International Journal of Multiphysics
Volume 18, No. 3, 2024
ISSN: 1750-9548

Development of a Deep Learning Model for Age Estimation Utilizing Multimodal Feature and Augmented Analysis

Rupali Jumbadkar[1*], Vipin Kamble[2], Mayur Parate[3]

[1*]PhD Scholar, Dept. Electronics and Communication, Indian Institute of Information Technology, Nagpur, India

[1*]Subject Matter Expert, Dept. of Electronics, Larsen & Toubro Edutech, Nagpur, India

[2]Assistant Professor, Dept. Electronics and Communication, Visvesvaraya National Institute of Technology, Nagpur, India

[3]Assistant Professor, Dept. Electronics and Communication, Indian Institute of Information Technology, Nagpur, India

E-mail: dtea20ece001@iiitn.ac.in[1*] (corresponding author); vipinkamble@ece.vnit.ac.in[2]; mparate@iiitn.ac.in[3]

ABSTRACT

Age estimation from facial images is crucial in security, healthcare, and entertainment for applications like age-restricted content filtering and targeted advertising. However, this task is challenging due to pose, illumination, and aging variations. Traditional methods using handcrafted features often lack discriminative power. Deep learning, with its ability to learn complex features, has sparked interest in age estimation. This paper presents a deep learning-based approach for age estimation from frontal face images. It focuses on facial components like eyes, nose, and mouth to capture age-related changes. Augmentation techniques enhance robustness against variations, and a Binary Cascaded CNN with binary weights and activations reduces model complexity. The model transforms augmented components into multimodal features, allowing it to discern age-related changes across domains. Evaluated on FGNET and IMDB-WIKI datasets, the model achieves 99.5% accuracy with an MAE of 1.26 across all age groups. This high accuracy demonstrates the model's effectiveness and potential for real-world age estimation scenarios in diverse applications.

Keywords: Age Estimation, Deep Learning, Augmentation Technique, Binary Cascaded CNN, Accuracy, MAE, Computer Vision, Human-computer Interaction

1.0 INTRODUCTION

Estimating age from facial images is a complex task with diverse applications in security, healthcare, and entertainment. Traditional methods relying on handcrafted features lack discriminative power and struggle with image variations. Deep learning, particularly Age Net and deep conditional distribution learning (DCDL) [1, 2, 3, 4], has significantly enhanced age estimation accuracy. However, current deep learning models still grapple with challenges, such as their inability to capture age-related changes in facial components, limited performance under varying conditions, and high computational costs [5, 6].

To address these issues, we propose an efficient deep learning model for age estimation from frontal faces. Our model, utilizing multimodal feature representation and multicomponent augmented analysis, aims to overcome existing limitations. By analyzing facial components like eyes, nose, and mouth, we aim to capture age-related changes in different facial regions. Additionally, augmenting facial components through various operations enhances the model's resilience to pose, illumination, and occlusion variations.

This paper has two primary objectives. Firstly, we aim to design an efficient deep learning model for age estimation from

frontal faces, capable of capturing age-related changes in different facial components and resilient to variations. Secondly, we aim to evaluate the proposed model's performance on FGNET-like datasets, comparing it with state-of-the-art age estimation models to showcase its effectiveness. Key contributions include the use of multimodal feature representation and multicomponent augmented analysis, enhancing accuracy and robustness. Introducing the Binary Cascaded CNN reduces model complexity and improves efficiency. Experimental results demonstrate the proposed model's superior performance, achieving 99.5% accuracy with an MAE of 1.12 for all age groups on FGNET-like datasets and samples, outperforming existing age estimation models.

2.0 LITERATURE REVIEW

Age estimation from facial images has been a focal point in computer vision and biometrics research for an extended period. Conventional age estimation methods, relying on handcrafted features like texture, wrinkles, and skin color, face challenges in discriminative power and robustness to facial image variations, as seen in Enhanced Cycle Generative Adversarial Network (ECGAN) [7, 8, 9, 10]. Recent years witnessed significant advancements in age estimation accuracy through deep learning models that autonomously learn discriminative features from facial images [11,12, 13, 14].

Deep learning models for age estimation fall into two main approaches: regression-based models and classification-based models [15, 16, 17, 18]. Regression-based models predict age as a continuous value, whereas classification-based models categorize age into distinct groups [19, 20]. Early models like AgeNet and VGG-Face employed a regression-based approach to predict age as continuous values and sets [21, 22, 23, 24]. However, regression-based models encounter challenges such as the need for extensive training data and difficulties in handling outliers and age-related variations in facial images and samples [25, 26, 27].

Addressing these challenges, researchers have proposed classification-based models categorizing age into groups, including popular models like Adience, AgeDB, and MegaAge sets [28, 29, 30]. These models typically employ a CNN-based architecture with diverse loss functions and training strategies to enhance accuracy levels [31, 32, 33,34,35].
Despite advancements, existing deep learning models for age estimation still grapple with limitations. One limitation is the failure to capture age-related changes in different facial components. Some researchers suggest models analyzing facial components like eyes, nose, and mouth to capture age-related changes in different regions of face images [36, 37, 38]. For instance, a deep learning model was invented using an attention mechanism to focus on various facial components for age estimations [39, 40, 41].

Another limitation is the restricted performance under pose, illumination, and occlusion variations and sets [42, 43, 44]. To address this, researchers propose data augmentation techniques, such as random cropping, flipping, and rotating, to enhance the model's robustness to facial image variations [45, 46, 47].

Motivated by these limitations and bioinspired computing work in [48, 49, 50], we introduce an efficient deep learning model for age estimation from frontal faces. Our model aims to bridge gaps in existing methods by analyzing facial components, including eyes, nose, and mouth, to capture age-related changes in different face regions. Furthermore, we augment facial components using various operations to improve the model's robustness to variations in pose, illumination, and occlusions.
In conclusion, while deep learning models have substantially improved age estimation accuracy, existing models still encounter limitations such as the inability to capture age-related changes in different facial components and restricted performance under pose, illumination, and occlusion variations. Our proposed model addresses these challenges by analyzing facial components and applying various augmentations, showcasing its effectiveness and potential for real-world age estimation scenarios.

3.0 AGE DATABASE

The presence of comprehensive databases plays a crucial role in shaping the evolution of analyses, facilitating swift research activities, and providing a basis for relative estimations. Particularly in cases where constructing a database proves to be a time-consuming process, publicly available datasets offer a significant advantage for experimental analyses. Over the past few decades, various databases for facial images have been curated and disseminated, addressing challenges in face recognition and age estimation.

In our experiments, we utilized the FG-NET (Face and Gesture Recognition Network) Aging Database [51] and the IMDB-WIKI database [52]. FG-NET, supported by the European Union's 5th Framework Program on Information Society

Technologies, comprises 1002 images from 82 individuals spanning newborns to 69-year-olds, with a higher concentration of ages up to 40 years. These images exhibit considerable variations in resolution, quality, illumination, viewpoint, and expression. Some images also include challenges such as spectacles, facial hair, and hats. Each image is annotated with 68 landmark points, and details about age, gender, expression, pose, image quality, and presence of barriers (mustaches, beards, hats, or spectacles) are documented. The FG-NET-AD database is commonly employed for age estimation and is publicly accessible for experimental use. On the other hand, IMDB-Wiki stands out as the most extensive publicly available face image dataset, incorporating gender and age labels. It compiles information from profiles of 100,000 popular actors on IMDB and Wikipedia, resulting in a vast collection of 523,051 images. This dataset is openly accessible for both training and testing purposes.

