

Article

A Hierarchical Approach toward Prediction of Human Biological Age from Masked Facial Image Leveraging Deep Learning Techniques

Md. Mahbubul Islam  and Joong-Hwan Baek *

Department of Electronics and Information Engineering, Korea Aerospace University, Goyang 10540, Korea; mahbubcse@cu.ac.bd

* Correspondence: jhbaek@kau.ac.kr

Abstract: The lifestyle of humans has changed noticeably since the contagious COVID-19 disease struck globally. People should wear a face mask as a protective measure to curb the spread of the contagious disease. Consequently, real-world applications (i.e., electronic customer relationship management) dealing with human ages extracted from face images must migrate to a robust system proficient to estimate the age of a person wearing a face mask. In this paper, we proposed a hierarchical age estimation model from masked facial images in a group-to-specific manner rather than a single regression model because age progression across different age groups is quite dissimilar. Our intention was to squeeze the feature space among limited age classes so that the model could fairly discern age. We generated a synthetic masked face image dataset over the IMDB-WIKI face image dataset to train and validate our proposed model due to the absence of a benchmark masked face image dataset with real age annotations. We somewhat mitigated the data sparsity problem of the large public IMDB-WIKI dataset using off-the-shelf down-sampling and up-sampling techniques as required. The age estimation task was fully modeled like a deep classification problem, and expected ages were formulated from SoftMax probabilities. We performed a classification task by deploying multiple low-memory and higher-accuracy-based convolutional neural networks (CNNs). Our proposed hierarchical framework demonstrated marginal improvement in terms of mean absolute error (MAE) compared to the one-off model approach for masked face real age estimation. Moreover, this research is perhaps the maiden attempt to estimate the real age of a person from his/her masked face image.

Keywords: deep learning; masked face age estimation; COVID-19; hierarchical approach



Citation: Islam, M.M.; Baek, J.-H. A Hierarchical Approach toward Prediction of Human Biological Age from Masked Facial Image Leveraging Deep Learning Techniques. *Appl. Sci.* **2022**, *12*, 5306. <https://doi.org/10.3390/app12115306>

Academic Editor: Anton Civit

Received: 27 April 2022

Accepted: 22 May 2022

Published: 24 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The respiratory disease COVID-19 is an unprecedented crisis that has led to an enormous number of mortalities and security-related issues. This pandemic has transformed the lives of millions of people around the world. Over the last twenty years, there have seen several outbreaks of different coronavirus diseases across the world, but people around the globe are facing an elusive health crisis due to the emergence of COVID-19 disease, a member of the coronavirus family. It is a well-known fact that respiratory viruses such as COVID-19 spread by person-to-person contact within and outside the community. One of the modes of transmission of COVID-19 is airborne transmission. This transmission occurs as humans breathe in droplets released by an infected person through breathing, speaking, singing, coughing, or sneezing. Hence, public health officials have mandated the use of face masks which can lessen virus transmission by a substantial margin. Physicians have suggested that social distancing and contact-free services in public spaces can be major protective measures against COVID-19 disease. Contact-free operations are highly preferable among the possible measures, particularly in public spaces such as chain stores, stations, etc. The use of face masks in public areas presents an arduous challenge since

face-related applications are typically trained with human faces devoid of masks, but now, due to the onset of the COVID-19 pandemic, they are forced to deal with faces with masks. Moreover, wearing a face mask causes the following problems: (1) community access control and face authentication become very difficult tasks when a large part of one's face is hidden by a mask, and (2) the whole face image might not be taken into consideration for feature description.

Human faces largely provide better traits to recognize a person's demographics as opposed to other common biometric features (e.g., irises and fingerprints), as 6 mA and SNARE proteins play roles in various cellular processes and disease pathogenesis [1,2] in the area of biomedical research. The human face contains features that determine the age of a person and are equally useful in various real-life applications such as security surveillance, human–computer interaction, access control, demographic data extraction, law administration, marketing intelligence, etc. Despite the progress made by the researchers, age predictions from unconstrained masked facial images are yet to meet the demands of business applications. Thus, as an attempt to meet the necessity of an enterprise solution during the pandemic, we proposed an automatic age estimation system from masked faces, which is visualized in Figure 1.

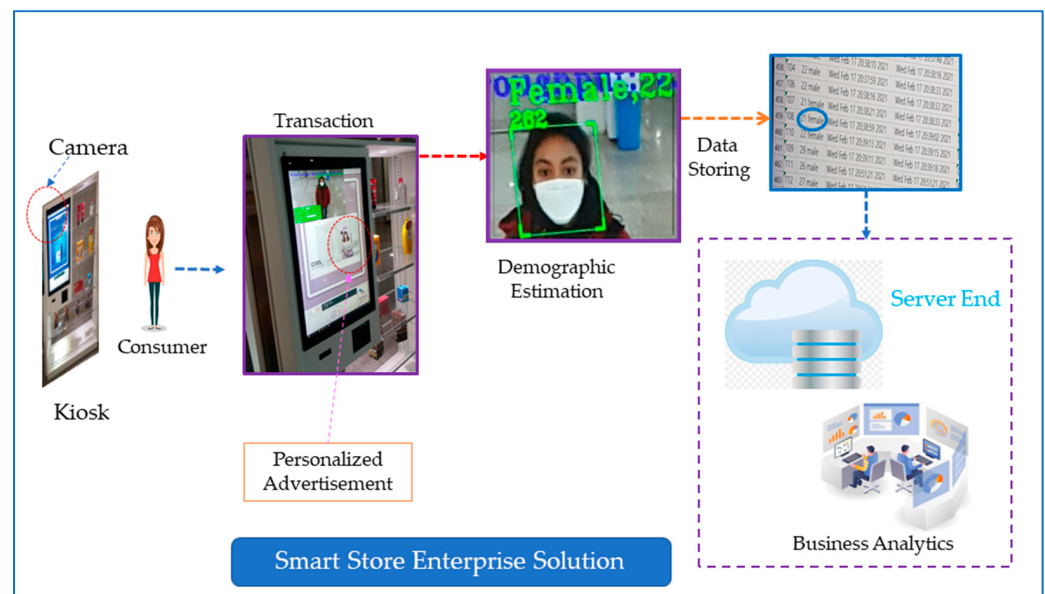


Figure 1. Masked face age estimation system for smart store customer relationship management.

