

Sclera-Based Biometric for Age Group Estimation Using a Hybrid ResNet-50 and Gradient Boosted Neural Network



P. O. Odion¹, M. N. Musa^{2*}, Z. Z. Ahmad¹

¹Department of Computer Science, Nigerian Defence Academy, Kaduna, Nigeria.

²Department of Cyber Security, Nigerian Defence Academy, Kaduna, Nigeria.



ABSTRACT: Age estimation using biometric features, particularly sclera biometrics, has emerged as a promising alternative to facial recognition in scenarios with obscured facial features, garnering attention in the fields of computer vision and security. This study proposed a hybrid sclera-based biometric model combining ResNet-50 for feature extraction and a Gradient Boosting Neural Network (GBNN) for classification to improve age group estimation. A dataset of 2,400 sclera images from 300 subjects, categorized into four age groups (*Children: 0–12, Teens: 13–19, Young Adults: 20–39, Older Adults: 40+*) was utilized, sourced from Nigerian populations to ensure regional representation and preprocessed using Otsu thresholding for precise sclera segmentation. Evaluated on a 60/20/20 train-validation-test split, the model achieved 94.15% accuracy, outperforming standalone ResNet-50 (92.08%). The GBNN's iterative refinement reduced misclassifications in challenging categories (e.g., *Children*) by 50%, attributed to its ability to learn residual errors and adaptively weight underrepresented samples. These results highlight the potential for deployment in biometric security systems and age-restricted content moderation, particularly in low-light environments where sclera visibility is consistent. While geographic homogeneity in the dataset and illumination-dependent on Otsu thresholding represent current limitations, this work establishes a foundation for future advancements in multimodal biometric integration and cross-population validation.

KEYWORDS: Age Estimation, Sclera Biometrics, Gradient Boosting Neural Network, Hybrid Models, Otsu Thresholding and Ensemble learning

[Received Jan. 27, 2025; Revised April 8, 2025; Accepted May 5, 2025]

Print ISSN: 0189-9546 | Online ISSN: 2437-2110

I. INTRODUCTION

Biometric characteristics, which are inherently unique to each individual, have emerged as fundamental components of modern identity verification, recognition, and authentication systems (Othmani et al., 2022). Technologies such as facial, ear, voice, and iris recognition have attracted considerable attention from researchers and practitioners in academic and industrial sectors. These biometric systems' secure nature presents a significant advantage over traditional password-based authentication methods (Narejo et al., 2016). Among the various biometric modalities, facial recognition is particularly notable for its versatility, as it can capture a range of attributes including age, gender, ethnicity, emotion, expression, and identity. Notably, age plays a critical role in influencing human interactions and perceptions. The aging process, characterized by dynamic alterations in facial features due to intrinsic and extrinsic factors, is reflected in changes in skin texture, wrinkles, variations in tissue composition, and shifts in coloration (Othmani et al., 2020). These changes have spurred significant research interest focused on age estimation, aiming to identify patterns that delineate the aging process.

Despite its advantages, the application of facial recognition for age estimation encounters several challenges. Variables

such as race, facial hair, cosmetic surgery, and skin tone introduce complexities that hinder accurate predictions. Consequently, there has been a growing interest in exploring alternative biometric regions, with particular emphasis on the human eye, specifically the iris and sclera as a promising avenue for research (Odion et al., 2022). The eye, a crucial component of human interaction and expression, undergoes age-related transformations, including changes in pupil size, scleral texture, and corneal shape, which can provide valuable insights for age estimation (Rajput and Sable, 2019). The sclera, which is the visible white part of the eye, is particularly advantageous due to its consistent visibility, ease of capture, and significant changes associated with aging and health, rendering it an attractive feature for biometric research (Taheri and Toygar, 2019). These attributes present an opportunity to develop more reliable and accurate age estimation systems utilizing eye-based biometrics.

Traditional methods for age estimation in facial and eye recognition have predominantly relied on conventional computer vision techniques, such as Local Binary Patterns (LBP), Biologically Inspired Features (BIF), and Active Appearance Models (AAM). These methodologies typically involved manual feature extraction processes, employing techniques like Histogram of Oriented Gradients (HOG) and

*Corresponding author: muhammadmusa2502@nda.edu.ng

Principal Component Analysis (PCA) to discern aging patterns, followed by age estimation through machine learning models such as Support Vector Machines (SVM) or Support Vector Regression (SVR) (Aggarwal, 2018; Othmani et al., 2020). However, these approaches often face challenges in generalization due to their limitations in capturing the nuanced details of smaller regions, such as the sclera.

The advent of deep learning has revolutionized computer vision, enabling the automatic extraction of complex and abstract features from raw image data. Convolutional Neural Networks (CNNs), in particular, have demonstrated exceptional performance in image recognition tasks, leveraging hierarchical data representations to eliminate the need for manual feature engineering (Taheri and Toygar, 2019). Nevertheless, the success of deep learning models relies heavily on access to large labeled datasets, which can be difficult to obtain in some contexts (Yi et al., 2019). To address this, hybrid models have emerged, combining pre-trained deep learning networks for feature extraction with machine learning algorithms for classification (Musa et al., 2024). These hybrid frameworks capitalize on the strengths of deep learning for robust feature learning while leveraging ensemble learning techniques to enhance classification accuracy.

Building on this foundation, recent studies have explored the potential of sclera-based age estimation. For instance, Irhebhude et al. (2023) achieved 99.27% accuracy using hand-crafted feature extraction in iris-based age and gender estimation, while Odion et al. (2022) attained 95% accuracy using ResNet-50 for sclera-based age classification. However, these studies typically categorized ages into broad classes, such as child, teen, and adult, leaving room for more granular classifications. The highlighted gap in knowledge includes advancing the field by introduction of a four-class age categorization scheme and leveraging a hybrid model that integrates ResNet-50 for feature extraction with a Gradient Boosting Neural Network (GBNN) for classification. This approach seeks to improve upon Odion et al.'s (2022) benchmark by enhancing the precision and generalizability of age estimation.

