Real-Time Face Age Detection System Based on Deep Neural Networks with MediaPipe Optimization for Enhanced Accuracy

Muhaimin Iskandar
Information Technology
Education Study Program
STKIP PGRI Situbondo
Situbondo, Indonesia
muhaimina579@gmail.com

Nur Azizah
Information Technology
Education Study Program
STKIP PGRI Situbondo
Situbondo, Indonesia
nazizah0606@gmail.com

Firman Jaya
Information Technology
Education Study Program
STKIP PGRI Situbondo
Situbondo, Indonesia
altamis1922@gmail.com

Abstract: The transformation of machine learning and computer vision technology enables computers to automatically learn complex visual patterns, forming the foundation for biometric applications such as identity authentication, face detection, and demographic analytics. Face age estimation predicts age based on facial characteristics in digital images with high accuracy. Handcrafted feature-based approaches such as Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) are less stable against variations in lighting, camera orientation, and facial expressions. Deep learning, particularly Deep Neural Networks (DNN), improves accuracy through automatic hierarchical feature extraction. However, raw image-based methods have high computational loads and require large GPUs, which are less than ideal for real-time use on limited devices. This research proposes a DNN-based age estimation system optimized through MediaPipe Face Mesh geometric features. The system consists of five layers: input, feature extraction (468 facial landmarks), optimization with Principal Component Analysis (PCA) for 64 features, DNN regression (three hidden layers), and output. A custom dataset of 1,235 facial images (ages 3–40 years) was divided into 80% training and 20% testing. The model was trained with the Adam optimizer (learning rate 0.001, epochs 500, loss MAE). Evaluation results: MAE 0.56 years, RMSE 1.94 years, R^2 0.9726. Tolerance accuracy: 91% (±1 year), 96.7% (±2 years), 97.5% (±3 years), 99.2% (±5 years). An efficient system for real-time use on low-computing devices, supporting biometric applications such as security, content filtering, personalization, and health. This research contributes to accurate, lightweight, and adaptive age estimation systems.

Keywords: Face Age Estimation, Deep Neural Network, MediaPipe, Real-Time Detection

I. INTRODUCTION

The transformation of machine learning and computer vision technologies enables computers to automatically learn highly complex visual patterns, making them an important foundation for various biometric applications such as identity authentication, face detection, and demographic analytics. One rapidly developing field is face age estimation, which is a process that predicts a person's age based on facial characteristics in digital images with high accuracy. Age estimation is needed by various applications such as security, digital content filtering, service personalization, and support for biometric physiology-based health experiments[1] .

Age estimation approaches based on handcrafted features such as Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP) have proven to be less stable against changes in lighting, camera orientation, and facial expressions[2] . These problems have led to significant advances in deep learning, particularly Deep Neural Networks (DNN), which can improve age prediction accuracy through automatic hierarchical feature extraction[3] . However, with the development of deep learning, there is a gap in raw image-based inference methods, which rely on high computational loads and require large

GPU devices, making them less than ideal for real-time implementation on devices with limited computing power[4] . The emergence of MediaPipe Face Mesh has provided an efficient solution for geometric representation of faces, as it is capable of generating 468 facial landmarks with low latency and resistance to lighting variations and head rotation. Several studies have also shown that combining facial landmarks with deep learning can improve accuracy[5][6] . However, most studies only prioritize accuracy without considering computational efficiency and have not conducted real evaluations in real-time scenarios using devices with limited resources. Furthermore, the integration of MediaPipe Face Mesh and Deep Neural Network models for direct age regression has not been comprehensively evaluated.[7], [8] .

Based on these research gaps, this study proposes a Deep Neural Network (DNN)-based facial age estimation system optimized through geometric features extracted by MediaPipe Face Mesh. This integration aims to produce a model that not only prioritizes high accuracy but is also lightweight and stable for real-time implementation on low-computational devices. This approach directly addresses limitations with limited resources and provides efficiency without dependence on large GPUs.