However, face-feature-based systems are not robust enough when used with masked faces, because the distinctive features of a face are largely covered by face masks. As a consequence, conventional face-centered solutions should transform into upgraded systems, as it is essential that they continue operating during the pandemic. There are two main approaches used to handle a mask-occluded face, which are the restoration approach [3] and the discard-occlusion-based approach [4,5]. The restoration approach tries to restore the occluded parts of images based on the images in the training. However, these approaches need a dictionary where each subject should have an adequate image in the gallery. The discard-occlusion-based approach rejects the occluded parts completely to avoid a bad reconstruction process, and then, the remaining parts of the face are used in the feature extraction and classification processes. These discard-occlusion approaches devote a lot of effort to detecting and discarding occlusion regions. Most current advanced age estimation approaches are designed based on deep learning, which depends on a large number of training samples. Accordingly, deep masked face estimation systems also need a rich masked face dataset, which is currently not available. Age estimation from masked face images is quite different from the orthodox age estimation approach. Firstly, there is

no large benchmark masked face dataset annotated with proper age labels. Secondly, the features of the mouth and nose regions are severely damaged so that effective features are greatly reduced. Finally, faces wearing masks are hard to detect.

To handle the above challenges, we built a synthetic masked face image dataset derived from the largest face image dataset, IMDB-WIKI [6], using the dlib machine learning library. MTCNN [7] was incorporated to detect face areas, face key point detection, and align the face. Our demonstrated method took a face mask as a feature in addition to the full face and deployed deep convolutional neural network (CNN)-based models to address the problem of masked face age estimation. It is obvious that CNN-based methods have strong robustness to illumination and facial occlusion changes. Finally, we conducted experiments on the hierarchical approach and compared it with the singular approach. We also released source codes and data freely available in GitHub repository https://github.com/MahbubSohel/Masked_Face_Age (accessed on 16 May 2022) for keen researchers to studies further.

Overall, the contribution of this study is summarized as follows:

- We derived a synthetic face mask image dataset annotated with real age from the largest in-the-wild face image dataset, IMDB-WIKI, toward the estimation of real age amid the COVID-19 pandemic circumstances.
- We somewhat mitigated the data sparsity problem from the existing in-the-wild benchmark dataset, IMDB-WIKI.
- We proposed a hierarchical approach to estimate real ages in a group-specific manner, where the intermediate result was a group age generated by the classification model, and the final result was a real age generated by the regression-through-classification techniques generated by the group-specific models.
- We compared our hierarchical approach results with a one-off model and empirically demonstrated that a hierarchical approach is marginally better than a singular approach.
- Moreover, this research is perhaps the maiden attempt toward real age estimation from masked face images.

The rest of the article is divided into the following sections: In Section 2, the literature related to age estimation is presented. In Section 3, we present our proposed masked age estimation approach, describing all the components of this research. Section 4 demonstrates the empirical results. The comparative discussion and conclusion of this research are presented in Sections 5 and 6.

2. Related Work

Even though there is no literature available regarding masked face age estimation (MFAE), this section succinctly reviews the associated works of facial age estimation. Facial age estimation (FAE) has been a trending research topic among computer vision researchers for more than a decade. Similarly, face recognition (FR) also deals with facial features to recognize a person by looking at his/her face. Due to the COVID-19 pandemic, some face recognition research has already expanded to the masked face recognition arena. We try to provide a cursory outline of the way in which these tasks were approached earlier by other researchers.

Currently, FR research approaches mainly focus on deep-learning-based solutions to recognize a person's face through significant systems such as DeepFace [8], DeepID [9], ArcFace [10], and CosFace [11]. Although deep-learning-based approaches achieved enviable performance, they are still deficient when dealing with faces in unconstrained settings (e.g., occluded faces). Facial occlusion due to wearing a mask is a challenging problem in the event of face recognition, facial age estimation, etc. The occluded face recognition problem is generally approached by researchers in the following ways: image reconstruction [3], occlusion discarding [4,5], and deep face recognition. In addition to the above-stated approached, some deep-learning-based approaches [12,13] address face recognition by uti-

lizing a generative adversarial network and attention-based approach convolutional block attention model [14] to simplify and focus on more discriminative and expressive features.

