Development of Facial Age Estimation Using Modified Distance-Based Regressed CNN Model

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

[1] Department of Electronics & Communication, Indian Institute of Information Technology, Nagpur 441108, India
[2] Department of Electronics & Communication, Visvesvaraya National Institute of Technology (V.N.I.T.), Nagpur 440010, India

Corresponding Author Email: dtea20ece001@iiitn.ac.in

ABSTRACT

Facial age estimation determines a person's age from their facial image, crucial for marketing and surveillance applications. Existing models face challenges due to variances in ethnic characteristics and facial expressions, making accurate age estimation difficult. To address these issues, we developed a new model using a modified distance-based regression Convolutional Neural Network (CNN). This innovative approach combines CNN architecture with linear distance-based regression, enhancing accuracy by identifying feature hierarchies and refining age predictions. Additionally, an entropy-based attention mechanism improves accuracy by focusing on the most meaningful parts of the image. This allows the model to effectively weigh different regions during the age estimation process. Our modified distance-RCNN model achieved lower error values for mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and root mean logarithmic squared error (RMSLE). Furthermore, the model showed the lowest error values during k-fold 10. This combination of advanced techniques results in a significant improvement in facial age estimation, addressing the limitations of previous models and providing more precise and reliable age predictions.

1. INTRODUCTION

The process of age estimation involves using different biometric attributes to determine a person's age. The human face is particularly useful in this regard as it provides valuable biological information that enables age estimation through the observation of facial features. Various factors, such as gender, genes, environment, and lifestyle, influence the diversity in facial appearance and the way individuals age across different age groups [1, 2]. Consequently, age estimation of facial images has gained importance in areas like human-computer interaction, identity verification, personalized information services, and more [3]. Different facial features such as wrinkles, freckles, age spots, hair color, face shape, facial hair, and skin texture are used to estimate a person's age [3, 4]. Various methods have been developed to estimate facial age, using image representation techniques like Active Appearance Model (AAM), Active Shape Models (ASM), Aging Pattern Subspace Model (AGES), age manifold-based models, appearance models, and hybrid models, collectively known as Anthropometric models. The extraction of facial features is crucial in age estimation, as it focuses on capturing the essential and significant characteristics [5].

Age estimation is a difficult task due to various factors such as the variation in facial appearance among individuals of the same age and the influence of genes and living conditions on the aging process [6, 7]. Additionally, the slow progression of aging results in subtle differences in facial appearance between adjacent ages, making them hard to perceive. To extract important features in facial image estimation, a range of techniques such as Gabor filters, Linear Discriminant Analysis (LDA) [8], Local Binary Patterns (LBP) [9], Local Directional Patterns (LDP), Local Tertiary Patterns (LTP), Gray-level co-occurrence matrix (GLCM), especially flexible patch (SFP), and Biologically Inspired Features (BIFs) are employed [10, 11]. Age estimation face databases are typically small, due to the challenge of collecting face images with accurate age labels [12-14]. This leads to a common problem of overfitting among existing age estimation algorithms [15].

In recent years, researchers have been working on various methods to enhance the recognition of attributes in facial images. In the past, age estimation and gender classification in face images were achieved by utilizing specifically created feature vectors and statistical or machine learning models. Nevertheless, these tailored features have demonstrated inadequate performance when tested on benchmark datasets containing unrestricted images [16, 17]. Consequently, researchers have shifted their focus towards employing Convolutional Neural Network (CNN) based deep learning architectures for age and gender prediction. These architectures can automatically extract features from the input images [18]. Various methods have been proposed for these tasks. The current research on soft ranking encoder methods has achieved commendable results, but the model is

susceptible to overfitting problems. To address these limitations, CNNs have been developed and have shown promising performance in various aspects of face-related analysis such as face alignment [19], face recognition, face verification, age estimation, and gender classification. Deep learning and CNN have proven to be efficient in predicting the age label based on facial images [20].

The modified distance-based regressed CNN model is a substantial achievement in facial age estimation because of various innovative factors incorporated into it. Unlike the conventional approach, such a method fuses CNN architecture and distance-based regression techniques. This model uses these approaches together to identify and learn to analyze hierarchical information in facial images, so it can generate results better. Further, the model has a preprocessing part before it which includes background subtraction and noise removal, displays better quality of the input data therefore it will likely improve the whole result. In addition, the model incorporates an entropy-based attention mechanism that makes it even more intelligent by allowing it to proactively highlight the significant regions of the face for the process of age estimation to work efficiently.

1.1 Modified distance-based regressed CNN

The Modified distance-based regressed CNN model represents a significant advancement in facial age estimation by combining several innovative elements like hybrid distance-based loss and combined entropy attention-based module. Unlike traditional methods, this model utilizes CNN architecture in tandem with distance-based regression techniques. By integrating these approaches, the model can effectively capture and analyze hierarchical features within facial images, allowing for more accurate age estimation. Furthermore, the inclusion of hybrid distance-based loss will help to control the loss of the happened in the model and the inclusion of an entropy-based attention mechanism adds another layer of sophistication, enabling the model to dynamically prioritize relevant facial regions during age estimation. The major role of the combined entropy-based attention module will help to tune the model for better efficiency.

Section 2 discusses current projects, interpreting their methodologies and the difficulties they encounter. Section 3 discusses the facial age estimation using a Modified distance-based regressed CNN model. Moving onto Section 4, we examine the outcomes of implementing the Modified distance-based regressed CNN model. Section 5 offers a comprehensive examination of the findings of the study, while section 6 explains the concluding summary.

2. MOTIVATION

The research on facial age estimation tackles some challenges such as capturing the diversity of aging patterns in populations, training with data of high quality and diversity; handling shape and illumination variations in the face, modeling in time appropriately, managing privacy and ethical concerns, removing biases, developing models which can be interpretable and dealing with adversarial attacks. However, there is a need to overcome these obstacles to deploy accurate, unfair, and reliable age estimation systems which have applications in biometrics, healthcare as well as personalized

queries.

2.1 Literature review

The following literature review discusses the topic of facial age estimation, highlighting its advantages and limitations.

Dagher and Barbara [1] introduced a novel hierarchical network that utilized transfer learning to achieve accurate age estimation in human faces. Although the network yielded high accuracy, training the model from scratch remained a difficult task. In their study, Liu et al. [2] introduced a direct and effective multi-task learning network that merged classification and regression methods to estimate age. By incorporating age category details into the multi-task learning method, the model's classification accuracy was improved. However, the model's large network architecture increased both its complexity and performance degradation. Zeng et al. [15] introduced a new approach to address the difficulties in facial age estimation by introducing a method called soft-ranking, that encodes two essential characteristics of age prediction from facial features. The incorporation of soft ranking in the encoding block improves the results compared to established methods. However, there is a risk of overfitting with this approach. Wen et al. [21] introduced a distinct approach known as adaptive variance-based distribution learning for age estimation. This technique employs meta-learning to dynamically modify the variation for each image in each iteration, thereby accurately determining the ages through facial images. Nevertheless, the accuracy of the approach's detection capability is affected by the image quality. Bao et al. [22] introduced a new approach to estimating facial age in diverse scenarios by using the Deep Domain-Invariant Learning (DDIL) model. This model effectively addresses domain discrepancies, as demonstrated through thorough experimentation on various age estimation datasets. However, the usefulness of the DDIL method is limited in real-life situations. Liu et al. [23] introduced an effective method for precise facial age estimation in their study. Their approach involved similarity-aware deep adversarial learning and was based on three key components: hard-negative mining, reconstruction loss, and the generation of fake samples using similarity-aware membership. Despite its robust performance, this model still faces challenges regarding scalability. A deep learning model was introduced by Garain et al. [24] to infer age and gender through facial images. This model achieved lower mean absolute error (MAE) and higher accuracy in identification. However, it was noted that a considerable amount of labeled data was needed for this model. Zaghbani et al. [25] introduced a new method for deep learning techniques for the prediction of age estimation. The supervised pre-training method generates better results as compared to the unsupervised method. However, in this model overfitting of data cannot be done is the major challenge. Pathan et al. [26] developed a CNN and LSTM-based CNN model are the deep learning algorithm to predict the human age estimation. The incorporation of time and frequency domain characteristics is a fundamental of this research. Nevertheless, only a calculated set of training data is used in this research and the introduced deep learning model fails to provide improved accuracy performance in predicting. Chandaliya and Nain [27] presented a novel attention with wavelet transformation-based AW-GAN for predicting the aging and d-aging of children and adults. It can detect global and long-range dependencies within internal representations of images with minimal parameters.

However, it suffers from model collapse and less training images. Hangaragi and Singh [28] introduced a new method focused on a deep neural network for the prediction of face recognition. The model detects and recognizes the faces in non-frontal images efficiently. However, it required more data to train the model.

2.2 Challenges

- CR-MT net model requires deeper architecture which increases the complexity of network architecture. Additionally, the model requires substantial training data and computational power [2].
- In the soft ranking method, the model may create misleading results due to the identity overlapping of training and testing datasets [3].
- The quality of the image impacts the efficiency of the model, robust and quality image pre-processing algorithms were needed to overcome these limitations [5].
- The computational complexity of the model is magnified by the need for a substantial amount of data for facial representation and subsequent learning.