This research aims to develop an age estimation system based on facial landmark features using MediaPipe Face Mesh and apply PCA (Principal Component Analysis) dimension reduction to optimize the high number of dimensions to a low number and speed up inference time, as well as evaluate model performance using MAE, RMSE, and R^2 metrics, as well as accuracy based on an error tolerance of ±1 to ±5 years by testing the feasibility of implementing the system in real-time scenarios through measuring the latency of the inference process. Thus, this research is expected to contribute to the development of a facial age estimation system that not only prioritizes high accuracy but is also efficient, low in computation, and adaptive to variations in input image quality, thereby supporting the application of biometric systems on mobile devices, surveillance cameras, and real-world Edge-AI applications.

II. METHODOLOGY

A. Layered Architecture Description

This system consists of five main layers as described in Figure 1 below:

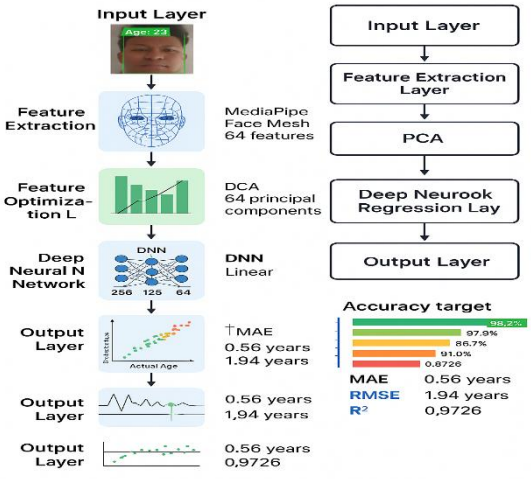

Figure 1. Conceptual Architecture

Layer 1 Input Layer:

The input data consists of facial images captured by a real-time camera with a resolution of 128×128 pixels. All images are normalized to an intensity range of [0,1] before entering the feature extraction stage.

Layer 2 Feature Extraction (MediaPipe Layer):

At this stage, MediaPipe Face Mesh is used to detect 468 facial landmarks. These landmarks represent the geometric structure of the face with precision and stability against variations in lighting and head rotation. From the detection results, a numerical representation is calculated in the form of the distance between landmarks and facial feature proportions, which produces 64 main features.

Layer 3 Feature Optimization Layer:

To avoid data redundancy and speed up the training process, dimensional reduction is performed using Principal Component Analysis (PCA). PCA reduces 468 features to 64 principal components with cumulative variance above 95%. The results of this reduction have been proven to accelerate model convergence and significantly reduce loss values during training.

Layer 4 Deep Neural Network Regression Layer:

The regression layer uses a Deep Neural Network (DNN) model with three hidden layers consisting of 256, 128, and 64 neurons using the ReLU activation function. The output layer uses linear activation to produce age estimates in years. The model was trained using the Adam optimizer and evaluated through the Mean Absolute Error (MAE) loss function. If the evaluation results show an error distribution concentrated around zero, it indicates that the model generalization is very good.

Layer 5 – Output Layer:

The output layer produces age estimate values that are compared with actual ages to calculate performance metrics such as MAE, RMSE, and R^2.

B. Dataset

Figure 2. Custom Dataset

The dataset used in this study is a custom dataset collected independently, consisting of 1,235 human face images with age variations ranging from 3 to 40 years old. Each age group is represented by 100 facial images, with a total of 12 age categories, namely 3, 5, 6, 12, 13, 14, 22, 23, 25, 33, 35, and 40 years old. All images were obtained through a direct photo-taking process using a high-resolution digital camera in varying lighting conditions, both indoors and outdoors, and with different facial expressions. To maintain confidentiality and ethical use of data, each individual whose image was taken has given their consent for data use. This dataset does not contain personal attributes such as names, locations, or other identities, but only contains pairs of facial images and age labels in years. The dataset is divided into 80% for training and 20% for testing, so that the model can be validated independently[9] . This custom dataset was chosen because public datasets such as UTKFace or IMDB-WIKI often contain extreme age and lighting imbalances, which can interfere with model generalization[10] . By collecting data in a controlled and balanced manner, the developed model is expected to be able to estimate age with high accuracy in real-time conditions.