In parallel, academicians also commit their best efforts to creating face-feature-based methods for demographic estimation. A recent survey of age estimation including all approaches from the last few decades can be found in [15]. There are various prominent age estimation models that use different image representation techniques, such as anthropometric models [16], the active appearance model (AAM) [17], active shape models (ASMs) [18], the aging pattern subspace model (AGES) [19], models based on age manifold [20], and appearance models [21]. From another viewpoint, the conventional classification approach is not solely explored as a method for facial age estimation. It can be categorized as either a multi-class classification problem [6], regression problem [22], or hierarchical problem [23,24] as an integration of classification and regression. The lion's share of age-estimation-based background studies that concern classification [25–27], regression [28–34], and hybrid [35–39] problems are reported in the references. In [19], aging features were extracted from an aging pattern representative subspace, and a robust regressor was used to determine age. In [25], the authors generated a face parametric statistical model and evaluated it with classifiers formulated on a shortest-distance neural network, and used quadratic function to model the correlation between age and face parameters. In [38], local binary pattern (LBP) histograms were used as features for classification. In [29], a multi-level local binary pattern (MLBP) was made up of several single-level LBPs and finally concatenated to extract local and global texture features. Guo et al. [34,35] introduced locally adjusted robust regression (LARR), where manifold learning was likely for feature extraction, and then, LARR was adopted for age prediction.

In the last decade, researchers have changed their direction of study toward deep learning, especially convolutional neural networks (CNNs), which are remarkably effective and promising compared to typical age estimation techniques. These methods can automatically learn different image representations and provide plausible results for demographic estimation [40]. The Deep Learned Ageing (DLA) pattern is a six-layer CNN framework proposed by [41], where Principal Component Analysis (PCA) is applied after CNN feature extraction. Manifold learning was employed to form the aging pattern and structure extracted from different layers. Ref. [42] demonstrated better performance in age group classification over the in-the-wild benchmark Adience dataset by using a simple CNN structure. Ref. [27] reviewed texture and appearance descriptors in detail to explain the significance of the fusion of both features to estimate age. Ref. [43] adopted the score fusion of regression and classification models based on real values. Additionally, a general-to-specific transfer learning strategy was applied to avoid the overfitting problem on a small apparent age estimation dataset. Ref. [44] designed an ensemble of CNN models wherein different models were trained for different age groups, and at the end, they were combined to achieve better efficiency. Ref. [45] proposed a three-stage cascaded CNN to deal with unconstrained face images in the order of age group classification, apparent age estimation by considering the mean relative age found from every single group, and an age error correction strategy. Ref. [46] classified multiple binary sub-problems and attained them through ordinal regression problems. Biologically inspired features (BIFs) were used for feature extraction tasks. Ranking of a CNN framework was proposed by [47], where a single image was considered while designing the order of age group sub-networks and ranking all these binary outputs to predict the final age label. Ref. [48] proposed an end-to-end age estimation method, incorporating a novel cumulative hidden layer to reduce sample imbalance problems from neighboring ages. In [49], the authors suggested a CNN model competent to work in an unconstrained environment, where attention CNN predicted the best attention grid and Patch CNN identified high-resolution patches for the task of classification using a multi-layer perceptron.

In recent years, Ref. [50] unified the strength of a CNN and an extreme learning machine (ELM) for age estimation and gender classification. Ref. [51] proposed a framework of five cascaded structures for age prediction using gender and race information in the

form of Gaussian Process Regression (GPR) instead of linear regression. Ref. [52] designed a conditional multi-task learning system with weak label expansion for real age predictions from the gender of a person. Ocular images captured from smartphones were utilized for age classification in [53,54], which combined multi-stage learned features in the form of score-level fusion and handcrafted features as the feature-level fusion of facial images. In [55], the authors presented and evaluated an age estimation approach in unconstrained images using facial parts (eyebrows, eyes, nose, and mouth), cropped from the input images using landmarks, to feed a compact multi-stream CNN architecture. A CNN-based architecture for a joint age–gender identification method was proposed by [56]. For the CNN network, Gabor filter responses were used as inputs. Ref. [57] proposed a two-stage approach, where the CNN predicted age and gender by using a modified MobileNet. Ref. [58] proposed a novel method based on attention long short-term memory (ALSTM) combined with a ResNet network for fine-grained age estimation. Ref. [59] introduced a novel end-to-end two-level CNN approach to identify age and gender from unfiltered faces. A lightweight CNN network with a mixed attention mechanism for low end-devices was proposed in [60], where the output layer was fused by the classification and regression approach. Another multi-task learning approach merging classification and regression concepts to fit the age regression model with heterogeneous data with the help of two different techniques for partitioning data towards classification was proposed in [61]. To resolve the problem of data disparity and ensure the generality of the model, a very recent method was proposed by Kim et al. [62], where a cycle-generative-adversarial-network-based race and age image transformation method was used to generate sufficient data for each distribution. All of the aforementioned CNN-based systems were evaluated based on the constrained dataset Morph for age estimation. In our recent paper [63], we demonstrated better MAE over the Morph dataset by pretraining a CNN network with the in-the-wild face image dataset, IMDB-WIKI. We ensured relative balancing among the 101 age classes that exist in the IMDB-WIKI dataset towards real age estimation. Although a lot of research has been carried out on age estimation from face images, none of the research focuses on masked face estimation. Being motivated by masked face recognition research amid the COVID-19 pandemic, we pursued our maiden attempt to estimate real ages from human masked face images.

3. Proposed Method

Our proposed method was utilized during our experiments for real age estimation. The proposed hierarchical approach was composed of two stages. In the first stage, we performed group age classification, where 101 age classes were divided into three broad age groups named child (0–15), adult (16–64), and elderly (65+) based on age dependency ratio 2020. After obtaining the predicted age group, the sample image was computed through a particular network trained for each group in the final stage, utilizing the technique of regression-through-classification. The idea behind this approach is to reduce the error space, as the masked face real age estimation system consists of a significant amount of age classes with less distinctive features. Moreover, the samples are not equal in every age class, so the training can be biased to specific groups with required image samples in a balanced fashion. Each step of the proposed approach shown in Figure 2 is depicted thoroughly in this section.