2.3 Problem statement

The existing research has used various approaches which consume more memory. Some models have suffered from overfitting issues, the limited amount of data preferred degradation in complexity, lack of concentration, and challenges regarding scalability. In this research, the model utilizes an adaptive attention mechanism that applies weight to different regions of a facial image, depending on their significance to the aging process. This increases the concentration of the model and results in accurate outcomes. So, the current research is based on the Modified distance-based regressed CNN Model developed to overcome the above-mentioned challenges of the existing methods.

3. PROPOSED METHODOLOGY FOR THE FACIAL AGE ESTIMATION USING MODIFIED DISTANCE-BASED REGRESSED CNN MODEL

The major aim of this research is to estimate the facial age by developing a Modified distance-based regressed CNN model. The input images will be obtained from the publicly available UTK face database [29]. Initially, the input images will go through preprocessing which includes background subtraction and noise removal. Face detection using the Viola-Jones algorithm will then be conducted, followed by face modeling using an active shape model commonly used in facial image analysis. The image will be analyzed for wrinkle detection-based statistical features, generating a heat map using a pre-trained deep Alex Net, and extracting hybrid weighted shape-based ResNet-101 features. These extracted features will be inputted into a modified distance-based regressed CNN classifier, which accurately determines the age of the facial image. The model's effectiveness is assessed by utilizing a distinct test face image. Figure 1 depicts the schematic representation of the proposed methodology, which optimizes the performance of the classifier by incorporating a combined entropy attention module [25] and a hybrid distance-based loss.
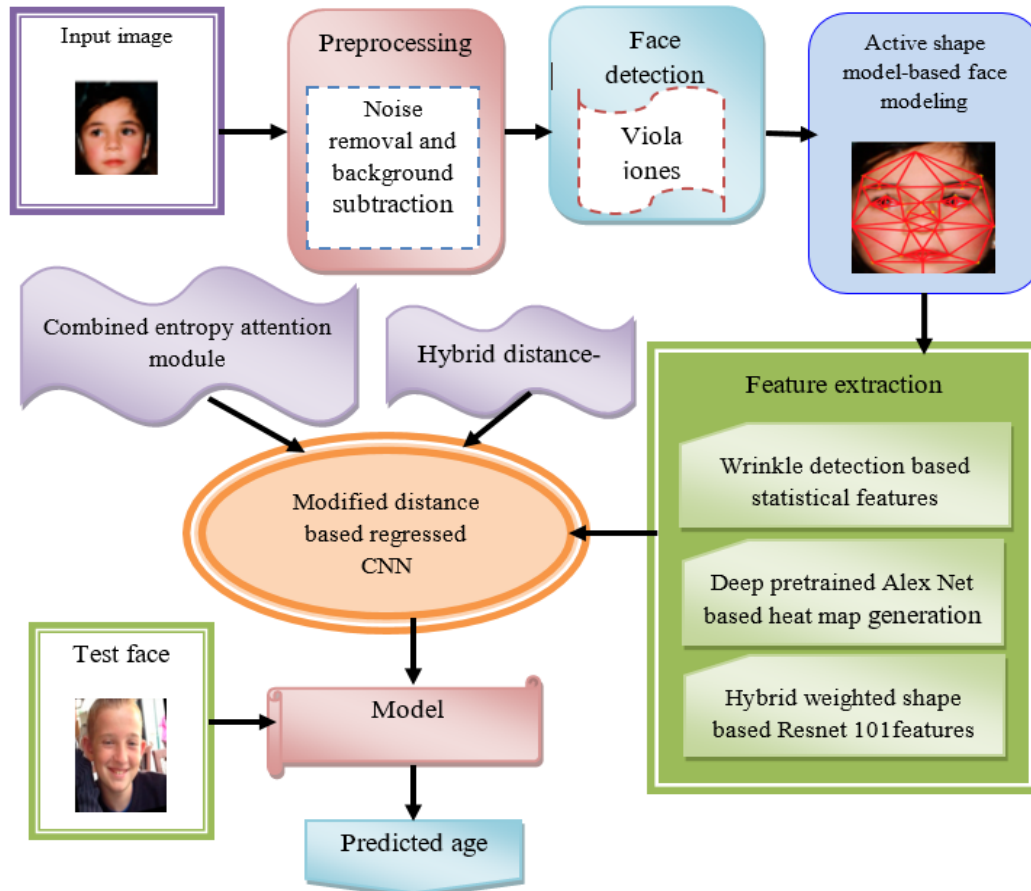

Figure 1. Schematic representation of the proposed method

3.1 Input

The input image is collected from the UTK face database. The mathematical representation of the input utilized for estimating facial age is derived from sources [29].

$$S = \sum_{n=1}^{d} S_q \qquad (1)$$

Here, S_q represents the database.

3.2 Preprocessing

The preprocessing is done in this research to achieve clear data that enables the processing further in a better manner. This specifically focuses on noise removal using a Gaussian filter and background subtraction using the MOG2 (Mixture of Gaussians) algorithm. Gaussian filtering effectively reduces noise and creates smooth images, making it suitable for preprocessing over the other filtering methods. The MOG2 algorithm is used to present multiple color distributions at a pixel. It is most adaptable for crucial backgrounds compared with other background subtraction techniques. The gaussian filter process is given below.

3.2.1 Gaussian filter for noise removal

Typically, an image contains a certain amount of noise, which refers to unwanted variations in brightness or colors. However, in certain cases, adding noise to an image can be beneficial during the other process, as it improves the overall perception of the image, despite degrading the signal-to-noise ratio. The sources of noise in an image can include thermal noise, changes in the scanner, subject placement, and analog-to-digital conversion. These types of noise can be classified as amplifiers or Gaussian noise, salt and pepper noise, shot or Poisson noise, and speckle noise. Gaussian noise is commonly encountered during image acquisition, such as in high temperatures or poor lighting conditions. The Gaussian de-noise algorithm can effectively eliminate Gaussian noise while applying a Gaussian filter can further enhance image smoothening [30]. This filter serves as the initial step in detecting and removing noise, based on the principles of Gaussian distribution. Then, the background subtraction is computed using the MOG2 algorithm. The MOG2 method takes into account each pixel as a mixture of multiple Gaussian distributions, each distribution denoting possible background values for the pixel [31]. Initially, the algorithm models these Gaussian distributions by fitting a histogram from the pixel intensity values in the input image sequence. The model updates the Gaussian models over time in response to changes in the background induced by variations in lighting, scene conditions, or camera movement whenever new frames are processed by the algorithm. After Gaussian distributions have been created, the algorithm computes the differences between the present pixel intensity values and the corresponding Gaussian distributions. Pixels whose values deviate largely from the Gaussian models are identified as foreground, suggesting that such pixels may be of the foreground objects or subjects of interest (such as faces) rather than the background. The approach is known as an iterative method which means that the background models are continuously updated and the comparison between the input image and the background region is repeated. Thus, the MOG2 algorithm separates the foreground which is the face from the background region.

The preprocessed output is denoted as

$$M = \frac{1}{n} \sum_{q=1}^{n} AS_i \qquad (2)$$

3.3 Face detection

Face detection is extensively used in different applications, including security systems and human-computer interfaces. Additionally, it serves as the first stage in the identification of human faces. The Viola-Jones method is highly popular in the field of face detection and is frequently utilized for this purpose [32]. This method is used to construct a face detection system, which is 15 times faster than the previous method and consumes less computation time to generate accurate detection. This method categorizes images using simple feature values instead of directly analyzing pixels. There are numerous reasons for utilizing features rather than pixels. One significant reason is that features enable the encoding of specialized knowledge in domains that are challenging to learn from limited training data. Moreover, when it comes to speed, features derived from the operating system outperform pixel-based systems. The classification of images depends on the values associated with features.

3.4 Active shape model-based face modeling

Active Shape Models (ASMs) and eventually Active Appearance Models (AAMs) are used for face modeling. These models can represent faces correctly, which includes the shape and appearance of facial features, which are very important for tasks like wrinkle detection, feature extraction, and age estimation from facial images. It is beneficial with other models, which consist of shape information and texture information. This modeling is employed because of its accurate feature extraction and robustness. Facial Landmarks Shape Predictor is an algorithm that performs face key points detection and positioning, it identifies and localizes eye corners the tip of the nose, and the mouth corners. Using such an algorithm, the input image is analyzed, and the landmarks are detected and linked to form a curve behind the facial outlines. This geometric shape provides a boundary for the facial area to extract its region from the raw image enclosing the background and other non-facial elements.

3.4.1 Facial landmarks shape predictor

This specific model is designed to predict custom shapes and identify 81 specific facial landmarks in any given image. The training method used is like Dlib's shape predictor for 68 facial landmarks [33]. Apart from the original 68 landmarks, an additional 13 landmarks have been incorporated to encompass the forehead area. This expansion enables accurate detection of the head and facilitates image manipulations that require points on the top of the head. To acquire these extra 13 points, utilized the modified version of Eos by Patrikhuber. Moreover, employed the Surrey Face Model and selected 13 specific points that best suited the intended purpose. After incorporating these adjustments, the model on the comprehensive bug large image database replaces each image's original 68 landmark coordinates with the updated 81 landmark coordinates.