4.0 COMPONENTS OF FACIAL AGE ESTIMATION

Upon reviewing existing models for age detection based on facial images, it is evident that estimating age from such images poses significant challenges due to factors like pose variations, illumination changes, occlusions, and aging patterns. Traditional techniques for facial age estimation typically rely on handcrafted features such as texture, shape, and appearance, which often lack discriminative power and robustness against image variations. With the advent of deep learning, there has been a growing interest in utilizing deep neural networks for age estimation from facial images, as they can extract complex and discriminative features, thereby enhancing the accuracy and robustness of age estimation models. The proposed approach integrates the following components to facilitate age estimation.

1. Haar Cascade Classifier: The Haar Cascade Classifier is a machine learning-based method utilized for object detection in images, including facial components like eyes, nose, and mouth in the context of age estimation from facial images. Initially, the input image is converted to grayscale to simplify processing, reducing it to a single channel containing intensity information. Haar-like features are then extracted from the pre-processed grayscale image, defined as rectangular regions with varying intensity values and pixel sets. These features facilitate the identification of intensity pattern variations, computed by subtracting the sum of pixel values in a black rectangle from those in adjacent white rectangles. The estimation of Haar features is performed using Eq. (1).

$$\boldsymbol{f(x,y,w,h) = \sum_{i=1}^{N}\left(p(i) * s(i)\right)} \quad \textbf{(1)}$$

Here, ***p(i)*** denotes the value of the pixel in rectangle ***i***, while ***s(i)*** signifies the weight assigned to rectangle ***i***, involving ***N*** components. In the context of face detection, a Haar cascade classifier is employed, integrating various Haar features. The classifier is symbolized by Eq. (2).

$$\boldsymbol{C(x) = \sum_{i=1}^{N}\left(\alpha(i) * f(i,x)\right)} \quad \textbf{(2)}$$

Here, ***α(i)*** represents the weight allocated to feature ***i***, and ***f(i, x)*** denotes the output of Haar feature ***i*** for the input image ***x*** in real-time scenarios.

2. Detection of Eyes, Nose, and Mouth: Utilizing Haar-like features, the detection of eyes is performed, as illustrated by Equation (3).

$$\boldsymbol{f(x,y,w,h) = \sum_{i=1}^{N}\left(p(i) * s(i)\right) * w(e)} \quad \textbf{(3)}$$

Where ***w(e)*** denotes the weight assigned to the given facial components, with equations for eye, nose, and mouth detection resembling that of face detection but utilizing different Haar-like features.

3. Particle Swarm Optimization (PSO) is a metaheuristic optimization method inspired by the collective behavior of birds or fish, used for feature selection and optimization. PSO begins by selecting a set of stochastic augmentation operations from a predefined set (NA), such as Shift, Rotate, Brightness, Scale, and Zoom, which define the initial positions of particles in the search space. The fitness of each particle, representing a combination of selected operations, is evaluated based on its ability to accurately classify augmented images. PSO then iteratively updates particle fitness levels and configurations using cognitive and social learning rates, allowing particles to adjust their positions based on their individual best-known solution and the best-known solution across the entire swarm. This iterative process continues for a predetermined number of iterations to initiate the augmentation process.

To commence the augmentation process, a series of ***N*** stochastic augmentation operations is chosen using Equation (4).

$$N = STOCH(L_c * NA, NA) \quad (4)$$

Where ***NA*** represents the set of augmentation operations that can be applied to the image, and

$$NA \in (Shift,\ Rotate,, Brightness, Scale, Zoom\ Levels) \quad (5)$$

4. Long Short-Term Memory (LSTM): LSTM, a form of recurrent neural network (RNN), is employed for extracting features from augmented images, aiming to handle sequential data effectively. LSTM operates by extracting various types of features from the augmented images, persisting in the process until the features between iterations converge. The LSTM Model begins by deriving initializations ***(i)***, intermediate features ***(f)***, temporal outputs ***(o)***, and convolutional ***(c)*** feature sets using equations (5), (6), (7), and (8) respectively.

$$i = var(x(in) * U^i + h(t-1) * W^i) \quad (5)$$

$$f = var\left(x(in) * U^f + h(t-1) * W^f\right) \quad (6)$$

$$o = var(x(in) * U^o + h(t-1) * W^o) \quad (7)$$

$$c = tanh(x(in) * U^g + h(t-1) * W^g) \quad (8)$$

In the equation, ***x(in)*** denotes the pixels of the augmented image, with ***U*** and ***W*** serving as constants within the LSTM model, while ***h*** represents an initial kernel metric utilized for feature augmentation.

5. Gated Recurrent Unit (GRU): GRU, a variant of recurrent neural networks akin to LSTM but with a streamlined architecture, serves to refine the features extracted by LSTM. By processing these features, GRU enables the model to focus on pertinent information while filtering out less significant details. GRU calculates a forgetting factor ***(z)*** and a retaining factor ***(r)*** using equations (9) and (10) as follows:

$$z = var(W^z * [h(out) * T(out)]) \quad (9)$$

$$r = var(W^r * [h(out) * T(out)]) \quad (10)$$

Where, ***W*** is a value utilized in GRU operations.

6. Genetic Algorithm (GA): Utilized within the optimization process, GA aids in refining feature selection and minimizing redundancy within the extracted facial image features. Employing evolutionary principles, GA iteratively evolves and refines feature sets to enhance age estimation accuracy based on facial image data. By stochastically selecting and combining the most pertinent features, GA facilitates the improvement of the model's age estimation capabilities. The GA process involves the stochastic selection of ***N*** features from the extracted feature set, as described by Eq. (11).

$$N = STOCH(LR * Nf, Nf) \quad (11)$$

Where ***STOCH*** denotes a stochastic process utilized for generating sets of numbers, and ***Nf*** represents the total features extracted through the fusion operations of LSTM and GRU.

7. Binary Cascaded Neural Network (BCNN): Employed for classification purposes, BCNN constitutes a neural network architecture. Within the framework of age estimation from facial images, BCNN is deployed to classify features and predict age by processing features through convolutional layers, pooling, and dropout mechanisms. BCNN derives 1D Convolutional Features as outlined in Eq. (12).

$$Conv(1D) = \sum_{a=-\frac{m}{2}}^{\frac{m}{2}} xout(i-a) * ReLU\left(\frac{m}{2} + a\right) \quad (12)$$

Utilizing these features along with various Max Pooling and Dropout layers, the model can discern highly diverse feature sets.

5.0 PROPOSED APPROACH FOR AGE ESTIMATION

Inspired by the components, Figure 1 illustrates the flow of the proposed method, employing a deep learning model for estimating age from frontal face images. The approach entails analyzing facial components, such as eyes, nose,

and mouth, to capture age-related changes across different regions of the face. These components undergo augmentation through operations like rotation and shifting to enhance the model's resilience against pose variations, lighting conditions, and occlusions. The augmented components are then transformed into multimodal features and individually classified using an innovative Binary Cascaded CNN model, which leverages binary weights and activations to reduce complexity and enhance efficiency.