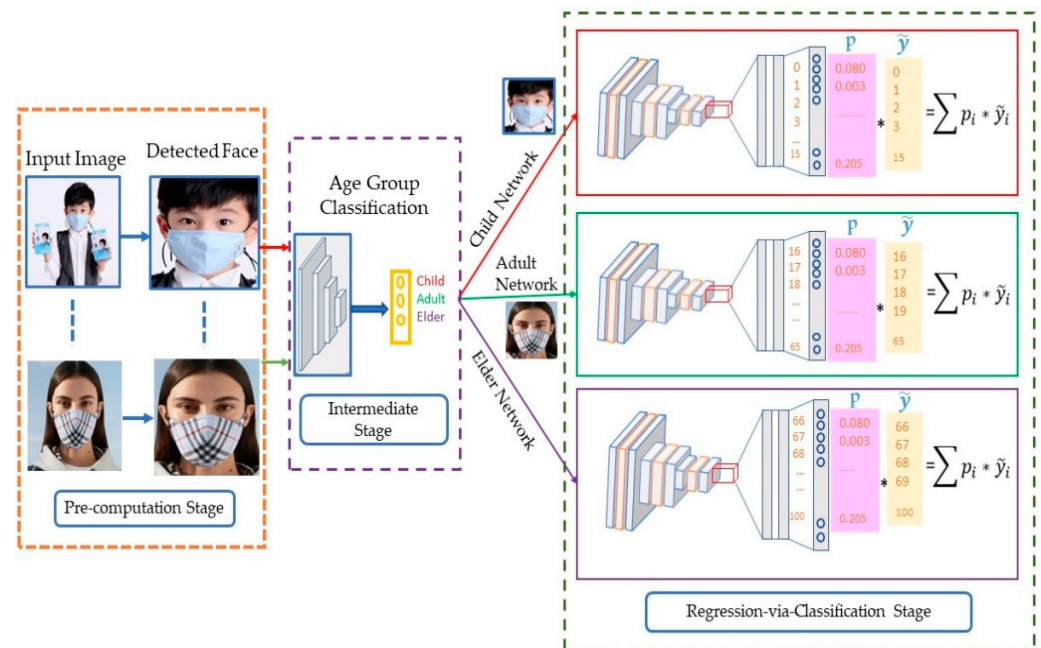


Figure 2. Schematic diagram of the proposed masked face real age estimation method.

3.1. Data Sparsity Mitigation

Computer vision scholars acknowledged that facial age estimation in masked and unmasked faces is a complex problem in this research area. In particular, masked face age estimation has evolved as an obligatory research topic because of the pandemic. It is known that solving a complex subject requires a deep structure of problem-defining components along with sufficient data to train. If not, the trained model will endure the overfitting issue. The main solution to the network overfitting problem is to ensure adequate data for model training. As a matter of fact, there is a solo competent dataset, IMDB-WIKI, which builds upon an unconstrained setting while having ample facial images. We found that the IMDB-WIKI dataset is well-suited for training our age estimation deep network. Although this dataset is rich enough in the context of samples, the inter-class disparity is a concern. After screening this dataset, we observed that data in the age range 15~65 were rationally incomparable with other age classes, which may have misled the empirical results. As a solution to the data sparsity problem that exists in the aforementioned dataset, the following steps can be useful:

- Set a threshold for choosing the image samples from each class considering the data ratio all over the age classes.
- Remove the erroneously annotated samples of each age class using human visual perception, which is shown in Figure 3.
- Enhance the image for the classes with very limited samples by (1) appending images from the very pertinent Adience dataset [64] and (2) oversampling through data augmentation operations such as flipping, rotating in the range of $-30^\circ \sim 30^\circ$ maintaining the steps of 5° , and scaling to ensure the threshold value for the class balancing.



Figure 3. Example of wrongly age annotated image samples in the existing IMDB-WIKI benchmark dataset.

Additionally, some online data augmentation was also performed by rescaling the images into 256×256 pixels, and a crop of 224×224 pixels from the center was input to the network during training as a measure to alleviate the overfitting issues as well as the robustness of the model. The subsequent data distribution of the underlying age classes prior to and after initiating parity in the IMDB-WIKI dataset is presented in Figure 4.