This study focuses exclusively on developing a hybrid sclera-based biometric model, combining ResNet-50 for feature extraction and a GBNN for classification to improve age group estimation. A dataset comprising images from Odion et al. (2022) for individuals aged 0–30 years and additional images from postgraduate students at the Nigerian Defence Academy, Kaduna for ages above 30 were utilized.

II. LITERATURE REVIEW

This section reviews the literature on the application of deep learning for age estimation across eye regions, with an emphasis on the initial exploration of sclera usage. It also examines the evolution of neural network architectures, transfer learning, convolutional neural networks (CNNs), and the Gradient Boosted Neural Network (GBNN) methodology, highlighting their significance in this domain.

A. Overview of Deep Learning

The field of deep learning has undergone significant evolution, with its roots tracing back to the 1940s. The McCulloch-Pitts

(MCP) model, developed by McCulloch and Pitts, introduced the concept of artificial neurons mimicking the brain's pattern generation (Rondeau, 2019). This foundational work paved the way for the development of artificial neural networks.

A significant turning point in the field of deep learning emerged when Hinton et al. (2006) introduced the concept of Deep Belief Networks, featuring stratified restricted Boltzmann machines. This innovation spurred advancements in deep architectures, fueled by curated datasets and the swift advancement in parallel GPU computing. This transition from CPU-based to GPU-based training has played a pivotal and expeditious role in facilitating the evolution of deep learning paradigms (Voulodimos et al., 2018). Deep learning, a subset of machine learning, excels at extracting features from large datasets, making it particularly effective for tasks like image recognition (Madan and Madhavan, 2020).

B. Transfer Learning

Transfer learning (TL) facilitates leveraging knowledge from pre-trained deep learning models for new tasks, particularly when data is insufficient. It reduces training time by utilizing parameters from earlier models (Yi et al., 2019). Contemporary implementations of TL predominantly adopt dual adaptation strategies: (i). *feature extraction*, which preserves pre-trained hierarchical representations while retraining task-specific classification layers (Chollet, 2018), and (ii). *Fine-tuning* is a more nuanced approach that selectively updates earlier network weights to align with target domain characteristics (Bengio et al., 2017; Chollet, 2018; Chen et al., 2020). As illustrated in Figure 1, this framework operationalizes the distillation of high-level features originally learned from large-scale datasets such as ImageNet into specialized architectures tailored for different tasks.

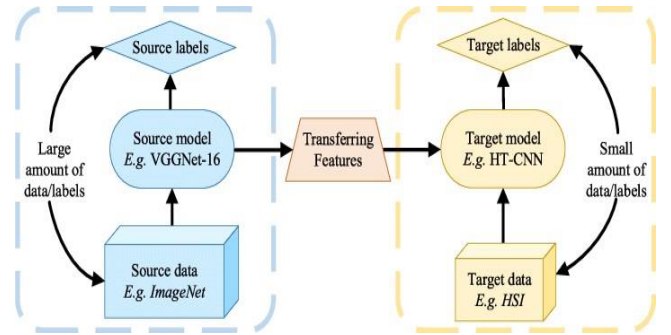


Figure 1: Transfer learning method (Thai et al., 2019).

C. Convolution Neural Network (CNN) Architecture

CNNs are essential for image classification, featuring convolutional, pooling, and fully connected layers. They employ sparse connections and weight sharing, reducing computational complexity (Wang et al., 2019). These components work synergistically to streamline the complex processes of image feature extraction and data reconstruction during image recognition and processing tasks. The convolution kernel extracts features, while pooling layers downsample data to prevent overfitting and the fully connected layer performs the classification (Zang et al., 2018; Su et al., 2019). This process is depicted in Figure 2.

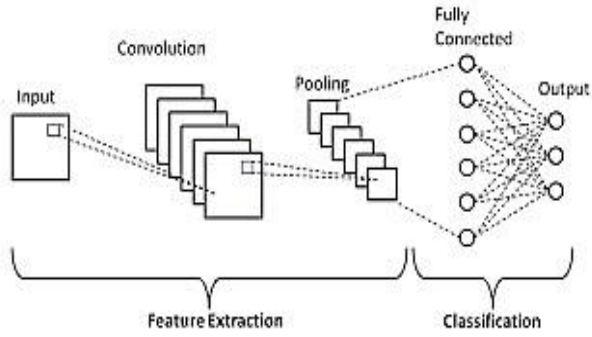


Figure 2: Convolutional Neural Network (Wang et al., 2019).

1) ResNet Model Architecture

ResNet (Residual Network), pioneered by He et al. (2016), introduced a transformative framework for deep learning by resolving the persistent challenge of vanishing gradients through its residual learning paradigm. Central to its

innovation are residual blocks equipped with shortcut connections, which enable the network to learn residual mappings rather than direct transformations. This architectural breakthrough circumvented the degradation problem that plagued earlier architectures—such as VGG and plain CNNs—which exhibited diminished accuracy as network depth increased due to architectural constraints. ResNet’s scalable design encompasses a spectrum of model variants (e.g., 34, 50, 101, and 152 layers, as depicted in Figure 3), demonstrating exceptional adaptability to diverse computational requirements. Remarkably, despite its exponential depth escalation, ResNet maintains parameter efficiency and computational tractability, attributable to its systematic use of bottleneck layers and dimensionality reduction strategies (He et al., 2016). This balance between depth and complexity makes its enduring relevance in tasks demanding hierarchical feature extraction, such as medical imaging and biometric analysis, where both precision and computational feasibility are critical.