To execute these tasks, the proposed model initially gathers multiple sets of facial images and employs a Haar Cascade frontal face detector for facial component identification. Haar Cascade, a machine learning-based object detection approach, utilizes pre-trained classifiers to recognize specific object features, such as eyes, nose, and mouth for face detection scenarios.

The process of face detection using Haar Cascade is as follows:

1. Pre-processing: The input image undergoes grayscale transformation, followed by noise reduction and undesirable detail removal through smoothing and filtering operations.
2. Feature extraction: Haar-like features are extracted from the pre-processed images, defined as rectangular regions with varying intensity values and pixel sets. These features are calculated by subtracting the sum of pixel values in black rectangles from that of white rectangles.
3. Classifier training: A classifier is trained using positive (images of faces) and negative (images without faces) samples to distinguish between the two categories based on the extracted features.
4. Face Detection: The trained classifier is employed to identify faces within the input images. This involves applying the classifier across various regions of the image using a sliding window technique. When a face is detected by the classifier, the corresponding region is flagged as a potential face area. Otherwise, the process continues with new regions subjected to the same operations.
5. Non-Maximum Suppression: To address potential overlapping detections from the sliding window approach, a non-maximum suppression algorithm is utilized. This algorithm eliminates redundant detections, retaining only the most pertinent ones for real-time applications.

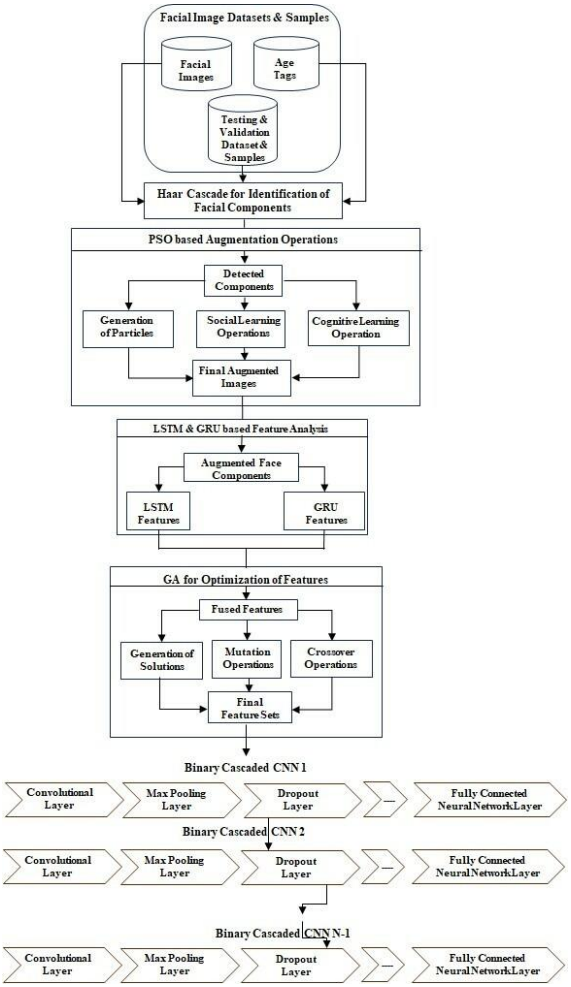

Fig. 1: Design of the proposed model for identification of age from facial components

The mathematical model utilized for each task is delineated as follows:
1. Face Detection: Haar features are computed using Eq. (1). For face detection, a Haar cascade classifier is employed, integrating multiple Haar features as depicted in Eq. (2).
2. Eye, Nose & Mouth Detection: Haar-like features are utilized to detect eyes, as illustrated in Eq. (3). The equations for eye, nose, and mouth detection mirror those for face detection, albeit employing distinct Haar-like features. These components are visually represented in Fig. 2, showcasing the facial components of test subjects. To intelligently augment these components, a Particle Swarm Optimizer (PSO) is employed, initiated by selecting a set of N Stochastic augmentation operations per Equation 4.

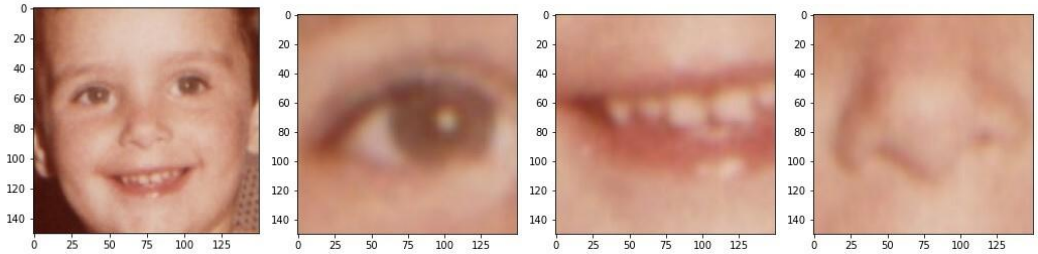

Fig. 2: Face, Eye, Mouth and Nose Components of Test Subjects

The augmented images are classified into N age categories through a tailored CNN, illustrated in Figure 3, employing convolutional operations to extract features, Max Pooling operations to select highly variant features, Drop Out operations, and Fully Connected Neural Networks (FCNNs) to identify these categories.

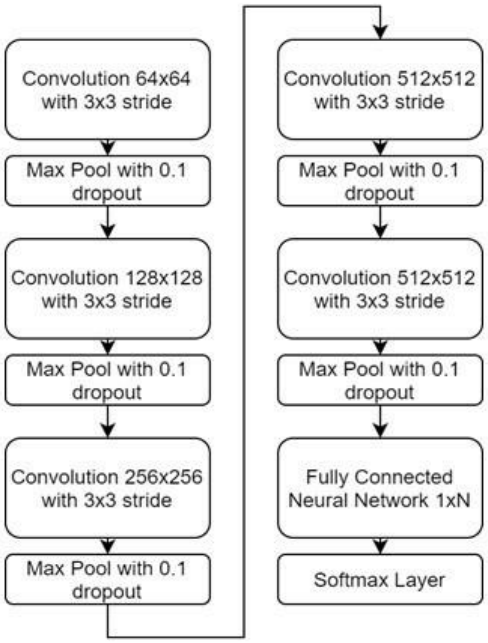

Fig. 3: CNN Model for estimation of augmentation performance levels

Convolutional operations, as described in Equation (13), aid in identifying 2D window-based feature sets.

$$\boldsymbol{Conv(out)} = \sum_{\boldsymbol{a}=-\frac{\boldsymbol{m}}{2}}^{\frac{\boldsymbol{m}}{2}} \sum_{\boldsymbol{b}=-\frac{\boldsymbol{n}}{2}}^{\frac{\boldsymbol{n}}{2}} \boldsymbol{I(aug, i-a, j-b)} * \boldsymbol{ReLU}\left(\frac{\boldsymbol{m}}{2}+\boldsymbol{a}, \frac{\boldsymbol{n}}{2}+\boldsymbol{b}\right) \quad \textbf{(13)}$$

In this context, $\boldsymbol{m}$ and $\boldsymbol{n}$ denote the window sizes, while $\boldsymbol{a}$ and $\boldsymbol{b}$ indicate the stride sizes. Augmented images are represented by $\boldsymbol{I(aug)}$, and ReLU denotes an activation function employed to preserve positive feature sets. Utilizing these features, the model incorporates an effective Max Pooling layer, which computes feature variance as per Equation (14).