3.4.2 Active Appearance Models

Implemented the Independent Active Appearance Model which consists of 120 annotated images of the frontal face belonging to 12 different subjects (10 images per subject). The images are annotated with 73 points, partitioned the dataset into 110 images for training and 10 images for testing. The Independent Active Appearance Model, as its name implies, separately models shape and texture. AAM model to the target face is a nonlinear optimization task, where the difference in texture between the current model estimate and the target image covered by the model is minimized [34]. This model is commonly applied for face modeling and recognition purposes due to its statistical approach to linearly modeling appearance and shape.

Shape Model. Here utilize annotated point sets as the basis for shaping models. By employing Procrustes Analysis, determine the average shape and covariance matrix for shape matching. The point sets are then transformed into pre-shape space, where we address an alternate optimization problem to calculate the average shape and rotation matrix (to align the shapes in pre-shape space with the average shape). Out of a total of 73×2 points (146 points in total), we select 3 eigenvectors corresponding to the highest 3 Eigenvalues for shape fitting. Thus, the shape model requires calculating 3 scalar values to fit a test image. To accomplish this, use standard linear regression to iteratively fit the 3 coefficients for a test image.

Shape Normalization. The frame of reference is the average shape. The region within the set of points is converted to the average shape mesh using Delaunay Triangulation, Bilinear Interpolation, and Piecewise affine warping for all images. By doing this, obtain a normalized image patch with a shape of 232×240 from the original 800×600 image.

Texture Model. The texture within the mean shape mesh is replicated. The 232×240 image is then reduced in size by a factor of 4 to 53×60. To determine the average texture, solve another optimization problem to obtain global light normalization parameters and the mean texture, which are defined for each image about the mean texture. To create the texture-fitting model, select and keep 60 eigenvectors out of 3480 (corresponding to the top 60 Eigenvalues) from the 58x60 image. Therefore, the texture model requires the computation of 60 scalar values to fit a test image. Once again utilize linear regression (a mathematical technique that solves the Maximum Likelihood estimate using pseudo inverse) for this purpose. As a result, the overall AAM consists of a total of 63 parameters.

3.5 Feature extraction

The first step in the feature extraction stage involves obtaining the necessary information from a facial image, which should accurately represent the image for classification purposes while allowing for a reasonable margin of error. There are three key steps involved in this stage. The input size for feature extraction is (224, 224, 3).

3.5.1 Wrinkle detection based statistical features

During the extraction of wrinkle detection-based statistical features, a group of various approaches are taken that help to get necessary data from facial images. The Canny edge detection algorithm is used to discover edges that match wrinkles that are on the face. Wrinkle edges give information about the spatial distribution and orientation of wrinkles. The initial Gaussian filtering step effectively removes noise from images leading to more accurate edge detection. In this detection the pixels with the maximum gradient magnitude along the edge are retained, producing the thin and well-defined edges. At the same time the principal horizontal wrinkles unit (PHU) method is used to identify the main horizontal wrinkles on our face such as forehead lines and laugh lines. These are the prominent regions of the wrinkles, which are the most particular of the wrinkle patterns that can be isolated. To improve the extracted features further, the bitwise operators are used to combine the result of the canny edge detection and the contour of the pebbles and holes in uncertainty. To illustrate, bitwise AND operations can clip the detected edges under the PHU mask to the contour, resulting in wrinkles being emphasized in the inferred PHU areas. These methods in the right combination offer the opportunity to gather valuable summary statistics to wrinkle patterns that can be involved in the analysis.

Mean. Measurements of the mean pixel intensity with consideration that it can be used to infer the average brightness of facial features would depend on age-related changes in skin texture, tone, or appearance [35].

$$M = \frac{1}{n}\sum_{q=1}^{n} AS_i \tag{3}$$

Therefore, here the overall pixel in the image is denoted as n the intensity value of the ith pixel in the image is signified by the activated shape-based model (AS).

Minimum pixel. The pixel intensity at its lowest is equivalent to the darkest pixel that can be found in the facial image and this darkest pixel is deducted from the region of interest.

$$Min\ pixel = \min(AS) \tag{4}$$

This is the notation that describes the operation that yields the lowest of all pixel value intensities of image *min(AS),* in the active shape model.

Maximum pixel. The brightest pixel corresponds to the pixel value which represents the brightest among those that were defined in a specific region of interest of face image. It marks the highest degree of brightness within the scope.

$$Max\ pixel = \max(AS) \tag{5}$$

This is the notation that describes the operation that yields the lowest of all pixel value intensities of image *max(AS),* in the active shape model.

Skewness. Skewness refers to how the pixel distribution around its mean is asymmetrical. A distribution with zero skewness is symmetrical, with the mean, median, and mode all having equal values. If the skewness is positive, it means the data is positively skewed or skewed to the right. This indicates that the right tail of the distribution is longer, the mean is greater than the median, and the mode is less than the median. Conversely, if the skewness is negative, the data is negatively skewed, meaning the left tail of the distribution is longer, the mean is less than the median, and the mode is greater than the median [35].

$$S = \frac{\left(\frac{1}{n}\sum_{q=1}^{n}(AS-M)^3\right)}{\sigma^3} \tag{6}$$

So, σ means the standard deviation of the pixel intensities in the image.

Entropy. Entropy is a means to assess the level of disorderliness in an image, serving as a gauge of its texture. It offers a numerical representation of the information present in the image. When all the pixels in the image possess identical grayscale values, the entropy is at a minimum. Conversely, if the pixels exhibit a balanced assortment of grayscale values or if the image undergoes histogram equalization, the entropy reaches its maximum level.

$$E = \left(-\sum_{q=1}^{M} p(q) \log_2 (p(q)) \right) \tag{7}$$

In this case, $p(q)$ Denotes the number of available pixel intensity values computed with the active shape model (AS).

Histogram. Histograms can be utilized to analyze the distribution of pixel intensities across facial images, offering insights into the texture, contrast, and overall appearance of the images.

$$H = h(0), h(1), h(2)....., h(M-1) \tag{8}$$

Kurtosis. Kurtosis is a measure that indicates the height of the intensity distribution array. Like skewness, it describes the shape of the probability distribution. There are three interpretations of kurtosis: meso kurtic, leptokurtic, and platy kurtic distributions. A kurtosis value of zero indicates a meso kurtic distribution, which has a normal peak around the mean. A positive kurtosis value represents a leptokurtic distribution, which has a sharper peak around the mean. On the other hand, a negative kurtosis value signifies a platy kurtic distribution, which has a flatter, wider peak around the mean [35].

$$K = \frac{\frac{1}{n} \sum_{q=1}^{M} (AS_i - M)^4}{\sigma^4} \tag{9}$$

The total output size obtained after the wrinkle detection is 1x6.

3.6 Deep pre-trained AlexNet-based heat map generation

High dimensional heat map generation using deep pre-trained AlexNet-based techniques entails the utilization of a pre-trained AlexNet model to extract high-level features from facial images, which are then used to generate heat maps showing regions of utmost importance for age estimation. To build a 1000 class label distribution, the output of the final fully connected layer is passed into the 1000-way softmax function. Eight completely connected nodes make up the network [36]. Following suit, Alex-Net, a state-of-the-art CNN architecture is first taught ImageNet dataset to identify the generic image features. In this context, the pre-trained AlexNet model is explicitly deployed to solve the age estimation applications. Feeding facial images into the network activates layers of increasing depth, which leads to the representation of the complex texture patterns and features of the skin using these activations. The activities generated are later utilized to create heat maps which provide information on the face regions that have a bigger contribution to age prediction. Such heat maps provide opportunities for the

model to detect which facial features, such as wrinkles or contours, are likely to be identified by the model as factors contributing to aging, and they offer valuable insights for the interpretation of the age estimation results and analysis of this data. Deep pre-trained AlexNet-based heat map generation output size: 1x1000. The Deep pre-trained AlexNet-based heat map generation is denoted as D_{AN}.

3.7 Hybrid weighted shape-based ResNet-101 features

Hybrid weighted shape-based ResNet-101 involves using ResNet-101, a very popular architecture because of its end-to-end connections and depth, to extract facial features that are very important for age estimation. By its residual block property, the CNN is distinguished to overcome the gradient that appears due to deep learning [37]. Here a model enfolds the advantage of deep learning with exclusive shape-separated features of face images. ResNet-101 is a model used for face age recognition, which learns hierarchical representations of features from previous datasets to capture dimensions in facial characteristics useful for age recognition. This is done by subsequently utilizing a mechanism that would emphasize the shape-specific features of aging like wrinkles, facial morphology, and especially facial structure; the weights would be combined in a hybrid manner to reinforce the discriminative power of each single attribute. The development of a new hybrid age estimation model which includes shape-based clues and deep learning representations as its components is ultimately aimed at boosting the precision, robustness, and correspondent understanding of the facial aging patterns which are associated with the specific age prediction. The obtained hybrid weighted shape-based Res net 101 features output size is 1×10. These hybrid weighted shape-based ResNet-101 features are denoted as HR_{101}.

The extracted features are concatenated to form an output and the output is fed as an input to the proposed Modified distance-based regressed CNN.

$$FE_{whole} = W_s \parallel D_{AN} \parallel HR_{101} \tag{10}$$

In feature extraction after the concatenation of Wrinkle detection-based statistical features, Deep pre-trained AlexNet-based heat map generation, and Hybrid weighted shape-based ResNet-101 features, the output size is obtained as 1×1018.