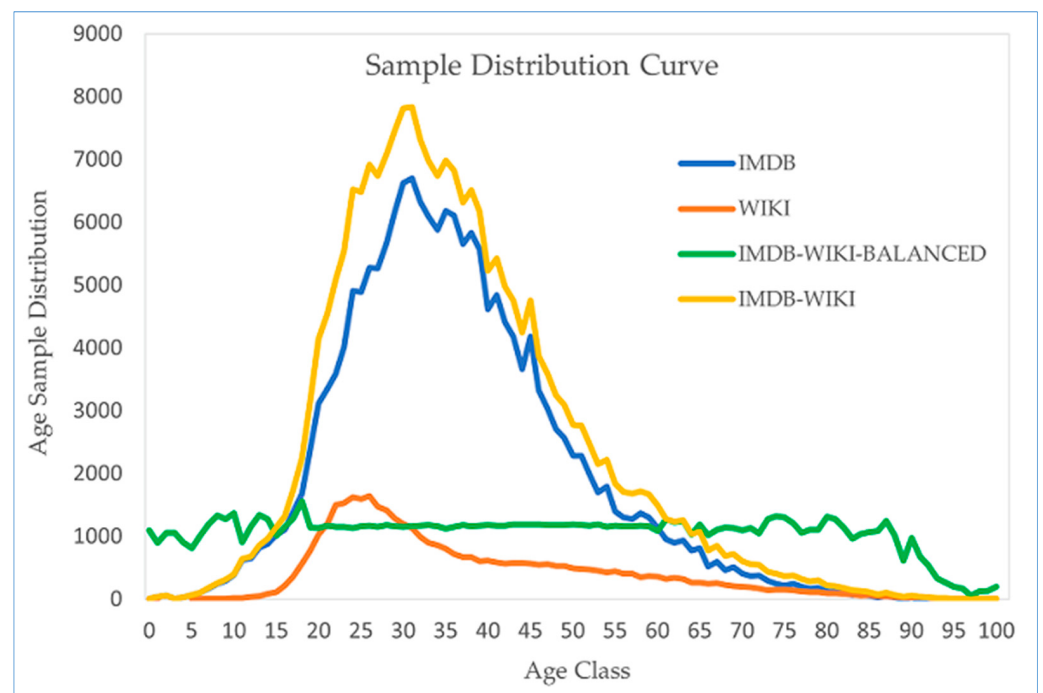


Figure 4. Sample distribution of the revised IMDB-WIKI dataset prior to and following uniformity amongst classes.

3.2. Face Detection and Alignment

Human face detection becomes an extremely difficult task when there are huge appearance variations and external factors. A standard network expects an input image of identical size, centered position, and less background. As a practice, the majority of age estimation systems perform face detection as an essential initial step because mining discriminative facial features helps the systems to make final decisions. There are two major

parts, named detection and alignment. Hence, a cascaded multi-task framework [7] is employed for face detection, where detection and alignment maintain an implicit relationship for performance gain. The multi-task cascaded convolutional neural network (MTCNN) face detection process is graphically visualized in Figure 5. Since deep CNNs are effectively capable of dealing with tiny alignment errors, our goal was to choose a robust face detector ensuring face alignment just via a marginal up-frontal rotation [6]. As a result, the overall age error was reduced while using the detected face as input rather than a whole image.

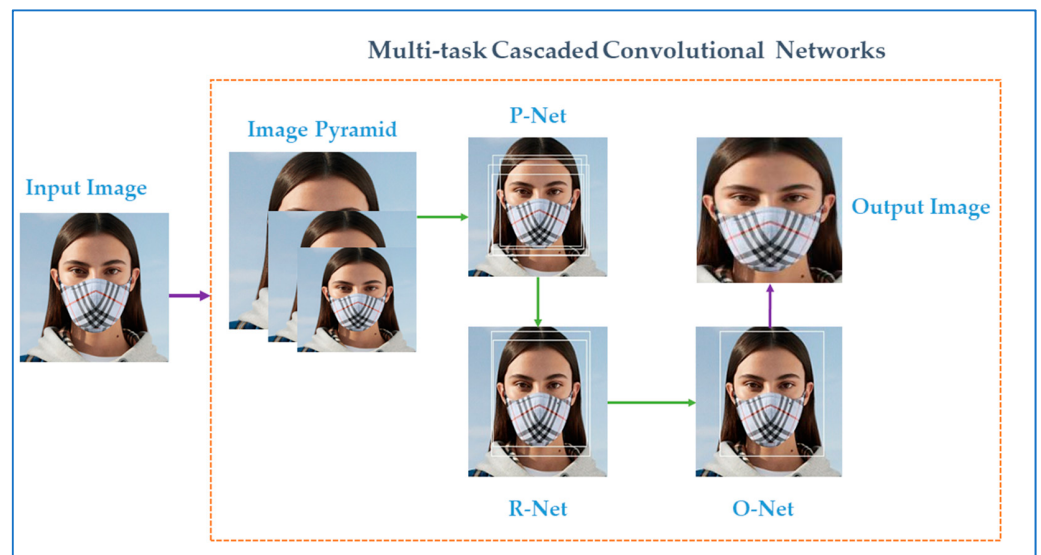


Figure 5. The entire face detection process over raw face images using MTCNN.

Although the chosen face detectors mostly successfully detected faces, they missed faces in a few cases. In the case of failure, the whole image was considered for network input. Furthermore, some extra context all around the face also boosted the system's performance. Thus, a 20% margin on every side was added with the detected face, and border pixels were merely repeated while there was no context.

3.3. Regression via Classification

Conventionally, the prediction of age falls into the category of a regression problem because age represents a continuous value, preferably a set of discrete classes. The off-the-shelf pre-trained models evaluated for ImageNet classification consist of neurons for every object class, normalized employing the Softmax function. In the case of the regression task, we employed a Euclidean loss function with the replacement of the output layer with one neuron. Unfortunately, training a CNN solely for age regression tasks experiences a high error rate due to the instability while handling outliers. Consequently, the model convergence had difficulty owing to high gradients and unstable predictions as well.

Under these circumstances, we considered the age regression task via a standard classification approach where the ages were discretized into K categories. Hence, the Softmax function was used in the output layer of the network to normalize the arbitrary output values into probabilities across predicted age classes. Equation (1) describes the procedure of the probability generation for each labeled class by the Softmax activation function, mentioned below:

$$S(v)_i = \frac{e^{v_i}}{\sum_{j=1}^N e^{v_j}} \quad (1)$$

where i denotes the present element index of the input vector v , N is the total number of classes for the particular task, and all v values are the elements of the input vector.

Following this procedure, the learning conducted by the CNN model for age classification and quantifying the regression value out of the expected value formulated by means of Softmax probabilities belongs to the K neurons, as shown in Equation (2).

$$E = \sum_{i=1}^{|k|} y_i \cdot p_i \quad (2)$$

where k belongs to 101 age categories, $y_i \in [0, 100]$, and $i = 1, 2, \dots, 101$. p_i denote the Softmax-normalized output probability of neuron i . From the experimental results, it is proven that robustness throughout training and accuracy during testing is ensured by this formula.