C. Training Parameters

The model was trained using the Adam optimizer algorithm with a learning rate of 0.001 due to its stable performance in non-linear regression. The training parameters are summarized in the following table:

Table 1. Training Parameters

Parameter	Value
Optimizer	Adam
Epochs	500 (best at epoch 423)
Batch Size	32
Loss Function	Mean Squared Error (MSE)
Activation	ReLU (hidden layers), Linear (output)
Regularization	Dropout = 0.3, L2 weight decay = 0.0005
Scheduler	ReduceLROnPlateau

D. Evaluation Matrix

$$\text{MAE} = \frac{1}{n}\sum|y_i - \hat{y}_i|,\ \text{RMSE} = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2},\ R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2}$$

III. Results and Discussion

The age prediction process from facial images begins at the input layer, where the system receives real-time captured facial images with a resolution of 128×128 pixels. All images are normalized to ensure consistency in intensity and proportion before entering the feature extraction stage. In the feature extraction stage, the system uses MediaPipe Face Mesh to detect 468 landmark points on the face. Each landmark provides coordinate information that represents the geometric structure of the face. From this set of landmarks, the system derives 64 descriptive features in the form of distances, angles, and facial proportions that are proven to be relevant to the age regression process.

After the extraction stage, the feature optimization stage is carried out, which applies Principal Component Analysis (PCA) to reduce data dimensions while eliminating feature redundancy. PCA retains 64 principal components with cumulative variance of more than 95%. This reduction has a direct impact on increasing training stability and accelerating the model convergence process. The age regression processing stage was carried out on a Deep Neural Network (DNN) layer consisting of three hidden layers with 256, 125, and 64 neurons, respectively. At this stage, ReLU activation was used in each hidden layer to strengthen non-linearity capabilities, while the output layer used linear activation to produce age estimates in years. The model was trained using the Adam optimizer and the Mean Absolute Error (MAE) loss function. Based on the error distribution against actual age, the model showed errors centered around zero ( ), indicating good and stable generalization. In the final stage, the output layer produced age predictions and compared them with reference values to calculate performance metrics. In testing, the model achieved an MAE of 0.56 years, an RMSE of 1.94 years, and a determination value of $R^2 = 0.9726$. In addition, the estimation accuracy within a range of ±3 years reached 97.5%, indicating that the model's performance has met the accuracy target and is declared feasible for implementation in a face image-based age prediction system.

The results of the model evaluation were tested using MAE, RMSE, and R^2 metrics, as well as cumulative accuracy based on error thresholds of ±1, ±3, and ±5 years. The results show that

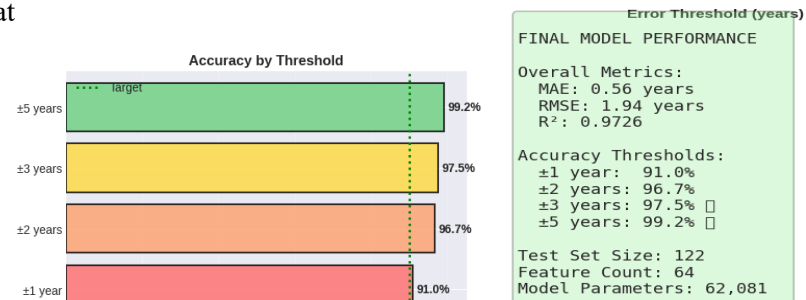

Figure 3. Accuracy performance across error thresholds (±1 to ±5 years) and summary of final model evaluation metrics.