4. MODIFIED DISTANCE-BASED REGRESSED CNN

Modified distance-based regressed CNN model is an age estimation method from facial images. The processed output enters the sequence of Convolutional layers within the CNN architecture, thus visual appearance features derived from patterns, textures, and other similar quality visual cues become easier to recognize and handle. The following step combines these features to input a fully connected layer that establishes the relationship between age labels using a modified distance-based regression approach so that it reduces the difference between predictions and actual age labels. Moreover, the inclusion of hybrid distance-based loss will divide the overall age class into independent age labels, and treating each age label independently will help to obtain the correlation between the adjacent ages into consideration and transform the age estimation problem into a simple binary sub-problem. The

inclusion of an entropy-based attention mechanism adds another layer of robustness, the key areas of an image are enhanced, and noised area can be ignored to enable the model to dynamically prioritize relevant facial regions during age estimation. The major role of the combined entropy-based attention module will help to tune the model for better efficiency. The final feature extracted output size forms the input for the facial age estimation mode. Figure 2 depicts the architecture of the developed Modified distance-based regressed CNN model.
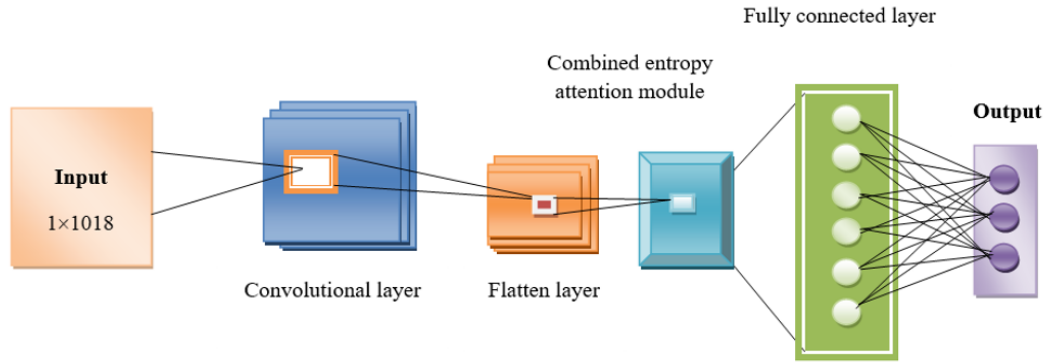

Figure 2. Architecture of the proposed modified distance-based regressed CNN model

4.1 Convolutional layer

The main component of a CNN is the Convolutional layer, which decides the outcome of connected inputs within smaller regions. This is achieved by using a collection of filters (kernels) that can be modified, which are then applied to the input data's width and height to calculate the product between the filter values and the input. Consequently, a two-dimensional activation map is generated for each filter. As a result, CNNs can acquire filters that activate when certain features of a particular spatial position in the input are recognized. Through this investigation after the input layer, 3 Convolutional layers are used consecutively, and these layers are motivated to extract hierarchical features from the input image. A group of filtering processes implemented by the Convolutional layer in each case is applied across an image to pick up special patterns, edges, textures, and various other visual attributes needed for age estimation. The network successively unravels the deeply concealed hiding layers and learns to represent the input image to more abstract levels.

$$D_{con} = Conv(T, L) + b \qquad (11)$$

Here, the input vector is denoted as T, the learnable filters are denoted as L, and the bias term is denoted as b.

4.2 Pooling layer

The CNN's pooling layer performs down sampling on the input's spatial dimension, effectively decreasing the number of weights in the activation. This down-sampling helps improve the invariance of the kernels by reducing the map size of the previous layer. Sub-sampling can be done through two methods: average pooling and maximum pooling. The ReLU activation function sets a threshold at zero, offering advantages over tanh/sigmoid functions. ReLU can be implemented simply by thresholding at zero, while tanh/sigmoid functions involve more costly operations like exponentials. Furthermore, ReLU prevents the loss of gradient error and significantly speeds up the convergence of stochastic gradient descent compared to tanh/sigmoid functions.

$$D_{pool} = Max\,pool(D_{conv}) \qquad (12)$$

The output from the Convolutional layer is denoted as D_{conv}.

4.3 Flatten layer

After Convolutional a set of feature maps will be generated that are multi-dimensional this map will be converted to a dimensional vector by flattening the layer. This is the primary treatment of the obtained features in preparation for the next fully connected layer's input.

$$D_{flatten} = flatten(D_{pool}) \qquad (13)$$

4.4 Combined entropy attention module

The combination of the combined entropy attention module is characterized as a module that has two attention mechanisms and entropy calculations all included. Here's an explanation of how it works based on the provided information:

4.4.1 Soft max input and attention mechanism
The emphasis slot in soft max stands for attention's output, the weights on that region or attribute for the task, in this case, indicate facial age estimation from the images. The attention mechanism only highlights the most relevant and informative image areas representing a face, such as facial contours or lines of expression, by assigning higher weights to these areas.

4.4.2 Entropy calculation
Entropy is a numerical value that shows the uncertainty or randomness of a group of probabilities. In the theoretical frame of attention mechanism, the entropy is computed based on the soft max output which is more like a distribution of consideration across diverse locations in the image. Using the entropy computation of attention weights the model can calculate the amount of randomness or distribution in the way of attention mechanism. The more uniform distribution probability of attention weights on all image pixels represents higher entropy, whereas the more focused or concentrated distribution probability of attention weights on a specific set of image pixels refers to lower entropy.

4.4.3 Multiplication and summation
The multiplication of the softmax inputs with the initial

input of an attention model is performed by a combined entropy attention module. The goal of this process is to emphasize these spots of the image that the attention mechanism highlights as important areas and are deemed relevant by the first input. Therefore, the final attention weights after the multiplication are computed as W. Next, W is executed with the entropy output contributing. The entropy information is therefore added to W. This may lead to a situation where the entropy is considered if the distribution has high uncertainty by influencing the attention mechanism. The main result of this module is $W + entropy$ feedstock, which combines the attention weights with entropy information to sustain the model's concentration on areas of the image while taking the uncertain part of the attention mechanism into account.

$$D_{attention} = soft \max(D_{flatten}) \times F + E \qquad (14)$$

Here, the first input is denoted as F, and entropy is denoted as E.

4.5 Fully connected layer

Fully connected layers resemble the neurons found in typical neural networks, as each neuron in these layers is connected to every neuron in the previous layer. These fully connected layers are considered conventional in deep neural networks and aim to generate predictions based on the activation function, which can be utilized for regression purposes. The model's attention-based output is then passed through one or more dense oblivious layers to yield the output. The networks detect the learned features in different layers and exhibit the classification decision. Concerning the age assessment, these networks of layers can link the extraction of features to the determinations of the age in the input.

$$D_{dense} = RELU(D_{attention}W + b) \qquad (15)$$

Here, weight is denoted as W, and bias is denoted as b.

To optimize the performance of the classifier the combined entropy attention module and a hybrid distance-based loss are integrated. After the combination it is fed into the proposed Modified distance-based regressed CNN model. The inclusion of hybrid distance-based loss will help to control the loss that occurs in the model and the inclusion of an entropy-based attention mechanism adds another layer of robustness, enabling the model to dynamically prioritize relevant facial regions during age estimation. The attention mechanism only highlights the most relevant and informative image areas representing a face, such as facial contours or lines of expression, by assigning higher weights to these areas. The combined entropy-based attention module will help to tune the model for better efficiency. Figure 3 illustrates the layer details for the modified distance-based regressed CNN model.
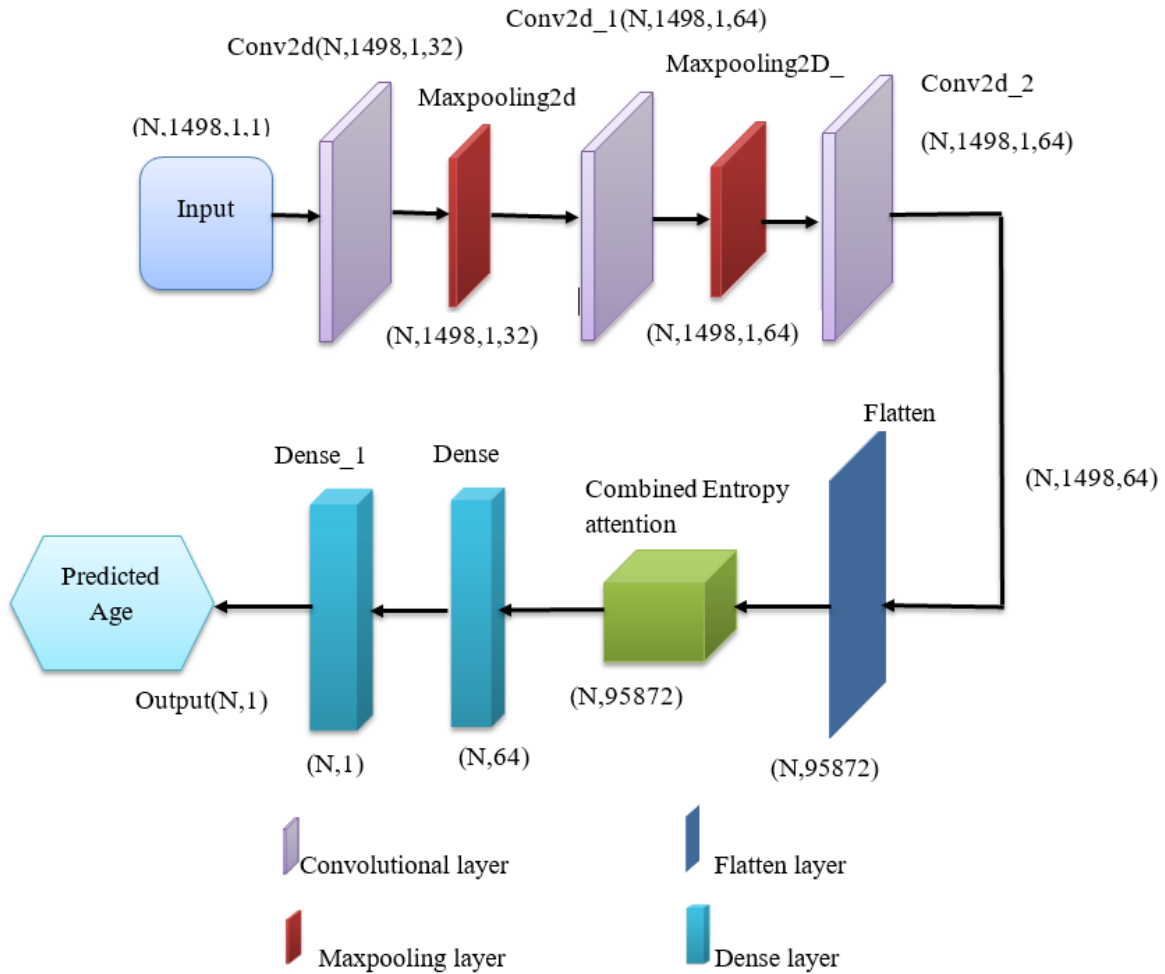