Figure 3: ResNet Architecture (He et al., 2016)

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

2) Gradient Boosted Neural Network (GBNN)

GBNN, proposed by Emami and Martínez-Muñoz (2023), combines gradient boosting and neural networks for classification. It iteratively trains small neural networks to fit the negative gradients (pseudo-residuals) of the cross-entropy loss computed from the previous iteration, enhancing flexibility and efficiency. This method allows dynamic adjustment of hidden units during inference, reducing computational costs without sacrificing accuracy. The GBNN framework provides an alternative to traditional training, offering robustness against overfitting and improved interpretability (Emami and Martínez-Muñoz, 2023).

D. Review of Related Works

Deep learning techniques have been extensively employed in the realm of age estimation and classification to extract and analyze biometric characteristics (Jang et al., 2017). However, limited efforts have focused on leveraging ocular images for age prediction. Early studies, such as that of Sgroi et al. (2013), investigated iris texture for categorizing individuals into distinct age groups. Using a dataset of 596 iris images (300 young, 296 old), segmentation and normalization

were followed by feature extraction and classification using a Random Forest classifier via the WEKA tool, achieving a classification rate of 64.68%. These findings suggested the presence of aging-related factors for slight age determination, with a 95% confidence interval of randomness.

Similarly, Erbilek et al. (2013) employed five geometric iris features from 210 subjects (ages 18–73) divided into young, adult, and older groups. The study implemented segmentation for detecting iris and pupil boundaries, followed by the extraction of twelve geometric features. Using Spearman’s rank correlation for feature selection and classification techniques like Support Vector Machines (SVM), K-Nearest Neighbor (KNN), and Multi-Layer Perceptron in MATLAB, the model achieved 75% accuracy. Rajput and Sable (2019) further advanced iris-based age classification by using pre-processing, segmentation with circular Hough Transform, and normalization through Daugman’s rubber sheet model. With Bagged Ensemble classifiers, their model reached an overall accuracy of 83.7%, outperforming other classifiers by reducing variance and mitigating overfitting.

Advancements in deep learning have further improved age prediction accuracy. Shigueoka et al. (2021) utilized peripapillary spectral-domain optical coherence tomography (SD-OCT) B-scans to predict age. Their model employed ResNet-50 convolutional neural networks (CNNs) to achieve a strong correlation between predicted and observed ages (MAE = 5.82 years; $r = 0.860$, $P < 0.001$), with an area under the curve (AUC) of 96.2%. Gowruju et al. (2022) tackled automatic age prediction from pupil images using UNet for segmentation without a dense layer, enhanced with skip connections and morphological processing. Their approach, tested on three benchmark datasets, achieved a remarkable accuracy of 96%, with optimization techniques such as Adaptive Moment Optimizer yielding the best results.

In the field of sclera biometrics, Das et al. (2022) and Odion et al. (2022) investigated age estimation based on scleral features. Das et al. (2022) framed the problem as a regression task, utilizing a modified VGG-16 network on a custom dataset, which resulted in a mean absolute error (MAE) of 0.06. In contrast, Odion et al. (2022) approached the task as a classification problem, employing a local dataset and pre-trained deep learning models. Among these pre-trained models, ResNet-50 achieved a state-of-the-art classification accuracy of 95%.

Rot et al. (2020) further emphasize the potential of deep learning for sclera recognition by utilizing the VGG architecture, achieving F1-scores exceeding 95.2% on standardized benchmarks. Their novel dataset, the Sclera Blood Vessel, Periocular, and Iris (SBVPI), includes ocular images captured from various gaze directions.

The choice of ResNet-50 for feature extraction was motivated by its proven effectiveness in computer vision tasks, particularly its ability to learn hierarchical representations of images through its deep architecture (Musa et al., 2024). Previous research has demonstrated ResNet-50's superior performance in feature extraction for biometric applications, including age estimation from facial and ocular images (Odion et al., 2022). The Gradient Boosting Neural Network (GBNN) was selected for its capacity to iteratively refine predictions by focusing on difficult samples and residual errors, which is particularly advantageous for handling the nuanced patterns present in sclera images. This ensemble approach allows the model to address class imbalances and complex decision boundaries that might challenge a single neural network architecture. This approach builds on the successes of previous research while addressing the limitations in age group granularity and classification accuracy.

III. METHODOLOGY

The study developed a model for carrying out feature extraction through ensembling pre-trained ResNet-50 architecture before finally applying a GBNN for age classification. So, this study adopted the standard image recognition pipeline as depicted in Figure 4.

A. Image Acquisition (Dataset)

The initial phase of this research involved data acquisition, a critical step in building the proposed model. Image data was sourced from two distinct populations to create a comprehensive dataset of eye images. The primary source was

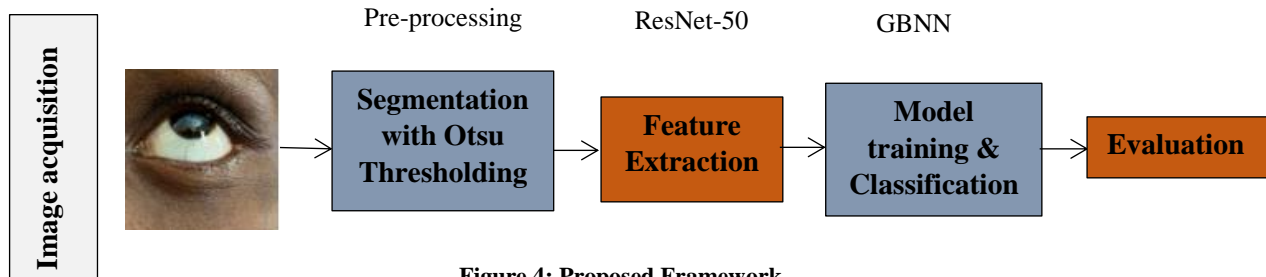


Figure 4: Proposed Framework

Beyond ocular-specific approaches, hybrid methodologies integrating traditional machine learning have shown promise. Irhebhude et al. (2023) utilized 3D histograms and Principal Component Analysis (PCA) for dimensionality reduction of iris, with SVM achieving an accuracy of 99.27%. Contemporary segmentation frameworks, including Otsu thresholding (2022), U-Net (Ronneberger et al., 2015), RefineNet (Lin et al., 2017), and SegNet (Rot et al., 2020) have demonstrated significant advances in sclera segmentation. Despite these advancements, there remains a gap in sclera-based age estimation research. Existing studies have primarily focused on classifying individuals into three broad age groups (e.g., child, teen, adult) (Odion et al., 2022; Irhebhude et al., 2023). To address this, efforts are being made to refine age classification by introducing four distinct age groups, offering a more granular and precise estimation compared to prior sclera-based models.