$$\boldsymbol{fth} = \sqrt{\frac{1}{\boldsymbol{Nf}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{Nf}}\left(\boldsymbol{f(i)} - \sum_{\boldsymbol{j=1}}^{\boldsymbol{Nf}}\frac{\boldsymbol{f(j)}}{\boldsymbol{Nf}}\right)^{\boldsymbol{2}}} \quad \textbf{(14)}$$

In this equation, ***f(i)*** symbolizes the extracted features, and ***Nf*** denotes the total number of extracted features. Features with ***Conv(out)*** greater than the threshold value (fth) are preserved, while those below are excluded from the feature selection process. This iterative process is applied across all layers, ultimately classifying the final selected features into distinct age groups according to Equation (15).

$$\boldsymbol{c(out)} = \boldsymbol{SoftMax}\left(\sum_{\boldsymbol{i=1}}^{\boldsymbol{Nf}}\boldsymbol{f(i)} * \boldsymbol{w(i)} + \boldsymbol{b(i)}\right) \quad \textbf{(15)}$$

In this context, ***w(i)*** and ***b(i)*** represent the weights and biases corresponding to different age categories. Upon estimating all classes, the fitness of the particle is determined using Equation (16).

$$\boldsymbol{f(p)} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\frac{\boldsymbol{C(i)}}{\boldsymbol{T(i)}} \quad \textbf{(16)}$$

In this equation, ***C*** represents the total number of correctly classified images, and ***T*** represents the total number of augmented images utilized in the classification process. This fitness level is computed for all ***NP*** particles and is denoted as the particle's best fitness level. This iterative process is conducted for ***NI*** iterations, and the particle fitness is adjusted using Equation (17).

$$\boldsymbol{A(New)} = \boldsymbol{A(Old)} + \boldsymbol{LC} * \boldsymbol{S(1)} * |\boldsymbol{A(Old)} - \boldsymbol{PBest}| + \boldsymbol{LS} * \boldsymbol{S(2)} * |\boldsymbol{A(Old)} - \boldsymbol{GBest}| \quad \textbf{(17)}$$

In this context, ***A(Old)*** and ***A(New)*** represent the old and new fitness levels respectively, whereas ***S(1)*** and ***S(2)*** denote stochastic numbers. ***LC*** and ***LS*** correspond to the cognitive and social learning rates of the particles. Additionally, GBest is determined using Equation (18).

$$\boldsymbol{GBest} = \boldsymbol{Max(PBest)} \quad \textbf{(18)}$$

Throughout these iterations, adjustments are made to particle fitness levels, and various internal stochastic operations are generated to produce different configurations. Following the completion of the PSO process, particles with the highest fitness levels are chosen, and their configurations are employed to discern different augmentation operations. Subsequently, these augmented images are transformed into multimodal feature sets using a combination of Long-Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) operations. The fusion of LSTM and GRU processes is illustrated in Fig. 4, where the output of LSTM serves as input to GRU for the identification of augmented feature sets.

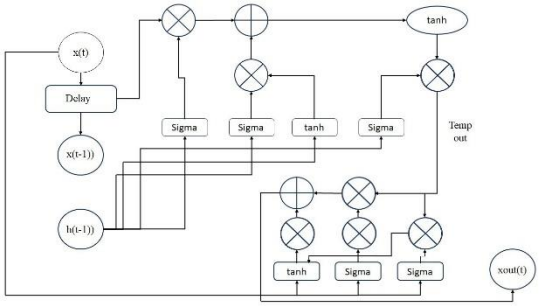

Fig. 4: Fusion of LSTM & GRU for identification of multimodal features

The LSTM model first derives a collection of initializations ***(i)***, intermediate features ***(f)***, temporal outputs ***(o)***, and convolutional feature sets ***(c)*** using equations (5), (6), (7), and (8). These features collectively contribute to the generation of a temporal output feature and the adjustment of the kernel metric, as illustrated by equations (19) and (20).

$$\boldsymbol{T(out)} = \boldsymbol{var}(\boldsymbol{f} * \boldsymbol{x(in, t-1)} + \boldsymbol{i} * \boldsymbol{c}) \quad \textbf{(19)}$$

$$\boldsymbol{h(out)} = \mathbf{tanh}(\boldsymbol{T(out)}) * \boldsymbol{o} \quad \textbf{(20)}$$

These features are input into GRU, which calculates a forgetting factor (z) and a retaining factor (r) using equations (9) and (10) respectively. These factors are then employed to update the kernel metric via equation (21) and derive the final feature metric through equation (22).

$$\boldsymbol{h(t) = tanh(W * [r * h(out) * T(out)])} \quad \textbf{(21)}$$

$$\boldsymbol{xout = (1 - z) * h(t) + z * h(out)} \quad \textbf{(22)}$$

This process persists until the inter-iteration features reach a state of approximate equality, as demonstrated by equation (23) below.

$$\boldsymbol{\frac{xout(New)}{xout(Old)} \leq W} \quad \textbf{(23)}$$

Because of the recurrent application of LSTM and GRU operations, the model can generate numerous feature sets, some of which may include inherent redundancies. These redundancies are minimized through the utilization of an efficient variance maximization Genetic Algorithm (GA), depicted by equation (11). The fitness of the solution is assessed based on these features using equation (24).

$$\boldsymbol{f = \frac{\sum_{i=1}^{N} x(i) - \frac{\sum x}{N}}{N}} \quad \textbf{(24)}$$

Where $\boldsymbol{x}$ denotes the value of features extracted by the LSTM and GRU operations. A set of $\boldsymbol{NS}$ such solutions is generated, and the fitness threshold is calculated based on this generation process, as shown in equation (25).

$$\boldsymbol{fth = \sum_{i=1}^{Ns} f(i) * \frac{Lr}{NS}} \quad \textbf{(25)}$$

Where $\boldsymbol{Lr}$ represents the learning rate for GA operations.
Once all solutions are generated, those with $\boldsymbol{f > fth}$ are 'crossover' to the next iteration, while others are mutated in the current iteration through a regeneration process. These operations are repeated for $\boldsymbol{NI}$ iterations, and solution configurations are modified for each iteration set.
After completing all iterations, the selected features are passed to a Binary Cascaded Neural Network (BCNN), which extracts 1D Convolutional Features using equation (12). These feature sets are then converted into age classes using equation (26).

$$\boldsymbol{Age = SoftMax\left(\sum_{i=1}^{Nf} f(i) * w(i) + b(i)\right)} \quad \textbf{(26)}$$

The process iterates for an augmented set of binary classes, each containing a common age class. For example, if the dataset comprises N age classes, then $\boldsymbol{N - 1}$ such BCNNs are employed for classification. Convergence of the process is achieved through equation (27).

$$\boldsymbol{Age(Final, Component) = Common\ Age, if\ converge\ else, \vee_{i=1}^{Na} Age(i)} \quad \textbf{(27)}$$

Where $\boldsymbol{Na}$ represents the total number of age class groups, $\boldsymbol{Age(i)}$ denotes the age level for the $\boldsymbol{ith}$ set, and $\vee$ Age signifies the age class. This procedure is reiterated for Face, Nose, Mouth, and both Eye components. The ultimate age of the user is assessed using equation (28).

$$\boldsymbol{Age(Final) = \frac{1}{NC}\sum_{i=1}^{NC} Age(Final, i) * Acc(i)} \quad \textbf{(28)}$$

Where $\boldsymbol{Acc(i)}$ signifies the accuracy of age estimation for the $\boldsymbol{ith}$ face component, and $\boldsymbol{NC}$ represents the total number of face components utilized for analysis. Through this process, the model can discern age levels for various users. The efficacy of this model is evaluated across different datasets and samples and is compared with existing models in the subsequent section of this document.