Figure 3. Layer details for the Modified distance-based regressed CNN model

5. COMBINED ENTROPY-BASED ATTENTION MODULE

The combination of the combined entropy attention module is characterized as a module that has two attention mechanisms and entropy calculations. The emphasis slot in soft max stands for attention's output; the weights on that particular region or attribute for the task, in this case, indicate facial age estimation from the images. The attention mechanism only highlights the most relevant and informative image areas representing a face, such as facial contours or lines of expression, by assigning higher weights to these areas. Entropy is a numerical value that shows the uncertainty or randomness of a group of probabilities. In the theoretical frame of attention mechanism, the entropy is computed based on the soft max output which is more like a distribution of consideration across diverse locations in the image. Using the entropy computation of attention weights the model can calculate the amount of randomness or distribution in the way of attention mechanism. The more uniform distribution probability of attention weights on all image pixels represents higher entropy, whereas the more focused or concentrated distribution probability of attention weights on a specific set of image pixels refers to lower entropy.

5.1 Hybrid distance-based loss

The integration of Euclidean distance and the Bhattacharya distance is known as the hybrid distance-based loss. The Euclidean distance is a well-known distance metric in machine learning. The Euclidean distance can be utilized to calculate the distance between any two points in two-dimensional space and also to measure the absolute distance between points in N-dimensional space. For face recognition, smaller values indicate more similar faces. The Bhattacharyya distance is a measure of the similarity between two probability density functions. It is closely related to the Bhattacharyya coefficient, which is a measure of the amount of overlap between two statistical samples or populations. This hybrid distance-based loss is utilized to minimize or control the loss that occurs in the model.

6. RESULT AND DISCUSSION

The purpose of this research is to create a proposed facial age estimation model that is based on the Modified distance-RCNN model. To evaluate its efficacy, the performance of this model is evaluated and compared to other existing models.

6.1 Experimental setup

The facial age estimation task is implemented in Python programming language in the Pycharm platform of the system with 8GB RAM storage and Windows 10 configuration. The activation function is ReLU, the filter size is 32, (1, 1) of kernel size, batch size of 128, Adam is used as the default optimizer.

6.2 Dataset description

6.2.1 UTK face dataset [29]
The UTK Face dataset is a comprehensive collection of face images featuring individuals of different ages, ranging from newborns to 116-year-olds. This dataset includes more than 20,000 images, each with annotations indicating the person's age, gender, and ethnicity. The images exhibit a wide range of variations in terms of pose, facial expression, lighting conditions, occlusions, and image resolution. This dataset can be utilized for various tasks such as face detection, estimating age, predicting age progression or regression, and localizing facial landmarks. Moreover, the dataset offers aligned and cropped versions of the faces, along with corresponding landmarks consisting of 68 points.

6.3 Experimental evaluation

The UTK face dataset images were evaluated using the proposed modified distance-based regression CNN model to generate the estimated facial age. The evaluation of the input image went through the process of preprocessing, face detection, Deep-trained AlexNet-based heatmap generation, hybrid weighted shape-based ResNet-101 features, and the output of estimation age is generated. Figure 4 shows the image results of the evaluation of the modified distance-RCNN model is given below.

6.4 Performance analysis based on TP

The outcomes of the Modified distance-RCNN approach in estimating facial age are displayed in Figure 5. The first graph, Figure 5(a), displays the MAE in training percentage at epoch values of 100, 200, 300, 400, and 500 are 3.33, 2.90, 2.24, 2.13, 1.74 respectively. The results from the Modified distance-RCNN are presented in Figure 5(b), demonstrating that for the same TP and epoch values, the values are 15.68, 12.84, 9.47, 8.59, and 6.04 along with an MSE. Similarly, Figure 5(c) showcases the RMSE values of 3.96, 3.58, 3.08, 2.93, and 2.46 generated by the Modified distance-RCNN using the mentioned training percentage and epoch values, with an RMSE.

6.5 Performance analysis based on k-fold

The outcomes of the Modified distance-RCNN approach in estimating facial age are displayed in Figure 6. The first graph, Figure 6(a), displays the MAE epoch values of 100, 200, 300, 400, and 500 as 3.29, 3.05, 2.74, 2.51 and 1.75 respectively, with a k-fold of 10. The results from the Modified distance-RCNN are presented in Figure 6(b), demonstrating that for the same epoch values, the MSE values are 14.83, 14.80, 13.33, 9.36, and 7.16 along with an MSE k-fold of 10. Similarly, Figure 6(c) showcases the RMSE values of 3.85, 3.85, 3.65, 3.06, and 2.68 generated by the Modified distance-RCNN using the mentioned epoch values, with an RMSE k-fold of 10.

6.6 Comparative methods

To emphasize the achievements of the Modified distance-RCNN, a comparative study is undertaken. This analysis involves employing various techniques including Deep Domain-Invariant Learning (DDIL) [22], Gated Residual Attention Network (GRANet) [24], Deep Learning [25], support vector machine [38], Deep Regression forest [39], Masked contrastive graph representation learning (MCGRL) [40] and random forest [41] is compared by the proposed modified distance-RCNN model for estimating facial age based on Training Percentage and K-fold, which is discussed in the below section.

Input image			
Preprocessed image			
Face detection			
AlexNet image			
ResNet image			
Output	**55**	**28**	**23**