the dataset published by Odion et al. (2022). This dataset contained images of individuals between the ages of 0 and 30, stratified into three age cohorts: children, adolescents, and young adults. Data collection for this portion involved capturing 2000 images from 250 participants using an iPhone 11 Pro Max at its highest resolution and image quality settings. The images were acquired at two locations: a local government primary school in Badarawa, Kaduna, and the Laraba Institute of Computing and International Languages, also in Kaduna. To facilitate robust recognition and classification, images were captured with variations in gaze direction. Participants were asked to maintain focus in four directions: straight ahead, upwards, leftwards, and rightwards, for each eye, yielding four distinct gaze angles per subject.

To broaden the age representation within the dataset and include older individuals, a secondary collection effort was undertaken. An additional 440 eye images were gathered from 55 postgraduate students at the Nigerian Defence Academy

(NDA). This supplemental data, collected using an iPhone 15 following the same image acquisition protocol described for the Odion et al. (2022) dataset, focused on individuals aged 30 and above. Throughout the data collection process, strict adherence to ethical principles was maintained, prioritizing participant privacy and securing informed consent. Images were captured in natural, uncontrolled environmental settings. Furthermore, a quality control measure was implemented, resulting in the exclusion of approximately 40 images from the original dataset due to incomplete capture of the sclera. The final dataset was organized into four age categories, detailed in Table 1. The sample size aligns with prior studies (Odion et al., 2022) and ensures representation across four age groups

Table 1: Age-group classes for the Dataset

Age Group	Age Range	Number of Subjects	Number of Images
Children	0 – 12	84	672
Teens	13 – 19	81	648
Young Adults	20 – 39	80	640
Older Adults	40 and above	55	440
Total		300	2,400

B. Pre-processing

It is very important for raw data to be transformed before carrying out the classification as it reduces the training time and to attain highly accurate results. Pal and Sudeep (2016), define image pre-processing as the process of cleaning, resizing, and transforming raw data images before it is fed into the network.

1) Image Segmentation

Image segmentation, a crucial preprocessing step in computer vision, image processing, and pattern recognition, employs algorithms to partition an image into distinct segments or pixel groupings based on shared characteristics (Murzova and Seth, 2020). The primary objective of this process is to isolate the region of interest from a potentially complex background. A range of segmentation techniques exists, including thresholding, clustering, edge-based methods, region-based approaches, artificial neural networks, and partial differential equation-based methods (Murzova and Seth, 2020). In this study, automatic segmentation was performed using the Otsu thresholding technique, as implemented by Odion et al. (2022). This method was chosen for its capacity to produce accurate, high-quality segmentation results, facilitating the precise extraction of relevant information from the sclera images under analysis.

C. Feature Extraction by ResNet-50

In this phase, the initial layers of pre-trained networks are employed as feature extractors, leveraging their capacity to identify fundamental patterns, shapes, and textures learned from their original training data (Priya, 2021). These pre-trained weights are then applied to the sclera images, effectively transferring the intrinsic features learned by the model without further fine-tuning of these initial layers, as illustrated on the left-hand side of Figure 5. The model is truncated after the fifth convolutional block, and the resulting

output is then used for classification. This methodology offers advantages such as a reduced risk of overfitting and faster training times, as the focus is shifted to training the later layers of the model while preserving the learned representations of the initial layers (Priya, 2021). Building on the foundational work of Odion et al. (2022), which demonstrated the effectiveness of pre-trained ResNet-50 models for age classification, this study introduces a streamlined hybrid framework. Here, ResNet-50 serves as a feature extractor: its pre-trained layers automatically identify key patterns in sclera images, and these unmodified features are then fed into an ensemble learning classifier. By directly repurposing ResNet-50's learned representations (originally trained on general image datasets like ImageNet), our method avoids the computational burden of re-training the entire network. This aligns with core principles of transfer learning—reusing existing knowledge for new tasks—while simplifying the workflow. The approach not only maintains high accuracy but also reduces training time and resource demands, making it practical for real-world applications where labeled sclera data is limited.

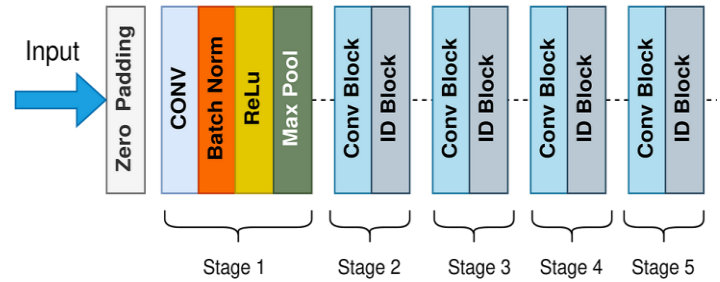


Figure 5: ResNet-50 for Feature Extraction

D. Model Training and Classification

This research classification approach involved the utilization of an ensemble machine learning algorithm known as a GBNN. This concept is adopted from the sequential training of neural networks with gradient boosting by Emami and Martínez-Muñoz (2023). As shown in Figure 6, in multi-class classification with K classes, the algorithm iteratively trains a single regression neural network with a few hidden neurons to fit the pseudo-residuals (negative gradients of the cross-entropy loss) computed from the previous iteration. This approach provides an alternative to standard backpropagation training of neural networks and allows dynamically adjusting the number of active hidden units during inference without significant accuracy loss.