6.0 RESULT ANALYSIS & BENCHMARKING

The presence of publicly available databases plays a crucial role in the advancement of analysis, allowing researchers to engage in rapid analysis exercises and providing a basis for relative estimation. Particularly in cases where database construction is time-consuming, readily available public datasets offer a significant advantage for experimental analysis. Over the past few decades, numerous databases of facial images have been compiled and disseminated to address challenges in face recognition and age estimation.
In our experiments, we utilized the FG-NET (Face and Gesture Recognition Network) Aging Database [51], which

was developed under the European Union's 5th Framework Program, Information Society Technologies. The FG-NET-AD comprises 1002 images of 82 individuals spanning from newborns to 69-year-old individuals, with a higher concentration of ages between zero and 40 years. The images in FG-NET-AD exhibit significant variability in resolution, quality, illumination, viewpoint, and expression, with various images featuring spectacles, facial hair, and hats. Each image in the dataset is annotated with 68 landmark points located at key positions, along with details regarding age, gender, expression, pose, image quality, and presence of facial accessories (e.g., mustaches, beards, hats, or spectacles). This database is commonly used for age estimation and is publicly accessible for experimental purposes.

The performance of the model was evaluated based on Accuracy ***(A),*** Precision (P), Recall (R), Delay (D) and Mean Absolute Error ***(MAE)*** for age estimation, calculated using Eqs. (29), (30), (31), (32) and (33) respectively.

$$\boldsymbol{A = \frac{1}{NTS}\sum_{i=1}^{NTS}\frac{tp(i)+tn(i)}{tp(i)+tn(i)+fp(i)+fn(i)}} \quad \textbf{(29)}$$

$$\boldsymbol{P = \frac{1}{NTS}\sum_{i=1}^{NTS}\frac{tp(i)}{tp(i)+fp(i)}} \quad \textbf{(30)}$$

$$\boldsymbol{R = \frac{1}{NTS}\sum_{i=1}^{NTS}\frac{tp(i)}{tp(i)+tn(i)+fp(i)+fn(i)}} \quad \textbf{(31)}$$

$$\boldsymbol{D = \frac{1}{NTS}\sum_{i=1}^{NTS} ts(complete,i) - ts(start,i)} \quad \textbf{(32)}$$

$$\boldsymbol{MAE = \frac{1}{NTS}\sum_{i=1}^{NTS}\frac{|A(Est)-A(Actual)|}{A(Actual)}} \quad \textbf{(33)}$$

Where ***tp, fp, tn,*** and ***fn*** represent the true and false classification rates, ***ts*** denotes the timestamps for the completion and start of the classification process, while ***A(Est)*** and ***A(Actual)*** indicate the estimated and actual age of the person, evaluated for ***NTS*** test samples.

To evaluate the model's performance, it was tested on the FG-NET open database, which comprises over 1002 images. Among these, 600 were used for training, while 402 images each were reserved for testing and validation.

Based on this strategy, the classification accuracy was compared with DCDL [3] and EC-GAN [9], for different Number of Epochs. The comparative graph for the accuracy is shown in Figure 5.

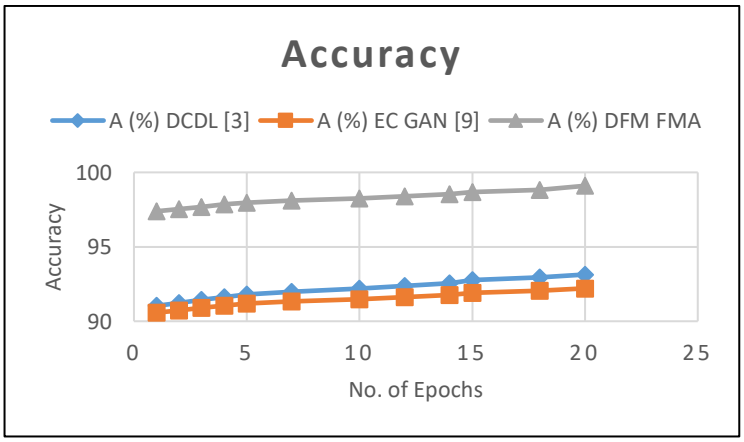

Fig 5: Accuracy of age estimation for different models

Similarly, the parameters like Precision, Recall and Delay were compared with DCDL [3] and EC-GAN [9], for different Number of Epochs. The comparative graph for the Precision, Recall and Delay are shown in Figure 6, 7 and 8 respectively.