Figure 4. Image result obtained using the Modified distance-RCNN model

(a) MAE

(b) MSE

(c) RMSE
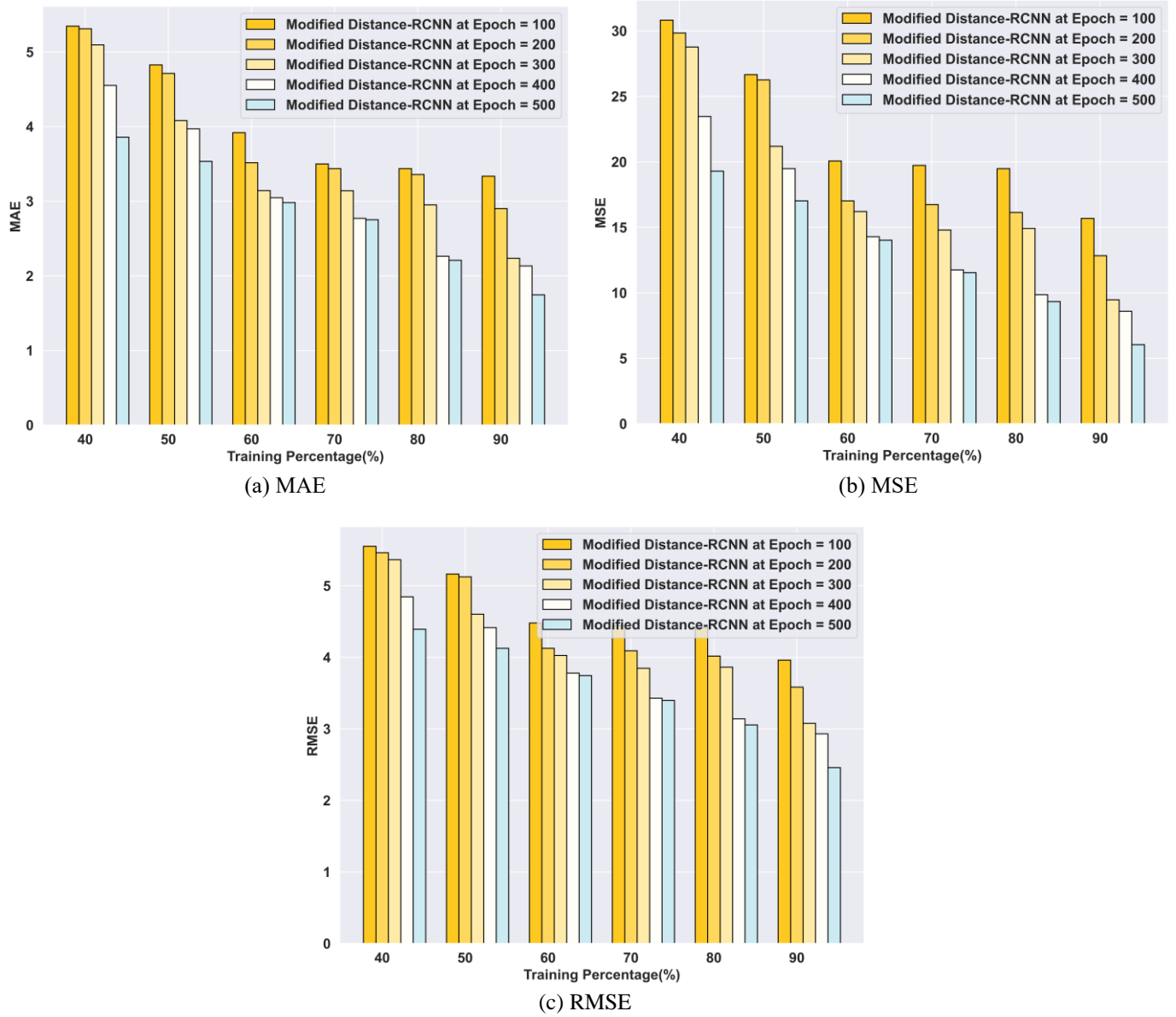
Figure 5. Performance evaluation of the Modified distance-RCNN model with respect to training percentage
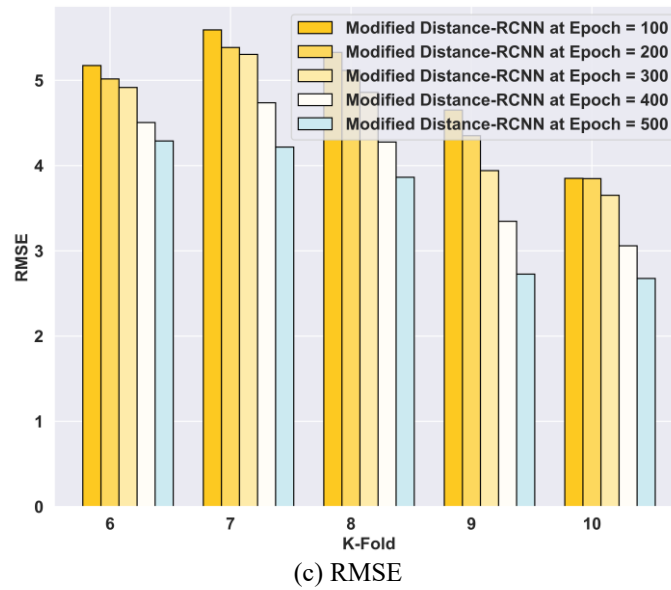

(a) MAE

(b) MSE

(c) RMSE
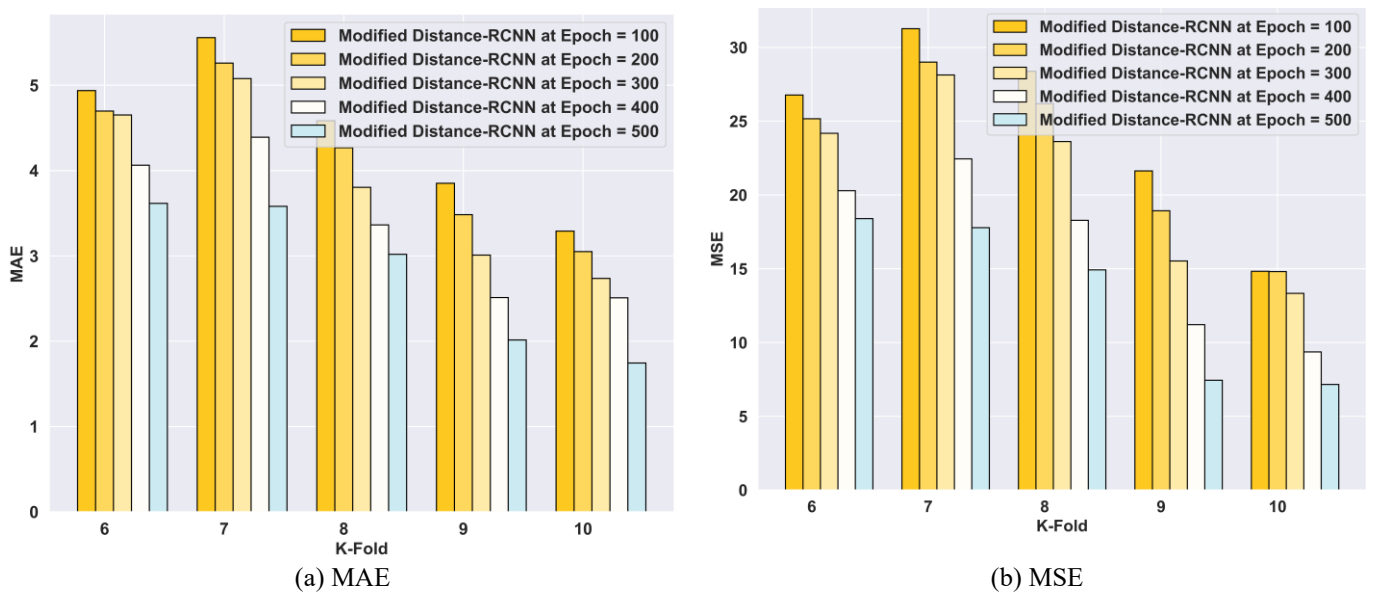
Figure 6. Performance evaluation with respect to k-fold Modified distance-RCNN model

6.6.1 Comparative analysis based on TP

To evaluate the efficacy of the Modified distance-RCNN in determining the age of a face, it is compared to alternative methods by measuring MAE, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) as the benchmark. As shown in Figure 7(a), the error difference of the Modified distance-RCNN model outperforms the other models in MAE of 1.74. The other models gaining the error difference of MAE over DDIL is 0.75, GRANet is 0.38, Deep learning is 0.79, SVM is 0.38, Deep Regression Forest is 0.34, MCGRL is 0.06 and RF is 1.28 respectively.

Furthermore, Figure 7(b) illustrates the superiority of the Modified distance-RCNN in facial age estimation through MSE, the MSE of the proposed model is 6.04. The other models gaining the error difference of MSE over DDIL is 3.26, GRANet is 2.36, Deep learning is 3.34, SVM is 1.90, Deep Regression Forest is 1.59, MCGRL is 1.20 and RF is 7.67 respectively.

Likewise, the Modified distance-RCNN model outperforms other models, as shown in Figure 7(c). The Modified distance-RCNN model achieves RNSE of 2.46 surpassing previous methods. The error difference of an RMSE over other models like DDIL is 0.59, GRANet is 0.44, Deep learning is 0.60, SVM is 0.36, Deep Regression Forest is 0.30, MCGRL is 0.23 and RF is 1.24 respectively.
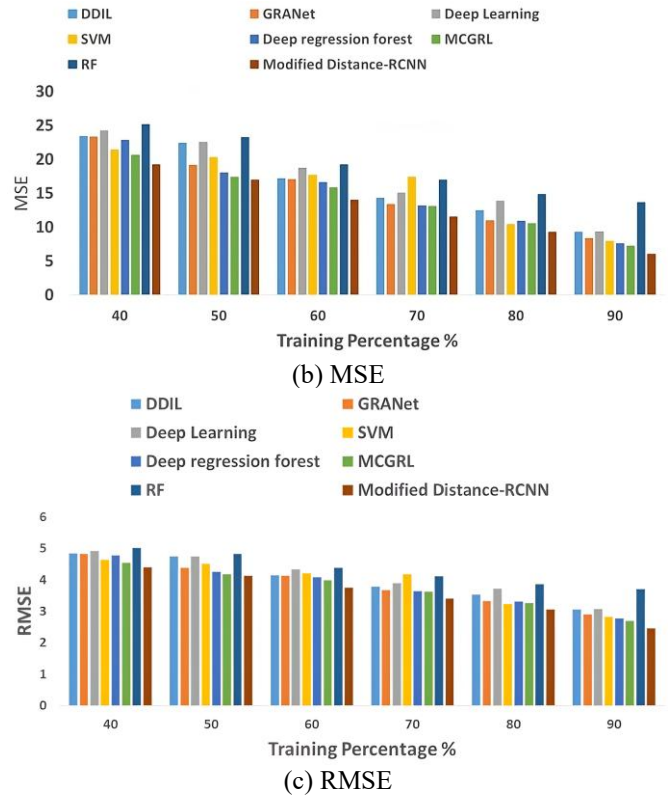

(b) MSE

(c) RMSE

Figure 7. Comparative evaluation of modified distance-RCNN model with respect to training percentage