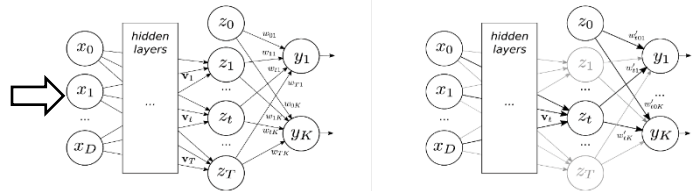


Figure 6: Boosting of Neural Network for Classification (Emami and Martínez-Muñoz, 2023).

This study incorporated the technique described by Emami and Martínez-Muñoz (2023) into the classification process to capitalize on its capacity to manage complex relationships and non-linear patterns present within the data. The GBNN ensemble learning strategy employs a sequential training approach, where each subsequent neural network is trained to mitigate the shortcomings of its predecessor. This iterative refinement enhances both the model's robustness and its predictive accuracy. Furthermore, the parameters of the GBNN were optimized for the specific task of age group classification, with the optimized parameters detailed in Table 2. This optimization process involved adjusting key parameters, including learning rates, tree depth, and the number of boosting iterations, to achieve the best possible performance.

Table 2: Fine-Tuned Parameters Used for the GBNN

Parameter	Value	Description
total_nn	30	Increased total number of nearest neighbours to enhance model accuracy.
num_nn_step	2	Step size for nearest neighbour search during feature aggregation.
eta	0.3	Learning rate adjusted to balance convergence speed and stability.
solver	'adam'	Optimizer switched from SGD to Adam for faster and more reliable convergence.
subsample	1.0	Full dataset used for training each tree in the ensemble (no subsampling).
tol	0.0	Early stopping disabled; training continues until max_iter is reached.
max_iter	30	Maximum iterations increased to ensure thorough convergence of the model.
random_state	42	Seed fixed for reproducibility of experimental results.
activation	'tanh'	Activation function changed to tanh for improved non-linear decision boundaries.

E. Model Evaluation

Model Evaluation is the process of assessing how well a model performs against real data. This specifically applies the model to test the data to determine. If the model, based on a training set, can be applied to other data. In particular, it helps eliminate the overfitting phenomenon that can occur whenever the same data is used for both training and testing and matches the built data too well, but underperforms on other data (Ritchie, 2018). The model implemented a train-test split on the dataset, dividing it into training, validation, and test sets in a 60/20/20 ratio. Subsequently, the test set was designated for the evaluation efforts. Apart from classification metrics, class-wise analysis was also performed.

The primary evaluation metric for the supervised learning models was accuracy as described in equation 1. It measures

the proportion of correct predictions made by a model compared to the total number of predictions.

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

1) Class-wise analysis

In this subsection, the class-wise analysis of the proposed method was performed. The metrics used include precision (2), recall (3), and F1-score (4) for each class on the corresponding dataset, defined as follows:

- a) **Precision:** Is about how precise or how often is the prediction correct? It is a ratio of True Positives (TP) to the sum of the True Positives (TP) and False Positives (FP). It calculated as follows:

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

- b) **Recall:** When the actual value is positive, how often is the prediction correct? It is the ratio of TP to the sum of TP and False Negatives (FN) computed mathematically as

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

- c) **F1-Score:** F1-Score is also known as F1 Score. It is the harmonic mean of precision and recall. Harmonic mean is appropriate for situations where the average of rates (a ratio between two related quantities) is desired. It is calculated as

$$F1\ Score = 2 \times \frac{Precision \times Recall}{(Precision+Recall)} \quad (3)$$

The obtained results across the metrics in 1, 2, 3, and 4 are presented in the result section.

IV. RESULTS AND DISCUSSION

This section presents the results and performance analysis of the proposed sclera classification system, which utilizes ResNet-50 for feature extraction and Gradient Boosting Neural Network (GBNN) for classification. The result of the segmentation process, employing Otsu thresholding, was also provided. The performance of the ResNet-50 and ResNet-50+GBNN models is evaluated using various metrics, including accuracy, precision, recall, and F1-score.

A. Result of Segmentation with Otsu Thresholding

To segment the sclera region from the original eye image, the Otsu thresholding technique was employed. This method automatically calculates a threshold value that effectively separates the image into foreground (sclera) and background pixels based on their intensity distribution. As illustrated in Figure 7, the Otsu thresholding technique successfully segmented the sclera region, generating a clear and accurate mask for subsequent analysis.



Figure 7: Image Segmentation using Otsu Thresholding Technique

The segmented sclera regions were then utilized for further analysis in the feature extraction stage, ensuring that only the relevant portion of the eye was processed for age estimation.

B. Model Performance of ResNet-50

The ResNet-50 model was initially applied to extract features from the segmented sclera images. These features were then passed through a fully connected softmax classifier to predict the age group. Figure 8 provides the classification report of ResNet-50 across the four age groups (Children, Teens, Young Adults, and Older Adults).

	precision	recall	f1-score	support
CHILD	0.93	0.88	0.90	134
OLDER ADULT	0.99	0.90	0.94	88
TEEN	0.95	0.97	0.96	130
YOUNG ADULT	0.85	0.93	0.89	128
accuracy			0.92	480
macro avg	0.93	0.92	0.92	480
weighted avg	0.92	0.92	0.92	480

Figure 8: Classification Report of ResNet-50

Based on reported findings by Odion et al. (2022), ResNet-50 demonstrated superior performance compared to other pre-trained models for age group classification using sclera with three age groups. Hence, the reason for experimenting with the model. The report summarizes the classification performance of ResNet-50 across the four age groups: Children, teens, young adults, and older adults. The results indicate that ResNet-50 alone performed well, with an overall accuracy of 92.08%. However, the model's performance was slightly less consistent across the age groups, particularly in the "teens" and "young adults" categories, where F1-scores are 96% and 89% respectively. The most successful classification was observed in the "teens".

