marker for determining AGE-modifications in Bruch's membrane using a Raman spectroscopy–based index of aging. *Investigative ophthalmology & visual science*, *52*(3), 1593-1598.

Das, S., De Ghosh, I., & Chattopadhyay, A. (2021). An efficient deep sclera recognition framework with novel sclera segmentation, vessel extraction and gaze detection. *Signal Processing: Image Communication*, *97*, 116349.

Erbilek, M., Fairhurst, M., & Abreu, M. C. D. C. (2013, December). Age prediction from iris biometrics. In *5th International Conference on Imaging for Crime Detection and Prevention (ICDP 2013)* (pp. 1-5). IET.

Galbally J, Haraksim R, Beslay L. "A Study of Age and Ageing in fingerprint Biometrics", IEEE Transactions on Information Forensics and Security, 14(5), 1351–1365, 2019.

Irhebhude, M. E., Kolawole, A. O., & Abemi, H. (2023). Age And Gender Classification from Iris Images of the Eye Using Machine Learning Techniques. *Academy Journal of Science and Engineering*, *17*(2), 54-69.

Kumar S, Rani S, Jain A, Verma C, Raboaca MS, Illés Z, Neagu BC. "Face Spoofing, Age, Gender and Facial Expression Recognition Using Advance Neural Network Architecture-Based Biometric System", Sensors, 22(14), 51-60, 2022.

Narejo, S., Pasero, E., & Kulsoom, F. (2016). EEG based eye state classification using deep belief network and stacked autoencoder. *International Journal of Electrical and Computer Engineering (IJECE)*, *6*(6), 3131-3141.

Odion, P. O., Musa, M. N., & Shuaibu, S. U. (2022). Age Prediction from Sclera Images using Deep Learning. *Journal of the Nigerian Society of Physical Sciences*, 787-787.

Olorunsola, o., & olorunshola, O. Deep Learning Ensemble Approach to Age Group Classification based on fingerprint pattern. Advances in Artificial Intelligence Research 3(2), 76-84.

Othmani, A., Taleb, A. R., Abdelkawy, H., & Hadid, A. (2020). Age estimation from faces using deep learning: A comparative analysis. *Computer Vision and Image Understanding*, *196*, 102961.

Oyewola, D. O., Dada, E. G., Ndunagu, J. N., Umar, T. A., & Akinwunmi, S. A. (2021). COVID-19 risk factors, economic factors, and epidemiological factors nexus on economic impact: machine learning and structural equation modelling approaches. *Journal of the Nigerian Society of Physical Sciences*, 395-405.

Rajput, M. R., & Sable, G. S. (2020). Age Group Estimation from Human Iris. In *Soft Computing and Signal Processing: Proceedings of 2nd ICSCSP 2019 2* (pp. 519-529). Springer Singapore.

Rajput, M., & Sable, G. (2019, June). Deep learning-based gender and age estimation from human Iris. In *Proceedings of the international conference on advances in electronics, electrical & computational intelligence (ICAEEC)*.

Russell, R., Sweda, J. R., Porcheron, A., & Mauger, E. (2014). Sclera color changes with age and is a cue for perceiving age, health, and beauty. *Psychology and Aging*, *29*(3), 626.

Sgroi, A., Bowyer, K. W., & Flynn, P. J. (2013, June). The prediction of old and young subjects from iris texture. In *2013 International Conference on Biometrics (ICB)* (pp. 1-5). IEEE.

Umarani, V., Julian, A., & Deepa, J. (2021). Sentiment analysis using various Machine learning and deep learning Techniques. *Journal of the Nigerian Society of Physical Sciences*, 385-394.

Vitek, M., Rot, P., Štruc, V., & Peer, P. (2020). A comprehensive investigation into sclera biometrics: a novel dataset and performance study. *Neural Computing and Applications*, *32*, 17941-17955.