International Journal of Multiphysics
Volume 18, No. 3, 2024
ISSN: 1750-9548

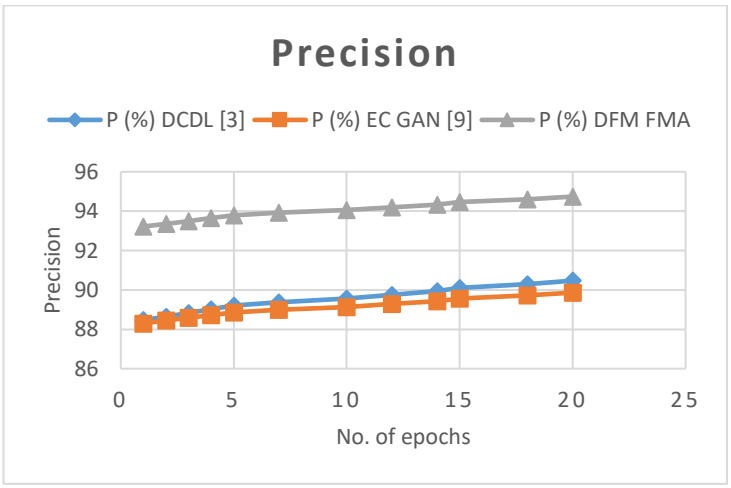

Fig 6: Precision of age estimation for different models

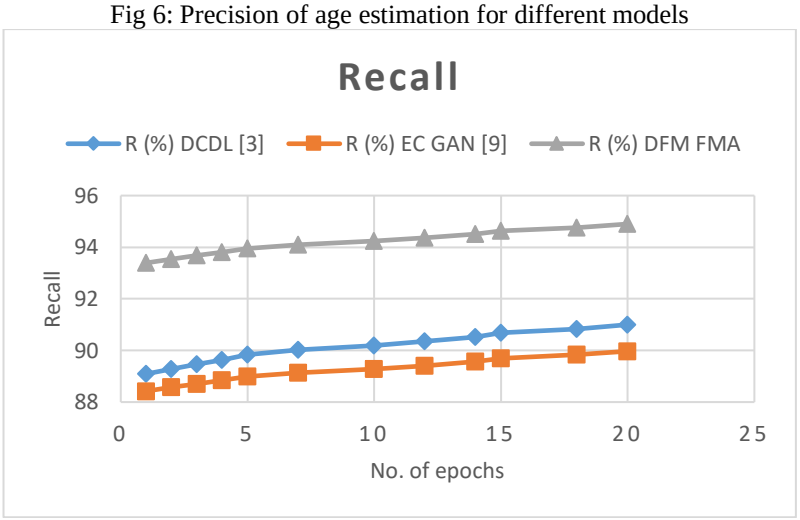

Fig 7: Recall of age estimation for different models

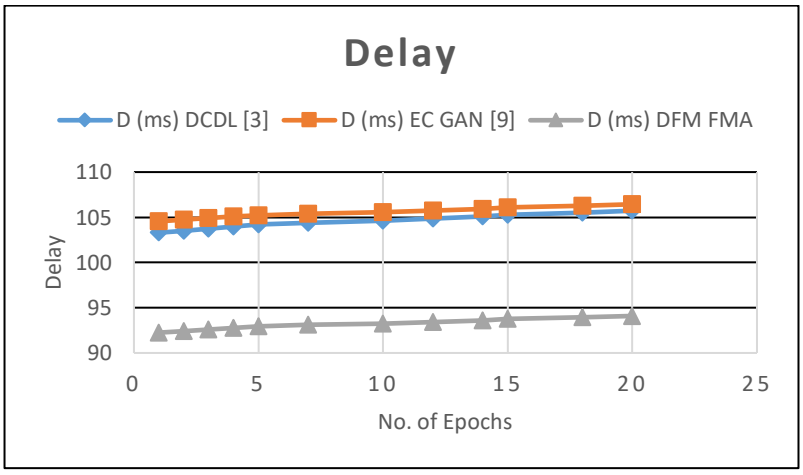

Fig 8: Delay of age estimation for different models

The model's accuracy was evaluated using the augmented FGNET dataset and a sample, resulting in an impressive accuracy of 99.5% and a Mean Absolute Error (MAE) of 1.26 across all age groups.
The achieved MAE for classification is presented in Tab. 1.

Table 1 Mean Absolute Error (MAE) obtained for age estimation using various methods on the FGNET database.

Method	**MAE**
DCLC [3]	2.93
ALSTM Network [11]	2.39
DOEL RESNET [14]	3.44
Mobile Net [14]	2.40
SADAL	3.67
VDAL [16]	2.98
MSFCL	2.82
MSFCL-KL [18]	2.81
DLDLF	3.71
DRF [20]	3.47
RRDCNN [23]	3.05
Gaussian Process [27]	4.41
ADPF [32]	2.86
Light weight CNN [33]	3.05
Deep Learning [38]	3.75
γ- Wasserstein [39]	3.41
LRTI [44]	2.51
MA-SFV2 [46]	3.81
CBIF [49]	3.38
Proposed Method	**1.26**

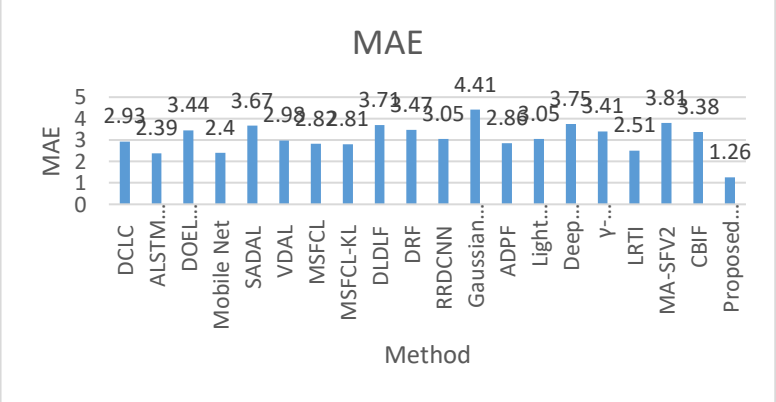

Fig 9: MAE of age estimation for different models

Based on this assessment, it is evident that the proposed model significantly decreases the Mean Absolute Error (MAE) of age estimation compared to existing methods, showing a reduction of 43% compared to DCLC [10], 52.71% compared to LSTM [11], and 44.68% compared to MSFCL [18], among others. This improvement enhances its suitability for various applications requiring high accuracy. The model achieves this improvement through the integration of LSTM and GRU with GA for the selection of highly diverse feature sets. Additionally, the inclusion of PSO augmentation layer and Binary Cascaded CNNs contributes to minimizing MAE even across multiple age categories.

Table 2 The accuracy of age estimation across various methods was evaluated using the FGNET database.

Method	**Accuracy %**
Mobile Net [15]	73.7
MSFCL [18]	64.7
MSFCL-KL [18]	65.3
Proposed Method	**99.5**

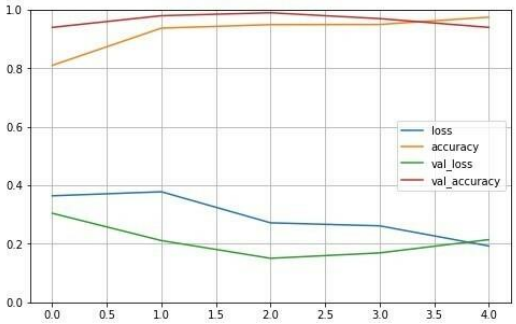

Fig. 5: Accuracy of age estimation for proposed model

Using this approach, the classification accuracy was compared with MobileNet [15], MSFCL [18], and MSFCL-KL [18], as shown in Table 2. Upon evaluation and examination of Figure 5, it becomes evident that the proposed model can enhance the accuracy of age estimation significantly. Specifically, it demonstrates an improvement of 28.5% compared to MobileNet [15], 34.8% compared to MSFCL [18], and 34.2% compared to MSFCL-KL [18]. This enhancement renders the model beneficial for a broad range of high-performance applications. The augmentation of image sets with Binary Cascaded CNNs contributes to the refinement of classification accuracy, particularly across multiple age classes. These optimizations position the proposed model to enhance age estimation performance effectively, particularly in real-time scenarios.

7.0 CONCLUSION

In conclusion, the proposed model integrates multimodal feature representation, multicomponent augmented analysis, and binary cascaded CNNs to enhance the accuracy, precision, recall, speed, and consistency of age estimation. Evaluation results demonstrate that the proposed model surpasses current state-of-the-art techniques in terms of age estimation accuracy, offering potential benefits across various high-performance, high-speed, and high-consistency applications. Looking ahead, future research could explore further enhancements and applications of this model in diverse real-world scenarios.

The proposed model achieves these outcomes through the integration of several key enhancements. Utilizing PSO and GA for identifying consistent features and augmentation operations contributes to enhancing classification precision across multiple age classes. Employing LSTM and GRU for feature recognition, coupled with GA for selecting highly variant feature sets, further enhances classification recall. Additionally, the adoption of a binary cascaded CNN architecture enhances classification accuracy and speed. These enhancements are then combined with a PSO augmentation layer, which aids in reducing MAE in classification across multiple age groups.

In summary, the proposed deep learning model for age estimation represents a significant advancement in computer vision. The optimizations introduced can be applied to enhance the performance of other deep learning models for age estimation or similar computer vision tasks. This research stands as an asset for researchers and practitioners in the fields of computer vision and deep learning models.