6.6.2 Comparative evaluation with respect to k-fold

As shown in Figure 8 the comparative analysis of the Modified distance-RCNN model outperforms the other models with k-fold values in estimating facial age. Figure 8(a) shows the MAE value of the proposed model is 1.75 at a k-fold value of 10. The other models gained the error difference like DDIL is 0.63, GRANet is 0.63, Deep learning is 0.84, SVM is 0.46, Deep Regression Forest is 0.46, MCGRL is 0.15 and RF is 0.94 respectively. Furthermore, Figure 8(9b) illustrates the superiority of the Modified distance-RCNN model gained an MSE of 7.16, it is compared with other models like DDIL is 2.43, GRANet is 2.03, Deep learning is

(a) MAE

4.64, SVM is 0.22, Deep Regression Forest is 0.94, MCGRL is 0.22 and RF is 4.87 respectively. Figure 8(c) emphasizes the impressive performance of the Modified distance-RCNN model generating the MMSE value of 2.68, when compared to the other models. It exhibits a significant improvement in error values over DDIL is 0.42, GRANet is 0.36, Deep learning is 0.76, SVM is 0.04, Deep Regression Forest is 0.17, MCGRL is 0.04 and RF is 0.79 respectively.
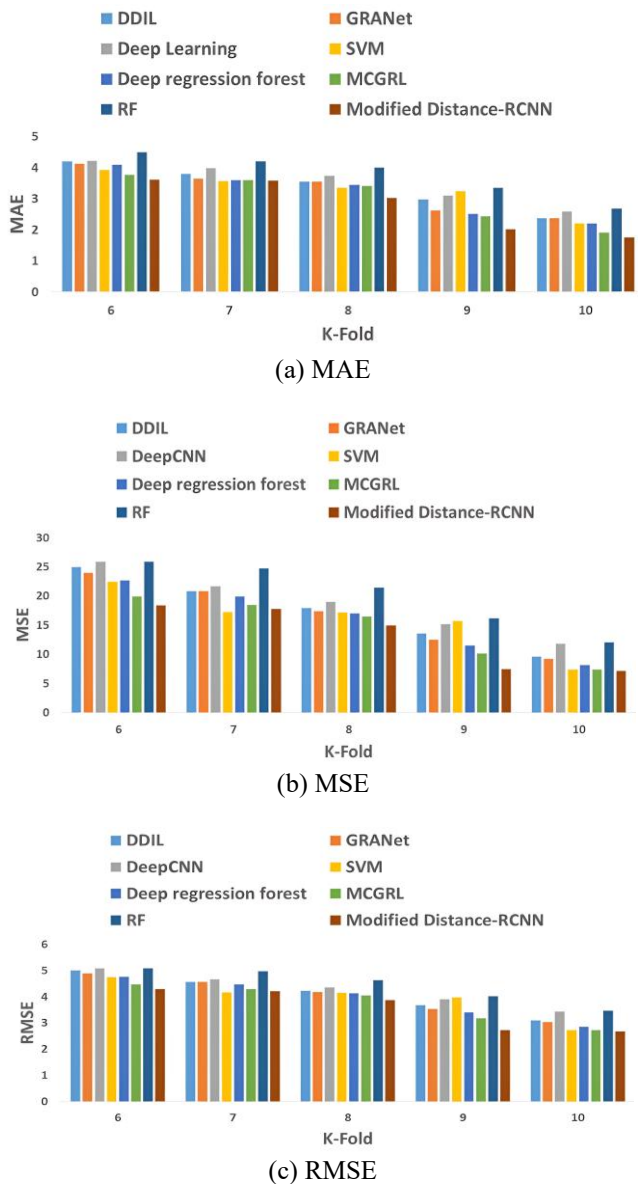

(a) MAE

(b) MSE

(c) RMSE

Figure 8. Comparative evaluation of Modified distance-RCNN model with respect to k-fold

6.7 Comparative discussion

The conventional approaches for the existing method's predictions have some limitations such as the support vector machine-based model being developed and it suffers from overfitting issues. The CNN-based deep learning model is created and it suffers from only a limited amount of data that can be used as training data in this research. A deep CNN classifier is introduced and it suffers from model collapse and only has a smaller number of training images. The previous models have limitations like overfitting issues and varying data loss in diverse backgrounds. A novel approach based on

deep neural networks is developed and it requires more data to train the model. The current study developed a Modified distance-RCNN model that achieves a better result by avoiding the above-mentioned issues. The proposed Modified distance-RCNN model with the convolution neural network architecture for hierarchical feature extraction performs better than the facial age estimation existing models. Such performances make the proposed model acceptable in real-time age estimation systems. Instead of the standard regression approach, the current technique can automatically learn complex relationships between face characteristics and age, thus creating more accurate predictions. Also, the model utilizes an adaptive attention mechanism that applies weight to different regions of a facial image, depending on their significance to the aging process. This increases the concentration of the model on notable aspects. The results demonstrate that the proposed model outperforms other models when it comes to MAE, MSE, and RMSE. Specifically, the Modified distance-RCNN model displayed notably minimum error values during training percentage 90 for MAE, MSE, and RMSE of 1.74, 6.04, and 2.46 respectively. Likewise, the model also achieved the minimum error values during k-fold 10 of 1.75, 7.16, and 2.68 respectively. The comparative discussion of the model results can be found in Table 1.
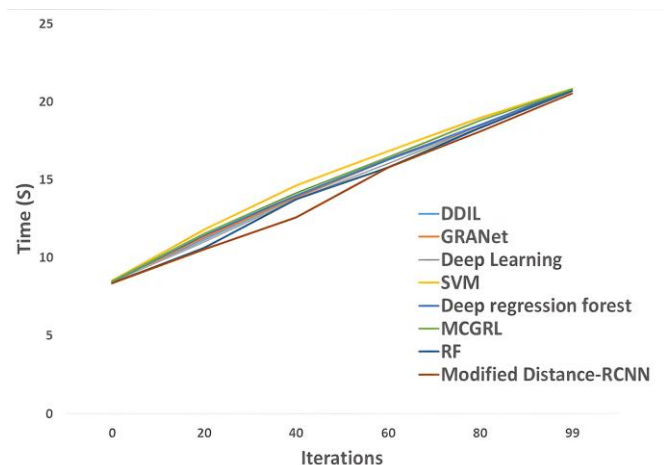

Figure 9. Computational complexity

6.8 Computational complexity

The computational complexity analysis shows the comparison between the modified Distance-based RCNN model with other existing models. Moreover, this model is trained over multiple iterations to evaluate the efficiency of the model. The computational complexity of the modified distance-based RCNN model proposes the analysis between the computational time and iterations of the existing model for estimating facial age. The computational time of the model-like support vector machine is 20.83, random forest is 20.69, deep CNN is 20.76, DDIL is 20.78, GRANet is 20.79, Deep Regression Forest is 20.79, MCGRL is 20.80, and computational time of the proposed model is 20.53 seconds. The Modified Distance-based RCNN model performs in minimum computational time over other models. Figure 9 illustrates the graphical representation of the computational complexity analysis of the modified distance-based RCNN model.

Table 1. Comparative discussion table for TP and k-fold

Models	TP -90			k-fold 10		
	MAE	MSE	RMSE	MAE	MSE	RMSE
DDIL [22]	2.50	9.30	3.05	2.38	9.59	3.10
GRANet [24]	2.12	8.40	2.90	2.38	9.19	3.03
Deep Learning [25]	2.54	9.38	3.06	2.59	11.80	3.44
Support Vector Regression [30]	2.12	7.94	2.82	2.20	7.39	2.72
Deep Regression Forest [39]	2.09	7.63	2.76	2.21	8.10	2.85
MCGRL [40]	1.80	7.24	2.69	1.90	7.38	2.72
Random Forest [41]	3.03	13.71	3.70	2.69	12.03	3.47
Modified distance-based RCNN	**1.74**	**6.04**	**2.46**	**1.75**	**7.16**	**2.68**

7. CONCLUSION

We propose a Modified distance-based regressed CNN model for the facial age estimation that incorporates the CNN architecture, distance-based regression, and entropy-based attention mechanisms. Through preprocessing which contains background subtraction and noise elimination, the model augments input images and hence makes it resilient to differentiate variables. Using CNN layers rich with information, facial images are analyzed hierarchically, thus leading to enhancing precision in the age estimation. Introducing the regression model based on distance allows building the relation between features and age labels helping to achieve even more precise classification. Moreover, the entropy-based attention mechanism behaves under attention distribution, which allows for the importance of different facial regions to be handled correctly. The resulting model is a fully structured system that is devoted to solving the most essential tasks in facial age estimation, thus providing age prediction accurately and reliably. Potential directions for future research should involve more optimization and testing of the model to ensure that it applies to varied contexts. The model should be tested towards different data trends and populations to evaluate its robustness. Specifically, the Modified distance-RCNN model displayed notable minimum error values of MAE is 1.74, MSE of 6.04, and RMSE is 2.46 respectively. Likewise, the model also achieved the minimum error values during k-fold 10 of MAE is 1.75, MSE is 7.16, and RMSE is 2.68 respectively compared with other existing methods. However, the facial age estimation methods performed on the UTK face datasets that the aging process can be impacted by factors such as gender, genes, race, health status, and life circumstances. It affects the generalizability of model results, which means the proposed modified distance-based RCNN model might not perform well in real-world problems like diverse populations, lighting, poses, and facial expressions. Further, incorporating the UTK dataset with another diverse dataset can improve the evaluation of the model performance, which impacts the generalizability of the proposed modified distance-based RCNN model.