C. Model Performance: ResNet-50 Plus GBNN

To enhance classification performance, a GBNN was integrated following feature extraction with ResNet-50. This integrated approach, combining ResNet-50 with GBNN, leveraged the advantages of ensemble learning to achieve improved classification accuracy and a more balanced performance across all age groups. The performance of this hybrid model is summarized in Figure 9.

[INFO] evaluating...				
	precision	recall	f1-score	support
CHILD	0.93	0.94	0.94	134
OLDER ADULT	1.00	0.91	0.95	88
TEEN	0.99	0.97	0.98	130
YOUNG ADULT	0.86	0.93	0.89	128
accuracy			0.94	480
macro avg	0.95	0.94	0.94	480
weighted avg	0.94	0.94	0.94	480

Figure 9: Classification Report for ResNet50-GBNN

Figure 9 detailed the performance metrics, presenting precision, recall, and F1-score for each age group classification. Compared to using ResNet-50 alone, the hybrid approach demonstrated improvements in both overall accuracy and F1-score across all age groups. Specifically, the overall accuracy increased from 92.08% to 94.15%. While the "teens" category maintained the highest performance, the improvements observed in the "children" and "older adults" groups were particularly significant, highlighting the synergistic effect of combining ResNet-50's deep learning capabilities with GBNN's ensemble learning strengths.

D. Class-Wise Comparative Analysis

To further evaluate the performance of the models, a class-wise comparative analysis was conducted between the ResNet-50 model employed by Odion et al. (2022) and the proposed ResNet-50 + GBNN hybrid model. This analysis, illustrated in Figures 10, 11, and 12, examines key metrics—precision, recall, and F1-score—across the four age groups: Children, Teens, Young Adults, and Older Adults. The comparison highlights the relative strengths and limitations of each model in accurately classifying individuals within these categories.

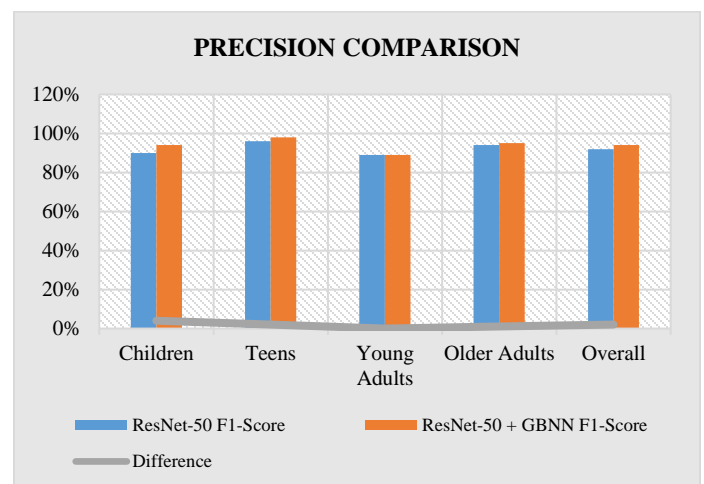


Figure 10: Precision Comparison

As shown in Figure 10, both models demonstrated comparable performance in the children group. However, the ResNet-50 + GBNN hybrid model outperformed ResNet-50 in the teens, young adults, and older adults groups, achieving precision gains of 4%, 1%, and 1%, respectively. This resulted

in an overall precision improvement of 2% for the proposed hybrid model.

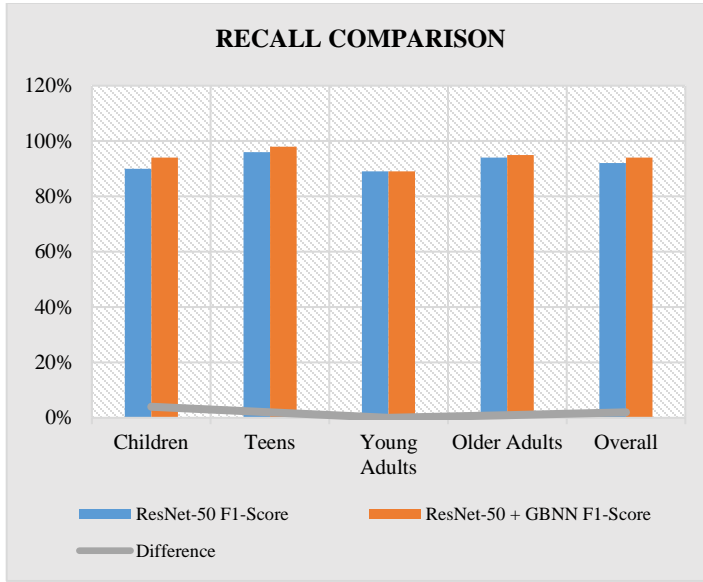


Figure 11: Recall Comparison

Figure 11 depicts the recall performance. The hybrid model exhibited a notable 6% increase in recall for the Children group. Recall remained consistent between the two models for the teens and young adults groups, while the older adults group experienced a modest improvement of 1%, culminating in a 2% overall enhancement in the recall.

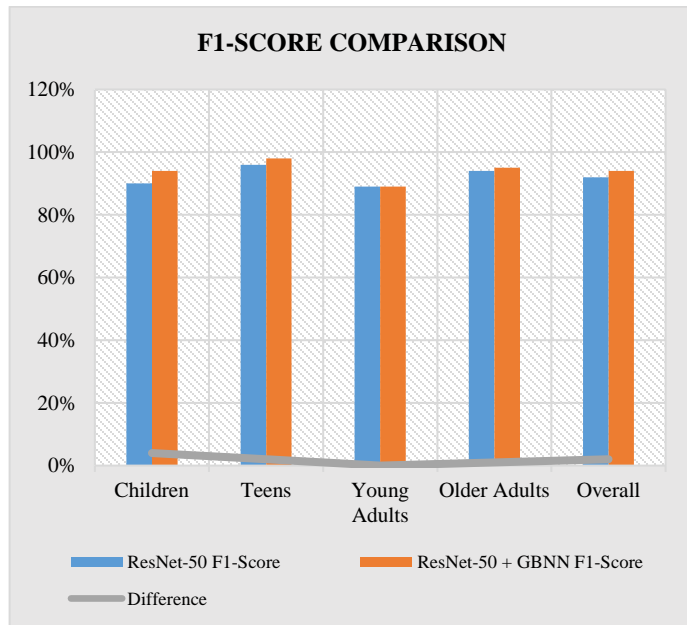


Figure 12: F1-Score Comparison

In Figure 12, the F1-score analysis revealed a 4% improvement for the children group in the hybrid model, signifying better harmony between precision and recall. Additionally, the teens and older adults groups showed gains

of 2% and 1%, respectively, while the young adults group maintained parity across both models. Overall, the hybrid model achieved a 2% increase in F1-score, further validating its superior performance in age group classification.