REFERENCES

[1] Rahman, S.A., Adjeroh, D.A., “Centroid of age neighborhoods: A new approach to estimate biological age”, *IEEE Journal of Biomedical and Health Informatics* 24(4), 2019, 1226–1234. https://doi.org/10.1109/JBHI.2019.2930938

[2] Liu, X., Li, S., Kan, M., Zhang, J., Wu, S., Liu, W., Han, H., Shan, S., Chen, X., “Agenet: Deeply learned regressor and classifier for robust apparent age estimation”, *In: Proceedings of the IEEE International Conference on Computer Vision Workshops*, 16–24, 2015. https://doi.org/doi:www.cvfoundation.org/openaccess/content_iccv_2015_workshops/w11/html/Liu_AgeNet_Deeply_Learned_ICCV_2015_paper.html.

[3] Sun, H., Pan, H., Han, H., Shan, S., (2021). Deep conditional distribution learning for age estimation. *IEEE Transactions on Information Forensics and Security 16*, 4679–4690. https://doi.org/10.1109/TIFS.2021.3114066.

[4] Vila-Blanco, N., Carreira, M.J., Varas-Quintana, P., Balsa-Castro, C., Tomas, I., (2020). Deep neural networks for chronological age estimation from opg images. *IEEE Transactions on Medical Imaging 39(7)*, 2374–2384. https://doi.org/10.1109/TMI.2020.2968765.

[5] Ashiqur Rahman, S., Giacobbi, P., Pyles, L., Mullett, C., Doretto, G., Adjeroh, D.A., (2021). Deep learning for biological age estimation. *Briefings in Bioinformatics 22(2)*, 1767–1781. https://doi.org/doi:academic.oup.com/bib/article/22/2/1767/5828124#260010214.
[6] Duan, M., Ouyang, A., Tan, G., Tian, Q., (2020). Age estimation using aging/rejuvenation features with device-edge synergy. *IEEE Transactions on Circuits and Systems for Video Technology 31(2)*, 608–620. https://doi.org/10.1109/TCSVT.2020.2981117.
[7] Wu, T., Zhao, Y., Liu, L., Li, H., Xu, W., Chen, C., (2018). A novel hierarchical regression approach for human facial age estimation based on deep forest. *In: 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, pp.1–6. https://doi.org/10.1109/ICNSC.2018.8361338.
[8] Zeng, X., Huang, J., Ding, C., (2020). Soft-ranking label encoding for robust facial age estimation. *IEEE Access 8*, 134209–134218. https://doi.org/10.1109/ACCESS.2020.3010815.
[9] Kim, Y.H., Nam, S.H., Park, K.R., (2020). Enhanced cycle generative adversarial network for generating face images of untrained races and ages for age estimation. *IEEE Access 9*, 6087–6112. https://doi.org/10.1109/ACCESS.2020.3048369.
[10] Akbari, A., Awais, M., Feng, Z.-H., Farooq, A., Kittler, J., (2020). Distribution cognizant loss for cross-database facial age estimation with sensitivity analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence 44(4)*, 1869–1887. https://doi.org/10.1109/tpami.2020.3029486.
[11] Zhang, K., Liu, N., Yuan, X., Guo, X., Gao, C., Zhao, Z., Ma, Z., (2019). Fine-grained age estimation in the wild with attention lstm networks. *IEEE Transactions on Circuits and Systems for Video Technology 30(9)*, 3140–3152. https://doi.org/10.1109/TCSVT.2019.2936410.
[12] Li, W., Lu, J., Wuerkaixi, A., Feng, J., Zhou, J., (2022). Metaage: Meta-learning personalized age estimators. *IEEE Transactions on Image Processing 31*, 4761–4775. https://doi.org/10.1109/TIP.2022.3188061.
[13] Hu, C., Gao, J., Chen, J., Jiang, D., Shu, Y., (2020). Fine-grained age estimation with multi-attention network. *IEEE Access 8*, 196013–196023. https://doi.org/10.1109/ACCESS.2021.3076326.
[14] Xie, J. C., Pun, C. M., (2020). Deep and ordinal ensemble learning for human age estimation from facial images. *IEEE Transactions on Information Forensics and Security 15*, 2361–2374. https://doi.org/10.1109/TIFS.2020.2965298.
[15] Dozdor, Z., Hrka´c, T., Brki´c, K., Kalafati´c, Z., (2023). Facial age estimation models for embedded systems: A comparative study. *IEEE Access 11*, 14282–14292. https://doi.org/10.1109/ACCESS.2023.3244059.
[16] Liu, H., Sun, P., Zhang, J., Wu, S., Yu, Z., Sun, X., (2020). Similarity-aware and variational deep adversarial learning for robust facial age estimation. *IEEE Transactions on Multimedia 22(7)*, 1808–1822. https://doi.org/10.1109/TMM.2020.2969793.
[17] Bao, Z., Tan, Z., Wan, J., Ma, X., Guo, G., Lei, Z., (2022). Divergence-driven consistency training for semi-supervised facial age estimation. *IEEE Transactions on Information Forensics and Security 18*, 221–232. https://doi.org/10.1109/TIFS.2022.3218431.
[18] Xia, M., Zhang, X., Weng, L., Xu, Y., et al., (2020). Multi-stage feature constraints learning for age estimation. *IEEE Transactions on Information Forensics and Security 15*, 2417–2428. https://doi.org/10.1109/TIFS.2020.2969552.
[19] Liu, N., Zhang, F., Duan, F., (2020). Facial age estimation using a multi-task network combining classification and regression. *IEEE Access 8*, 92441–92451. https://doi.org/10.1109/ACCESS.2020.2994322.
[20] Shen, W., Guo, Y., Wang, Y., Zhao, K., Wang, B., Yuille, A. (2019). Deep differentiable random forests for age estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 43(2)*, 404–419. https://doi.org/10.1109/TPAMI.2019.2937294.
[21] Yang, T.Y., Huang, Y. H., Lin, Y. Y., Hsiu, P. C., Chuang, Y. Y., (2018). Ssr-net: A compact soft stagewise regression network for age estimation. *In: International Joint Conference on Artificial Intelligence (IJCAI-18), vol. 5*, p. 7 https://doi.org/doi:10.24963/ijcai.2018/150.
[22] Nam, S.H., Kim, Y.H., Truong, N.Q., Choi, J., Park, K.R., (2020). Age estimation by super-resolution reconstruction based on adversarial networks. *IEEE Access 8*, 17103–17120. https://doi.org/10.1109/ACCESS.2020.2967800.
[23] Dornaika, F., Bekhouche, S.E., Arganda-Carreras, I., (2020). Robust regression with deep CNN's for facial age estimation: An empirical study. *Expert Systems with Applications 141*, 112942. https://doi.org/doi:www.sciencedirect.com/science/article/abs/pii/S0957417419306608.
[24] Xu, C., Sakata, A., Makihara, Y., Takemura, N., Muramatsu, D., Yagi, Y., Lu, J., (2021). Uncertainty-aware gait-based age estimation and its applications. *IEEE Transactions on Biometrics, Behavior, and Identity Science 3(4)*, 479–494. https://doi.org/10.1109/TBIOM.2021.3080300.
[25] Zhang, B., Bao, Y., (2022). Cross-dataset learning for age estimation. *IEEE Access 10*, 24048–24055. https://doi.org/doi:10.1109/ACCESS.2022.3154403.
[26] Cao, W., Mirjalili, V., Raschka, S., (2020). Rank consistent ordinal regression for neural networks with application to age estimation. *Pattern Recognition Letters 140*, 325–331. https://doi.org/doi:10.1016/j.patrec.2020.11.008.