The facial age estimation technologies are particularly employed in artificial intelligence and deep learning. Facial features are utilized in various applications like demographics, affective computing, security, and authentication. Facial age estimation is commonly used in human-computer interactions, Forensic investigations, face verification, detecting spoofing, and facial attacks. Among these applications, facial age estimation stands out as a key field due to its broad applications and the challenges that come with it.

REFERENCES

[1] Dagher, I., Barbara, D. (2021). Facial age estimation using pre-trained CNN and transfer learning. Multimedia Tools and Applications, 80: 20369-20380. https://doi.org/10.1007/s11042-021-10739-w

[2] 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

[3] Song, Z., Ni, B., Guo, D., Sim, T., Yan, S. (2011). Learning universal multi-view age estimator using video context. In 2011 International Conference on Computer Vision, Barcelona, Spain, pp. 241-248. https://doi.org/10.1109/ICCV.2011.6126248

[4] Shu, X., Tang, J., Lai, H., Liu, L., Yan, S. (2015). Personalized age progression with aging dictionary. In Proceedings of the IEEE International Conference on Computer Vision, pp. 3970-3978. https://doi.org/10.1109/ICCV.2015.452

[5] Geng, X., Zhou, Z.H., Zhang, Y., Li, G., Dai, H. (2006). Learning from facial aging patterns for automatic age estimation. In Proceedings of the 14th ACM international conference on Multimedia, New York, United States, pp. 307-316. https://doi.org/10.1145/1180639.1180711

[6] Sun, Y., Wang, X., Tang, X. (2013). Deep convolutional network cascade for facial point detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, USA, pp. 3476-3483. https://doi.org/10.1109/CVPR.2013.446

[7] Rizwan, S.A., Jalal, A., Gochoo, M., Kim, K. (2021). Robust active shape model via hierarchical feature extraction with SFS-optimized convolution neural network for invariant human age classification. Electronics, 10(4): 465. https://doi.org/10.3390/electronics10040465

[8] Yang, Z., Ai, H. (2007). Demographic classification with local binary patterns. In Advances in Biometrics: International Conference, ICB 2007, Seoul, Korea, pp. 464-473. https://doi.org/10.1007/978-3-540-74549-5_49

[9] Geng, X., Yin, C., Zhou, Z.H. (2013). Facial age estimation by learning from label distributions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(10): 2401-2412. https://doi.org/10.1109/TPAMI.2013.51

[10] Rahman, M.A., Aonty, S.S., Deb, K., Sarker, I.H. (2023). Attention-based human age estimation from face images to enhance public security. Data, 8(10): 145. https://doi.org/10.3390/data8100145

[11] Punyani, P., Gupta, R., Kumar, A. (2020). Neural

networks for facial age estimation: A survey on recent advances. Artificial Intelligence Review, 53(5): 3299-3347. https://doi.org/10.1007/s10462-019-09765-w

[12] Kwon, Y.H., da Vitoria, L. (1994). Age classification from facial images. 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, pp. 762-767. https://doi.org/10.1109/CVPR.1994.323894

[13] Lanitis, A., Taylor, C.J., Cootes, T.F. (2002). Toward automatic simulation of aging effects on face images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(4): 442-455. https://doi.org/10.1109/34.993553

[14] Geng, X., Yin, C., Zhou, Z.H. (2013). Facial age estimation by learning from label distributions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(10): 2401-2412. https://doi.org/10.1109/TPAMI.2013.51

[15] 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

[16] Bouchrika, I., Harrati, N., Ladjailia, A., Khedairia, S. (2015). Age estimation from facial images based on hierarchical feature selection. In 2015 16th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), Monastir, Tunisia, pp. 393-397. https://doi.org/10.1109/STA.2015.7505156

[17] Chen, S., Zhang, C., Dong, M., Le, J., Rao, M. (2017). Using ranking-CNN for age estimation. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, pp. 742-751. https://doi.org/10.1109/CVPR.2017.86

[18] Drobnyh, K.A., Polovinkin, A.N. (2017). Using supervised deep learning for human age estimation problem. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42: 97-100. https://doi.org/10.5194/isprs-archives-XLII-2-W4-97-2017

[19] Levi, G., Hassner, T. (2015). Age and gender classification using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, USA, pp. 34-42. https://doi.org/10.1109/CVPRW.2015.7301352

[20] 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

[21] Wen, X., Li, B., Guo, H., Liu, Z., Hu, G., Tang, M., Wang, J. (2020). Adaptive variance based label distribution learning for facial age estimation. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, pp. 379-395. https://doi.org/10.1007/978-3-030-58592-1_23

[22] Bao, Z., Luo, Y., Tan, Z., Wan, J., Ma, X., Lei, Z. (2023). Deep domain-invariant learning for facial age estimation. Neurocomputing, 534: 86-93. https://doi.org/10.1016/j.neucom.2023.02.037

[23] 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

[24] Garain, A., Ray, B., Singh, P.K., Ahmadian, A., Senu, N., Sarkar, R. (2021). GRA_Net: A deep learning model for classification of age and gender from facial images. IEEE Access, 9: 85672-85689. https://doi.org/10.1109/ACCESS.2021.3085971

[25] Zaghbani, S., Boujneh, N., Bouhlel, M.S. (2018). Age estimation using deep learning. Computers & Electrical Engineering, 68: 337-347. https://doi.org/10.1016/j.compeleceng.2018.04.012

[26] Pathan, R.K., Uddin, M.A., Nahar, N., Ara, F., Hossain, M.S., Andersson, K. (2021). Human age estimation using deep learning from gait data. In International Conference on Applied Intelligence and Informatics, pp. 281-294. https://doi.org/10.1007/978-3-030-82269-9_22

[27] Chandaliya, P.K., Nain, N. (2023). AW-GAN: Face aging and rejuvenation using attention with wavelet GAN. Neural Computing and Applications, 35(3): 2811-2825. https://doi.org/10.1007/s00521-022-07721-4

[28] Hangaragi, S., Singh, T. (2023). Face detection and recognition using face mesh and deep neural network. Procedia Computer Science, 218: 741-749. https://doi.org/10.1016/j.procs.2023.01.054

[29] UTK-face dataset. https://susanqq.github.io/UTKFace/.

[30] Suthaharan, S. (2016). Support vector machine. In Machine Learning Models and Algorithms for Big Data Classification: Integrated Series in Information Systems, vol. 36, pp. 207-235. https://doi.org/10.1007/978-1-4899-7641-3_9

[31] Shreyamsha Kumar, B.K. (2013). Image denoising based on gaussian/bilateral filter and its method noise thresholding. Signal, Image and Video Processing, 7: 1159-1172. https://doi.org/10.1007/s11760-012-0372-7

[32] Wibowo, H.T., Wibowo, E.P., Harahap, R.K. (2021). Implementation of background subtraction for counting vehicle using mixture of Gaussians with ROI optimization. In 2021 Sixth International Conference on Informatics and Computing (ICIC), Jakarta, Indonesia, pp. 1-6. https://doi.org/10.1109/ICIC54025.2021.9632950

[33] Dabhi, M.K., Pancholi, B.K. (2016). Face detection system based on Viola-Jones algorithm. International Journal of Science and Research (IJSR), 5(4): 62-64.

[34] Palaniappan, M., Sowmia, K.R., Aravindkumar, S. (2022). Real time fatigue detection using shape predictor 68 face landmarks algorithm. In 2022 International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, pp. 1-5. https://doi.org/10.1109/ICITIIT54346.2022.9744142

[35] Sabina, U., Whangbo, T.K. (2021). Edge-based effective active appearance model for real-time wrinkle detection. Skin Research and Technology, 27(3): 444-452. https://doi.org/10.1111/srt.12977

[36] Choi, Y.H., Tak, Y.S., Rho, S., Hwang, E. (2013). Skin feature extraction and processing model for statistical skin age estimation. Multimedia Tools and Applications, 64: 227-247. https://doi.org/10.1007/s11042-011-0987-7

[37] Khan, N., Singh, A., Agrawal, R. (2023). Enhancing feature extraction technique through spatial deep learning model for facial emotion detection. Annals of Emerging Technologies in Computing (AETiC), 7(2): 9-22. https://doi.org/10.33166/aetic.2023.02.002

[38] Khader, A., Alquran, H. (2023). Automated prediction of osteoarthritis level in human osteochondral tissue using histopathological images. Bioengineering, 10(7): 764. https://doi.org/10.3390/bioengineering10070764

[39] Zhang, B., Bao, Y. (2022). Age estimation of faces in videos using head pose estimation and convolutional neural networks. Sensors, 22(11): 4171. https://doi.org/10.3390/s22114171

[40] Shou, Y., Cao, X., Liu, H., Meng, D. (2025). Masked contrastive graph representation learning for age estimation. Pattern Recognition, 158: 110974. https://doi.org/10.1016/j.patcog.2024.110974

[41] Guehairia, O., Dornaika, F., Ouamane, A., Taleb-Ahmed, A. (2022). Facial age estimation using tensor based subspace learning and deep random forests. Information Sciences, 609: 1309-1317. https://doi.org/10.1016/j.ins.2022.07.135