3.4. Deep Learning Models

Since the second decade of the twenty-first century, deep learning methods have shown their pervasiveness in all sectors of academia and industry. Accordingly, these technologies proved primacy in learning multiple levels of facial representations that were persistent to facial changes due to external factors. Convolutional neural networks are amongst the most effective deep learning class and have shown primacy in a wide range of image-based applications such as image classification, face recognition, object detection, etc. A convolutional neural network comprises several computational basic building blocks named convolution, pooling, and dense layers, projected to learn spatial hierarchies of features automatically and adaptively through a backpropagation algorithm.

Therefore, we employed several convolutional neural networks to predict the age of a human exclusively from a single face image. This network used an aligned face with the background as input, and the output was the predicted real age. In our system, we used multiple popular, lightweight, and high-accuracy-based pre-trained CNN architectures named ResNet [65], WideResNet [66], DenseNet [67], MobileNet [68], ShuffleNet [69], and SqueezeNet [70]. The reasons for choosing these architectures were that they (1) are deep and tractable, (2) were winners of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in different years, (3) maintain a trade-off between performance and prediction time, and (4) concede reasonably good starts for model training. Table 1 reviews the prime attributes of the widely held CNN models used in human age estimation from masked face images.

Table 1. A summary of deployed pre-trained convolutional neural networks.

Model	Parameters	Size	Depth
ResNet 50	25.6 M	96 MB	50
Wide ResNet-50-2	68.9 M	131 MB	-
DenseNet-121	8.1 M	33 MB	121
MobileNet-v2	3.5 M	13 MB	53
ShuffleNet	1.4 M	5.4 MB	50
SqueezeNet	1.24 M	5.2 MB	18

3.5. Synthetic Masked Face Generation

In order to overlay face masks, we needed to perform face detection first. For face detection, we used dlib's frontal face detector which is based on histograms of oriented gradients (HOGs) and a linear SVM. Subsequently, we detected facial landmarks using facial landmark predictor `dlib.shape_predictor` from dlib library [71], which estimates the location of 68 (x, y) -coordinates that map to facial structures on the face. Facial landmarks are used to localize and represent salient regions of the face, such as eyes, eyebrows, nose, jawline, mouth, etc. It is a technique that has been applied to applications such as face alignment. Once the facial landmarks were detected, we were then able to start overlaying the face masks on the faces by joining the required points using the OpenCV drawing

functions. Some of the face images after overlaying masks from IMDB-WIKI datasets are presented in Figure 6.

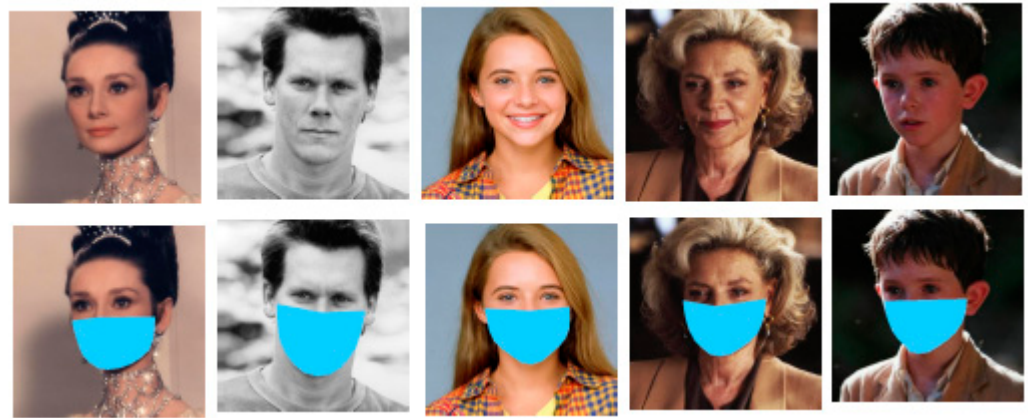


Figure 6. Exemplar images from IMDB-WIKI dataset after face mask overlay.

3.6. Performance Metric

Our masked face age estimation system was evaluated in a qualitative and quantitative fashion. As a quantitative performance measure, we used two standard metrics named mean absolute error (MAE) and cumulative score (CS). The predicted age and corresponding age ground-truth with the face image are presented as qualitative standards for the facial age estimation task.

MAE: The system performance is reported as the mean absolute error in years. It calculates the mean of the absolute error amongst the predicted and ground-truth age values. A de facto standard for evaluating the performance of age estimation is standardized as MAE because most researchers choose it for the system evaluation. The MAE is calculated using Equation (3).

$$MAE = \frac{\sum_{i=1}^N |y_i^p - y_i^{gt}|}{N} \quad (3)$$

where N is the sum of test images, y_i^{gt} denotes the ground-truth age, and y_i^p implies the predicted age of the i th image, respectively.

CS: The cumulative score is a measure based on the age error threshold value. It represents the percentage of test samples with an absolute age error less than a threshold value t among overall test samples. The cumulative score is described below by Equation (4):

$$CS = \frac{N_{ae < t}}{N} * 100\% \quad (4)$$

where $t \in [0, 10]$ and $N_{ae < t}$ is the count of test images that possess an absolute age error less than the corresponding threshold value. The overall count of test images is denoted as N .