E. Confusion Matrix Analysis

The confusion matrix analysis, presented in Figure 13, further revealed the performance improvement achieved by the ResNet-50 + GBNN hybrid model over the standalone ResNet-50 model, particularly in the children category. The hybrid model correctly classified 126 out of 134 children, surpassing the ResNet-50 model, which correctly identified 118. This advancement reduced the number of misclassified children, particularly those incorrectly categorized as young adults. The ResNet-50 model misclassified 12 children into the young adults group, whereas the hybrid model reduced this error to just seven, demonstrating its enhanced discriminatory capability.

Furthermore, the hybrid model exhibited marginally better performance in the older adults category, accurately classifying 80 out of 88 instances, compared to 79 by the ResNet-50 model. This improvement, though slight, highlights the hybrid model's robustness in handling challenging classifications within this group. Both models performed equally well for the Teens and Young Adults categories, with the same number of correct classifications, demonstrating the hybrid model's consistency across different age groups.

	CHILD	OLDER ADULT	TEEN	YOUNG ADULT
CHILD	126	0	1	7
OLDER ADULT	0	80	0	8
TEEN	0	0	126	4
YOUNG ADULT	9	0	0	119
	Predicted label			

a. Confusion Matrix for ResNet50-GBNN

	CHILD	OLDER ADULT	TEEN	YOUNG ADULT	
CHILD	118	0	4	12	True label
OLDER ADULT	0	79	2	7	
TEEN	1	1	126	2	
YOUNG ADULT	8	0	1	119	
	Predicted label				

b. Confusion Matrix for ResNet50

Figure 13: Confusion Matrix for Both ResNet-50 and ResNet50-GBNN

In terms of misclassifications, the ResNet-50 + GBNN model reduced the number of errors in the children group by nearly 50%, significantly improving the model's precision and recall for this age group. While both models had some difficulty distinguishing older adults and young adults, the hybrid model introduced fewer errors across all categories. Overall, the hybrid model's use of ensemble learning contributed to a more accurate classification, particularly in age groups with overlapping features, such as Children and Young Adults, making it a better choice for handling challenging datasets.

F. Discussion

The hybrid ResNet-50 + GBNN model demonstrated superior performance compared to standalone ResNet-50 used by Odion et al. (2022), achieving an overall accuracy of 94.15% versus 92.08% and reducing misclassifications by 50% in challenging categories such as children and older adults. The superior performance of the hybrid model can be attributed to its ability to address specific challenges associated with different age groups. For children aged 0–12 years, the estimation process faces difficulties due to subtle sclera texture changes that can be easily confused with features of young adults. The hybrid approach overcame these challenges through the GBNN's iterative training on pseudo-residuals, which enabled dynamic prioritization of underrepresented samples and refinement of decision boundaries. By focusing on residual errors from ResNet-50's initial predictions, the ensemble corrected systematic biases, such as misinterpreting youthful sclera brightness as adult-like homogeneity. This refinement reduced misclassifications in this critical group from 12 to 7 errors, enhancing applications like child safety

systems where accurate age verification is essential for restricting access to age-inappropriate content.

For older adults (40+ years), the estimation process encounters heterogeneous patterns resulting from aging processes such as sclera yellowing and blood vessel prominence, which vary widely across individuals. The hybrid model addressed these challenges through ResNet-50's deep features that captured localized texture details (e.g., microvascular changes), combined with the GBNN's ensemble approach that aggregated diverse weak learners to model global patterns (e.g., overall discoloration). This dual capability improved robustness against intra-class variability, making the model more reliable for healthcare applications such as non-invasive screening for age-related ocular conditions (e.g., scleritis) or demographic studies tracking aging populations.

In contrast, both models performed comparably for teens and young adults (F1-score: 96% vs. 96%), likely due to clearer physiological distinctions in sclera texture that emerge post-adolescence. This consistency suggests that the hybrid approach maintains performance across different age groups while providing significant improvements in more challenging categories.

The practical implications of these findings are substantial across multiple domains. In biometric security, the model's resilience to facial occlusion (e.g., masks, scarves) makes it ideal for secure access control in environments where traditional facial recognition systems fail. For demographic analytics, the granular age-group classification capabilities aid policymakers in resource allocation, such as tailoring healthcare initiatives for older adults or educational programs for teenagers. Furthermore, the reduced bias in classifying underrepresented groups aligns with fairness goals in automated systems, addressing ethical concerns in AI development and deployment.

V. CONCLUSION

This study developed and evaluated a hybrid age estimation model, leveraging ResNet-50 for feature extraction and Gradient Boosting Neural Network (GBNN) for classification, with sclera images as the biometric input. By focusing on precise sclera segmentation through Otsu thresholding, the model effectively classified individuals into four age groups: Children, teens, young adults, and older adults. The results demonstrated that the hybrid ResNet-50 + GBNN model significantly outperformed the standalone ResNet-50 in terms of accuracy, precision, recall, and F1-score, achieving an overall accuracy of 94.15%. Notably, the hybrid approach reduced misclassification rates, particularly in challenging age categories such as children and young adults, and exhibited a balanced performance across all groups. The findings highlighted the effectiveness of combining deep learning with ensemble methods for biometric applications. This research underscores the potential of sclera-based age estimation as an alternative to traditional facial recognition, particularly in scenarios where facial features are obscured or unreliable.