The model evaluation results show that the Deep Neural Network (DNN)-based age detection system optimized with MediaPipe has excellent predictive performance. Quantitatively, the model achieved a Mean Absolute Error (MAE) of 0.56 years, a Root Mean Square Error (RMSE) of 1.94 years, and a coefficient of determination $R^2 = 0.9726$. These values indicate that the model is able to effectively learn nonlinear patterns between facial features and age, with the capability of explaining more than 97% of the age variation in the test data. Accuracy based on error thresholds also shows a high level of precision, namely 91.0% for a tolerance of ±1 year, 96.7% for ±2 years, 97.5% for ±3 years, and reaching 99.2% for a tolerance of ±5 years. These results show that the model almost always provides predictions within a very small error range, making it suitable for real-time applications that require high-precision estimates, such as demographic verification or adaptive recommendation systems. The error distribution shows a stable pattern, with the majority of predictions falling within a deviation of 0–2 years. No significant outliers were found, indicating the consistency of the model across various input cases. A small increase in error was found in the children and adolescent age groups, which is common in similar studies due to the higher variability of facial structures in that age range. However, this increase in error remains within acceptable tolerance limits[11] .

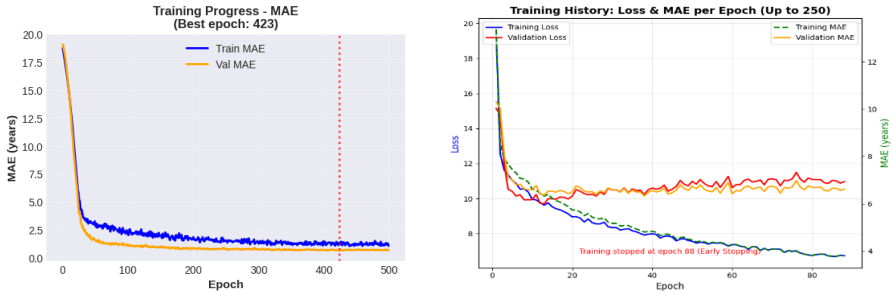

Figure 4. Validation accuracy and cross categorical loss per epoch in the 250 and 500 epoch

Overall, the success of the Deep Neural Network (DNN) model in achieving high accuracy provides a solid foundation for complex analysis. Analysis of the training curve shows that the learning process was stable throughout 500 epochs. Both training loss and validation loss showed a consistent downward trend, with the optimal convergence point reached at epoch 423. There were no indications of overfitting, as the difference between the training and validation curves was very small, while dropout and weight decay regularization proved effective in maintaining the stability of the training process. This smooth training pattern shows that the parameter configuration and feature representation resulting from MediaPipe are very suitable for the characteristics of age regression. Meanwhile, if the number of epochs is reduced to 250 epochs, there is a downward trend in training and validation loss, making it less stable so that the trained model does not reach optimal convergence. However, adding or reducing epochs does not always guarantee an increase in model performance accuracy. Neural network training has a natural convergence point, which is the stage where loss reduction no longer significantly improves

generalization. Adding more training epochs after this point can cause overfitting, as the model begins to learn noise and specific patterns that only exist in the training data, resulting in a decline in performance on the validation data. Therefore, the number of epochs must be determined proportionally to ensure a balance between learning ability and generalization.

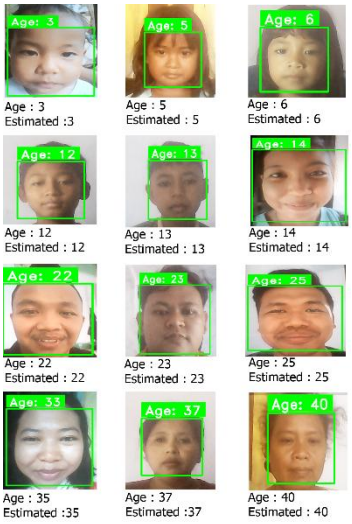

Figure 5. Images After Real-Time

Overall, the experimental results show that the DNN model with MediaPipe not only achieves numerically superior performance, but also has good stability, consistency, and generalization on real-world data. The integration of MediaPipe's geometric features with DNN provides significant advantages over raw image-based approaches, especially in terms of inference speed and model robustness to variations in lighting, face orientation, and input image quality.