4. Experiments and Results

4.1. Implementation Details

We learned six different CNN models separately in singular and hierarchical manners for the facial age classification task. The deep learning framework PyTorch [72] from Facebook's artificial research lab was used for network training. For faster processing and to facilitate a parallel computing platform, the Nvidia GeForce GTX 1080 Ti graphics processing unit (GPU) with 3584 CUDA cores and 11 GB of video memory was utilized. The overall training time for the large IMDB-WIKI dataset took almost half a day to train a deep network, whereas a light network only took a couple of hours.

For the age classification, we computed the expected age value using the Softmax probabilities corresponding to output neurons, similar to [6]. We trained the models designed in a singular and hierarchical fashion with a relatively balanced, synthetically generated masked face IMDB-WIKI dataset. Every age was considered as an individual class that ranged from 0 to 100 in the singular approach. Likewise, one model for age groups and three specific models were trained for individual groups in the hierarchical approach. We finetuned the deployed models with our experimented dataset and reshaped the network with a new output layer. In finetuning, we started with a pre-trained model and updated all the model's parameters for our task. Although both the source and target network performed classification, we performed finetuning rather than feature extraction because real age classification is a bit different from other object classification techniques. Hence, the output layer was trained from scratch, while the parameters of all the other layers were finetuned based on the parameters of the source model.

For all experiments regarding masked face real age estimation, the CNN was initialized with the weights trained on ImageNet [73]. This pre-trained model was then further finetuned on the IMDB-WIKI image dataset for real age prediction. Finally, the CNN was tested with the in-the-wild synthetically masked face image dataset IMDB-WIKI. Every training set in either age estimation approach utilized 80% of the images for training, and the rest, 20%, was used for the system evaluation as a validation set and test set. We used separate validation and test sets that equally comprised 20% of the total samples to validate the system. We maintained the de facto data split strategy to ensure that the validation and testing set minimally contained data from all the age classes. We used weighted cross-entropy to ensure the same weight for all classes during the course of learning. The training was optimized with Adam optimizer, keeping the momentum value of 0.9, and a standard weight decay rate was maintained to ensure the regularization. The learning rate was adjusted, while the loss values did not decrease for five consecutive epochs.

4.2. Dataset

In this paper, we generated a synthetic masked face image dataset from one of the largest in-the-wild face image benchmark datasets for masked face real age estimation. A brief description of this dataset that mentions the specifications is primarily introduced in this section.

IMDB-WIKI: IMDB-WIKI is a leading publicly available demographic annotated face image dataset. It is mainly a celebrity face image dataset crawled from the IMDb website and Wikipedia, and comprises nearly 500 k samples covering a wide age range of 0~100. The age is determined by the person's date of birth and the crawling timestamp. Around 88% of the half-a-million images are collected from the IMDb website, and the rest from another source. This dataset has a percentage of inferior images (i.e., funny, sketch, occluded, several subjects in the same frame, cluttered background, and black) that may mislead the system training. As the first step of our experiment, we removed the noisy labels and the images containing multiple subjects from each age class. Thenceforth, we followed the down-sampling and up-sampling strategies to maintain a close ratio of the samples in every class. Finally, we generated a synthetic face mask dataset from the existing samples in the revised IMDB-WIKI dataset. Hence, a total of 107k images remained available for the model training (80%) and evaluation (20%).

4.3. Results

The quantitative results of our proposed masked face real age estimation system are reported in this section. Our reported age estimation result is in the form of MAE. We evaluated our system on the synthetically generated masked IMDB-WIKI dataset to estimate the real/biological age of a person. In Table 2, we present the mean absolute error values achieved after a massive experiment was conducted, utilizing prominent off-the-shelf deep convolutional neural network models. From the comparison table, it is clearly evident that in all cases of experiments, hierarchical approaches perform marginally

better than the singular approach. The comparative results of deployed CNN architectures are graphically presented in Figure 7.

Table 2. Mean absolute error values regarding real age estimation over synthetic masked IMDB-WIK dataset.

Model Name	Singular Approach	Hierarchical Approach
ResNet 50	5.13	4.35
WideResNet	4.81	4.20
DenseNet	5.38	4.52
MobileNet	5.85	5.25
ShuffleNet	6.17	5.70
SqueezeNet	6.88	6.70