AUTHOR CONTRIBUTIONS

P. O. Odion: Supervised the research work, provided technical insights, and contributed to the refinement of the methodology to ensure the research meets academic standard. **M. N. Musa:** Designed and implemented the codebase, enhanced the methodology, and contributed to the overall program execution and evaluation. **Z. Z. Ahmad:** Drafted the manuscript, provided technical contributions, and ensured the coherence of the research documentation.

REFERENCES

- Aggarwal, C. C. (2018).** *Neural Networks and Deep Learning*. (Vol. 10, No. 978, p. 3). New York, USA: Springer International Publishing.
- Chollet, F. (2018).** Keras: The python deep learning library. *Astrophysics source code library*, ascl-1806. ui.adsabs.harvard.edu
- Das, S., Ghosh, I. D. and Chattopadhyay, A. (2022).** Deep age estimation using sclera images in multiple environment. *Applied Information Processing Systems*, 93(2), 34.
- Emami, S., and Martínez-Muñoz, G. (2023).** Sequential Training of Neural Networks with Gradient Boosting. *IEEE Access*, 11, 42738-42750.
- Erbilek, M., Fairhurst, M., and 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.
- Bengio, Y., Goodfellow, I., and Courville, A. (2017).** *Deep learning* (Vol. 1). Cambridge, MA, USA: MIT press.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016).** Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- Irhebhude, M. E., Kolawole, A. O., and 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.
- Jang, J., Jeon, S. H., Kim, J., and Yoon, H. (2017).** Robust Deep Age Estimation Methods Using Artificially Generated Image Set. *ETRI Journal*, 39(5), 643-651.
- Musa, M., Sanusi, M. B., Odion, P. O., and Shiti, S. S. (2024).** MRI-Based Brain Tumor Classification using ResNet-50 and Optimized Softmax Regression. *JURNAL INFOTEL*, 16(3), 598-614.
- Thai, L. H., Hai, T. S., and Thuy, N. T. (2012).** Image classification using support vector machine and artificial neural network. *International Journal of Information Technology and Computer Science*, 4(5), 32-38.
- Madan, P., and Madhavan, S. (2020, March 9).** *An Introduction to deep learning*. Retrieved from IBM Developer: <https://developer.ibm.com/technologies/artificialintelligence/articles/an-introduction-to-deep-learning/>
- Murzova, A., and Seth, S. (2020, August 5).** *Otsu's Thresholding with OpenCV*. Retrieved from learnopencv: <https://learnopencv.com/otsu-thresholding-with-opencv/>
- Narejo, S., Pasero, E., and Kulsoom, F. (2016).** EEG Based Eye State Classification using Deep Belief Network and Stacked AutoEncoder. *International Journal of Electrical & Computer Engineering*, 6(6), 3131-3141.
- Odion, P. O., Musa, M. N., and Shuaibu, S. U. (2022).** Age prediction from sclera images using deep learning. *Journal of the Nigerian Society of Physical Sciences*, 787-787.
- Othmani, A., Taleb, A., Abdelkawy, H., and Hadid, A. (2020).** Age Estimation from Faces Using Deep Learning: A Comparative Analysis. *Computer Vision and Image Understanding*, 196, 102961. doi:<https://doi.org/10.1016/j.cviu.2020.102961>
- Priya, T. S. (2021).** ResNet based feature extraction with decision tree classifier for classification of mammogram images. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(2), 1147-1153.
- Rajput, M. R., and Sable, G. S. (2019).** Age Group Estimation from Human Iris. In *Soft Computing and Signal Processing: Proceedings of 2nd ICSCSP 2019 2* (pp. 519529). Springer Singapore.
- Ritchie, N. (2018, October 10).** Evaluating a classification model. *Ritchie Ng*. <https://www.ritchieng.com/machine-learning/>
- Rondeau, J. (2019).** *Deep Learning of Human Apparent Age for the detection of Sexually Exploitative Imagery of Children [Msc Dissertation]*. The University of Rhode Island, United State of America.
- Ronneberger, O., Fischer, P., and Brox, T. (2015, October).** U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention-MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (pp. 234-241). Springer international publishing.
- Rot, P., Vitek, M., Grm, K., Emeršič, Ž., Peer, P., and Štruc, V. (2020).** Deep sclera segmentation and recognition. In *Handbook of vascular biometrics*, 395-432. Springer, Cham.
- Sgroi, A., Bowyer, K. W., and 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.
- Shigueoka, L. S., Mariottoni, E. B., Thompson, A. C., Jammal, A. A., Costa, V. P., and Medeiros, F. A. (2021).** Predicting age from optical coherence tomography scans with deep learning. *Translational Vision Science and Technology*, 10(1), 12-12.
- Su, J., Vargas, D. V., and Sakurai, K. (2019).** One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, 23(5), 828-841.
- Taheri, S., and Toygar, O. (2019).** On the use of DAG-CNN Architecture for Age Estimation with Multi-Stage Features Fusion. *Neurocomputing*, 329, 300-310.
- Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. (2018).** Deep Learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018. Article ID 7068349.
- Wang, W., Yang, Y., Wang, X., Wang, W., and Li, J. (2019).** Development of convolutional neural network and its application in image classification: a survey. *Optical Engineering*, 58(4), 040901-040901.

Yi, D., Lei, Z., and Li, S. Z. (2014, November). Age estimation by multi-scale convolutional network. In *Asian conference on computer vision* (pp. 144-158). Cham: Springer International Publishing.

Zang, H., Cheng, L., Ding, T., Cheung, K. W., Liang, Z., Wei, Z., and Sun, G. (2018). Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network. *IET Generation, Transmission and Distribution*, 12(20), 4557-4567.