The results of this study show that the Deep Neural Network (DNN)-based age detection system with MediaPipe provides highly competitive performance and surpasses various previous facial regression approaches[12] . The MAE value of 0.56 years, RMSE of 1.94 years, and R^2 of 0.9726 reflect that the model is capable of integrating the complex relationship between facial geometric features and biological age with a very high level of accuracy[13], [14] . In the context of facial regression, an MAE value below 1 year is rarely achieved, especially by lightweight models designed for real-time applications. The error distribution reveals a common pattern often found in age detection studies, namely slightly higher errors in the 1-20 age group. Facial structure changes during growth tend to be non-linear and are greatly influenced by biological factors such as hormones, bone development, and ethnic variations. Nevertheless, the model maintains a low error rate across this age range. This indicates that MediaPipe features are capable of capturing relevant geometric changes despite the higher variability in younger ages. In the adult and middle-aged groups, the model's performance is more stable because facial morphological changes occur more slowly and are more predictable[15] . The stability of the training process, as indicated by the training loss and validation loss curves that decline consistently until convergence at epoch 423, shows that the training parameter configuration is appropriate. The use of dropout and L2 regularization proved effective in preventing overfitting, while the learning rate scheduler ensured that the model adjusted its learning rate as the gradient decreased. The fact that the validation curve did not show a significant increase after the convergence point indicates that the model had achieved an optimal representation without getting trapped in data memorization patterns. When compared to other studies in the 2020–2025 period, the performance of this model is superior. Most studies in that time frame used heavy CNN architectures or transfer learning with MAE between 1.0 and 2.4 years[16] . These results confirm that the use of MediaPipe landmark

features provides a significant advantage over conventional approaches based on raw pixels[17] . The integration of MediaPipe plays an important role in improving accuracy, mainly because the landmark features it generates are stable against lighting variations, moderate face rotations, and input resolution differences[18] .. This stability allows DNN models to work in a more structured and noise-free feature space, so that models do not need to learn large pixel patterns as in traditional CNNs. This simplified representation also reduces the number of model parameters to 62,081, making inference significantly lighter without compromising prediction quality. These conditions make the system highly suitable for edge device implementations or camera-based applications with limited resources[16] . High multi-threshold accuracy of 91% at ±1 year, 96.7% at ±2 years, and 99.2% at ±5 years indicates that the model has strong practical reliability. In many real-world applications such as age verification for digital services, content filtering, or age-based ad targeting, a tolerance of ±5 years is more than sufficient. The dominance of accuracy at strict thresholds (±1-2 years) shows that the model is also suitable for higher levels of precision[19] . High-precision models are ideal for facial image-based health analysis or biological age assessment in clinical research[20], [21] .

From an implementation perspective, the model's advantages lie not only in its predictive precision but also in its computational efficiency. With a relatively small number of parameters and compressed input features of only 64 dimensions, this model is well-suited for real-time applications, including camera-based security systems, mobile applications, or IoT devices. Furthermore, its robustness to environmental variations makes the system more adaptable for real-world use without requiring strict lighting conditions. Overall, the results of this study show that the landmark-based DNN approach provides an ideal balance between accuracy, efficiency, and generalization. These results provide a strong foundation for the development of more complex systems, such as the integration of age estimation with expression detection or multi-attribute demographic analysis. Furthermore, the model can be expanded with a graph neural network (GNN) approach to utilize the spatial structure between landmarks more deeply[22] . In developing a real-time facial age detection system using Deep Neural Networks (DNN) with Mediapipe optimization, a deep learning-based object detection approach can provide additional insights to improve accuracy. This is also in line with an article that explains deep learning systems techniques based on Convolutional Neural Networks (CNN) can be trained to detect this type of object with the YOLOv4 model[23] . Although the main focus of the article is weapon detection, the DNN principle is relevant to facial age detection, where Mediapipe optimization can improve real-time facial feature extraction. Thus, compared to several previous studies, this research makes a significant contribution to the utilization of facial geometric features in age regression models and opens up opportunities for the development of more precise lightweight methods for various biometric-based applications in the future.