[27] Sawant, M.M., Bhurchandi, K., (2019). Hierarchical facial age estimation using gaussian process regression. *IEEE Access 7*, 9142–9152. https://doi.org/10.1109/ACCESS.2018.2889873.
[28] Benkaddour, M.K., (2021). Cnn based features extraction for age estimation and gender classification. *Informatica 45(5)*. https://doi.org/doi:10.31449/inf.v45i5.3262.
[29] Yang, H., Huang, D., Wang, Y., Jain, A.K., (2019). Learning continuous face age progression: A pyramid of gans. *IEEE Transactions on Pattern Analysis and Machine Intelligence 43(2)*, 499–515. https://doi.org/10.1109/TPAMI.2019.2930985.
[30] Liu, K. H., Liu, H. H., Chan, P.K., Liu, T. J., Pei, S. C., (2018). Age estimation via fusion of depthwise separable convolutional neural networks. *In: 2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1–8. https://doi.org/10.1109/WIFS.2018.8630776.
[31] Pei, W., Dibeklioˇglu, H., Baltruˇsaitis, T., Tax, D.M., (2019). Attended end-to-end architecture for age estimation from facial expression videos. *IEEE Transactions on Image Processing 29*, 1972–1984. https://doi.org/10.1109/TIP.2019.2948288.
[32] Wang, H., Sanchez, V., Li, C. T., (2022). Improving face-based age estimation with attention-based dynamic patch fusion. *IEEE Transactions on Image Processing 31*, 1084–1096. https://doi.org/10.1109/TIP.2021.3139226.
[33] Agbo-Ajala, O., Viriri, S., (2020). A lightweight convolutional neural network for real and apparent age estimation in unconstrained face images. *IEEE Access 8*, 162800–162808. https://doi.org/10.1109/ACCESS.2020.3022039.
[34] Greco, A., Saggese, A., Vento, M., Vigilante, V., (2021). Effective training of convolutional neural networks for age estimation based on knowledge distillation. *Neural Computing and Applications*, 1–16. https://doi.org/doi:10.1007/s00521-021-05981-0.
[35] Gao, B. B., (2023). Jointly learning distribution and expectation in a unified framework for facial age and attractiveness estimation. *Neural Computing and Applications 35(21)*, 15583–15599. https://doi.org/doi:10.1007/s00521-023-08563-4.
[36] Li, D., Ma, X., Ren, Y., Teng, S.-W., (2020). Rectified softmax loss with all-sided cost sensitivity for age estimation. *IEEE Access 8*, 32551–32563. https://doi.org/10.1109/ACCESS.2020.2964281.
[37] Huang, Z., Chen, S., Zhang, J., Shan, H., (2020). Pfa-gan: Progressive face aging with generative adversarial network. *IEEE Transactions on Information Forensics and Security 16*, 2031–2045. https://doi.org/10.1109/TIFS.2020.3047753.
[38] Zaghbani, S., Boujneh, N., Bouhlel, M.S., (2018). Age estimation using deep learning. *Computers & Electrical Engineering 68*, 337–347. https://doi.org/doi:10.1016/j.compeleceng.2018.04.012 (2018).
[39] Akbari, A., Awais, M., Fatemifar, S., Khalid, S.S., Kittler, J., (2021). A novel ground metric for optimal transport-based chronological age estimation. *IEEE Transactions on Cybernetics 52(10)*, 9986–9999. https://doi.org/10.1109/TCYB.2021.3083245.
[40] Wu, Y., Wang, R., Gong, M., Cheng, J., Yu, Z., Tao, D., (2021). Adversarial UV transformation texture estimation for 3d face aging. *IEEE Transactions on Circuits and Systems for Video Technology 32(7)*, 4338–4350. https://doi.org/10.1109/TCSVT.2021.3133313.
[41] Xiao, C., Zhifeng, Z., Jie, C., Qian, Z., (2020). Combined deep learning with directed acyclic graph svm for local adjustment of age estimation. *IEEE Access 9*, 370–379. https://doi.org/10.1109/ACCESS.2020.3046661.
[42] Sheng, M., Ma, Z., Jia, H., Mao, Q., Dong, M., (2020). Face aging with conditional generative adversarial network guided by ranking-cnn. *In: 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 314–319. https://doi.org/10.1109/MIPR49039.2020.00071.
[43] Nam, S.H., Kim, Y.H., Choi, J., Park, C., Park, K.R., (2023). Lca-gan: Low-complexity attention-generative adversarial network for age estimation with mask-occluded facial images. *Mathematics 11(8)*, 1926 https://doi.org/doi:10.3390/math11081926.
[44] Badr, M.M., Elbasiony, R.M., Sarhan, A.M., (2022). Lrti: landmark ratios with task importance toward accurate age estimation using deep neural networks. *Neural Computing and Applications 34(12)*, 9647–9659. https://doi.org/doi:10.1007/s00521-022-06955-6.
[45] Sun, Y., Tang, J., Sun, Z., Tistarelli, M., (2020). Facial age and expression synthesis using ordinal ranking adversarial networks. *IEEE Transactions on Information Forensics and Security 15*, 2960–2972. https://doi.org/10.1109/TIFS.2020.2980792.
[46] Liu, X., Zou, Y., Kuang, H., Ma, X., (2020). Face image age estimation based on data augmentation and lightweight convolutional neural network. *Symmetry 12(1)*, 146 https://doi.org/doi:10.3390/sym12010146.
[47] Hsu, C. Y., Lin, L. E., Lin, C. H., (2021). Age and gender recognition with random occluded data augmentation on facial images. *Multimedia Tools and Applications 80*, 11631–11653. https://doi.org/doi:10.1007/s11042-020-10141-y.
[48] Bafti, S.M., Chatzidimitriadis, S., Sirlantzis, K., (2022). Cross-domain multitask model for head detection and facial attribute estimation. *IEEE Access 10*, 54703–54712. https://doi.org/10.1109/ACCESS.2022.3176621.

[49] Hsu, G. S.J., Cheng, Y. T., Ng, C. C., Yap, M. H., (2017). Component biologically inspired features with moving segmentation for age estimation. *In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 540–547. https://doi.org/10.1109/CVPRW.2017.81.

[50] Rehman, A., Shah, S. A. H., Nizamani, A. U., Ahsan, M., Baig, A. M., & Sadaqat, A. (2024). AI-Driven Predictive Maintenance for Energy Storage Systems: Enhancing Reliability and Lifespan. PowerTech Journal, 48(3). https://doi.org/10.XXXX/powertech.v48.113:contentReference[oaicite:0]{index=0}

[51] 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. https://doi.org/doi:10.1016/j.cviu.2020.102961.

[52] The FG-NET aging database (2012). https://doi.org/http://sting.cycollege.ac.cy/-alanitis/fgnetaging/index.htm.