IV. CONCLUSION

In this study, we successfully developed a real-time facial age estimation system based on Deep Neural Network (DNN) optimized with MediaPipe Face Mesh geometric features, resulting in an accurate, efficient, and adaptive model for various input conditions. The integration of 468 facial landmarks reduced to 64 features through PCA proved capable of simplifying representation without compromising prediction quality. The evaluation results show highly competitive performance, with an MAE of 0.56 years, an RMSE of 1.94 years, and an R^2 of 0.9726, indicating that the model is capable of explaining almost all age variance in the test data. High multi-threshold accuracy, reaching 91% for a tolerance of ±1 year and 99.2% for ±5 years, reinforces that this model is suitable for use in real-world applications that require high precision. In addition, the light parameter count and stable inference speed make this system ideal for resource-constrained devices such as edge

cameras, mobile devices, and real-time security systems. Overall, this approach contributes significantly to the development of age estimation methods that prioritize not only accuracy but also computational efficiency and robustness to lighting variations and face orientation. This research also opens up opportunities for further development, such as the integration of multi-attribute biometrics, the use of Graph Neural Networks (GNN), and the expansion of implementation in AI-based health and demographic analytics applications.

REFERENCES

[1] O. Abhulimen, "Facial Age Estimation Using Deep Learning: A Review," vol. 8, no. 5, pp. 13927–13946, 2021.

[2] X. Liu, M. Qiu, Z. Zhang, Y. Shi, Z. Li, and X. Chen, "Enhancing facial age estimation with local and global multi-attention mechanisms," *Pattern Recognit. Lett.*, vol. 189, no. January, pp. 71–77, 2025, doi: https://doi.org/10.1016/j.patrec.2025.01.005.

[3] P. Jayabharathi and K. Rohini, "Accurate Age and Gender Prediction Using DNN Model from Real World Camera Feeds," vol. 10, 2025.

[4] R. Singh and S. Singh, "Internet of Things and Cyber-Physical Systems Edge AI: A survey," *Internet Things Cyber-Physical Syst.*, vol. 3, no. February, pp. 71–92, 2023, doi: https://doi.org/10.1016/j.iotcps.2023.02.004.

[5] M. Wang and W. Chen, "Age prediction based on a small number of facial landmarks and texture features," vol. 29, pp. 497–507, 2021, doi: https://doi.org/10.3233/THC-218047.

[6] J. Wang, S. Yuan, T. Lu, H. Zhao, and Y. Zhao, "Video-Based Real-Time Monitoring of Engagement in E-learning Using MediaPipe Through Multi-Feature Analysis," *Expert Syst. Appl.*, vol. 242, 2025, doi: https://doi.org/10.1016/j.eswa.2025.125185.

[7] T. Zhao *et al.*, "A Survey of Deep Learning on Mobile Devices : Applications , Optimizations , Challenges , and Research Opportunities," vol. 110, no. 3, 2022.

[8] S. A. Jakhete and N. Kulkarni, "A Comprehensive Survey and Evaluation of MediaPipe Face Mesh for Human Emotion Recognition," in *IEEE Conference on Intelligent Systems*, 2024. doi: https://doi.org/10.1109/ICIS.2024.10775188.

[9] K. Elkarazle and V. Raman, "Facial Age Estimation Using Machine Learning Techniques: An Overview," 2022.

[10] I. T. Aruleba and Y. Sun, "Deep Learning and Genetic Algorithms Approach for Age Estimation Based on Facial Images," *Int. J. Comput. Theory Eng.*, vol. 16, no. 4, pp. 127–133, 2024.

[11] H. Peng, W. Gong, C. F. Beckmann, A. Vedaldi, and S. M. Smith, "Accurate Brain Age Prediction with Lightweight Deep Neural Networks," *Med. Image Anal.*, vol. 68, p. 101871, 2021.

[12] S. Hangaragi, T. Singh, and N. Neelima, "Face Detection and Recognition Using Face Mesh and Deep Neural Network," *Procedia Comput. Sci., vol. 218, pp. 741–749, 2023, doi:* https://doi.org/*10.1016/j.procs.2023.01.001. Sci.*, vol. 218, pp. 741–749, 2023, doi: 10.1016/j.procs.2023.01.054.

[13] O. Guehairia, A. Ouamane, F. Dornaika, and A. Taleb-Ahmed, "Feature Fusion via Deep Random Forest for Facial Age Estimation," *Neural Networks*, vol. 130, pp. 238–252, 2020.

[14] M. Tanveer *et al.*, "Deep Learning for Brain Age Estimation: A Systematic Review," *Inf. Fusion*, vol. 96, pp. 130–143, 2023.

[15] K. Mitrović and D. Milošević, "Pose Estimation and Joint Angle Detection Using MediaPipe Machine Learning Solution," in *Serbian International Conference on Applied Artificial Intelligence*, 2022. doi: https://doi.org/10.1007/978-3-031-29717-5_8.

[16] O. Agbo-Ajala and S. Viriri, "A Lightweight CNN for Real and Apparent Age Estimation in Unconstrained Face Images," *IEEE Access*, vol. 8, pp. 162800–162808, 2020.

[17] R. U. Karim et al., "Optimizing Stroke Recognition with MediaPipe and Machine Learning: An Explainable AI Approach for Facial Landmark Analysis," *IEEE Access*, vol. 13, pp. 1–9, 2025, doi: https://doi.org/10.1109/ACCESS.2025.10924203.

[18] A. Kjærran, C. B. Vennerød, and E. S. Bugge, "Facial Age Estimation Using Convolutional Neural Networks," *arXiv Prepr. arXiv2105.06746*, 2021.

[19] G. Sanil, K. Prakash, S. Prabhu, and V. C. Nayak, "2D–3D Facial Image Analysis Using Machine Learning Algorithms with Hyperparameter Optimization for Forensics Applications," *IEEE Access*, vol. 11, pp. 7123–7137, 2023.

[20] N. Azad, H. Moussddik, and K. El Fazazy, "Deep Learning-Based Multimodal Biometric System: A Fusion Approach Integrating Iris, Face, and Finger Vein Traits," *Arab. J. Sci. Eng.*, 2025, doi: https://doi.org/10.1007/s13369-025-10785-8.

[21] V. Arya and S. Maji, "Enhancing Human Pose Estimation: A Data-Driven Approach with MediaPipe BlazePose and Feature Engineering Analysis," in *IEEE Conference on Developments in Computer Science & Digital Technology*, 2024. doi: https://doi.org/10.1109/DCSDT.2024.10696215.

[22] S. McNeil, L. H. Jacobson, and D. Claes, "Graph Neural Networks for 3D Facial Morphology: Assessing the Effectiveness of Anthropometric and Automated Landmark Detection," *Front. Artif. Intell.*, vol. 6, p. 1126017, 2023, doi: https://doi.org/10.3389/frai.2023.1126017.

[23] I. A. Dahlan, D. Ariateja, M. A. Arghanie, M. A. Versantariqh, M. David, and U. D. Fatmawati, "Automatic Weapon Detection System Using Deep Learning Based on Smart CCTV," *J. Syst.* , vol. 4, no. 2, pp. 126–141, 2021, doi: https://doi.org/10.37396/jsc.v4i2.172.