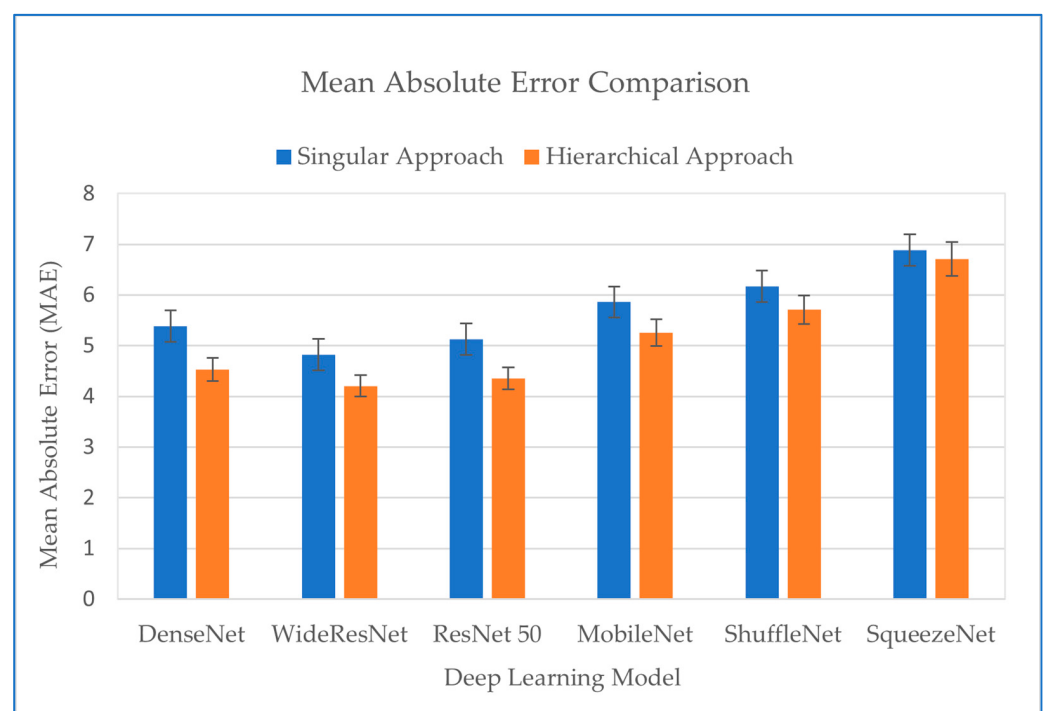


Figure 7. The performance comparison of different CNN models in real age estimation.

We investigated the performance of the deployed pre-trained CNN models for each particular age group (child, adult, and elder), which is quantitatively shown in Table 3 and graphically presented in Figure 8. A pictorial presentation that mentions the quantitative performance (MAE) found over the synthetic masked face dataset, IMDB-WIKI, through both approaches is reported in Table 4.

Table 3. Age-group-wise mean absolute error comparison of the experimented convolutional neural networks regarding real age estimation over synthetic masked IMDB-WIK dataset.

Model Name	Child Group	Adult Group	Elder Group
ResNet 50	0.80	4.70	1.52
WideResNet	0.81	4.41	1.64
DenseNet	0.92	4.63	2.04
MobileNet	1.03	5.13	2.49
ShuffleNet	1.10	5.98	2.05
SqueezeNet	1.35	6.41	3.23

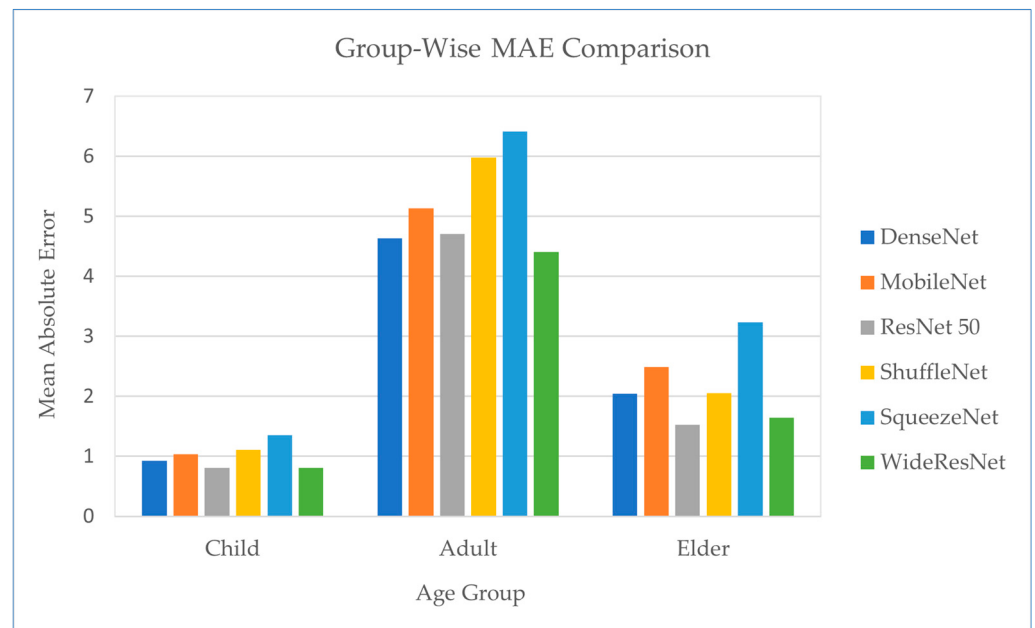


Figure 8. The performance of deployed models over age groups in terms of mean absolute error.

Table 4. Sample images with corresponding mean absolute error evaluated by the best model.

Real Age	25	14	72	93
				
Singular/Hierarchical	21.61/25.01	17.24/14.94	63.57/73.06	90.52/92.99

In Table 5, we compare our system performance with the existing methods concerning the mean absolute error achieved irrespective of whether faces are masked or non-masked. We only compare this with systems that estimate the biological age of a person rather than the apparent age or age group. We achieved better MAE in contrast to the works AGES [19], Kim et al. [63], OHRank [74], and CA-SVR [75], although our system handles face images with masks.

Table 5. Performance comparison with existing prominent facial age estimation methods.

Method	Dataset	Number of Images	Number of Subjects	Wearing Mask	MAE
AGES [19]	Morph	55,134	13,618	No	8.83
Kim et al. [63]				No	4.29
OHRank [74]				No	6.07
CA-SVR [75]				No	5.88
Proposed	Imdb-Wiki	523,051	20,284	Yes	4.20

As stated in the calculation procedure of the cumulative score (CS), the CS values on the synthetic masked face image dataset under distinct error thresholds are plotted in Figure 9. From the figure, a steady growth in CS value is observed if the allowable error thresholds increase. A comparison of cumulative scores between singular and hierarchical approaches is graphically presented in Figure 10. In this comparison curve, the score achieved using the WideResNet model is plotted, where the hierarchical approach exhibits

steady performance across all error threshold values. It is worth mentioning that in the early stage of the curve, the hierarchical approach contains more test samples with less errors than the singular approach and gradually performs steadily considering more error threshold values.

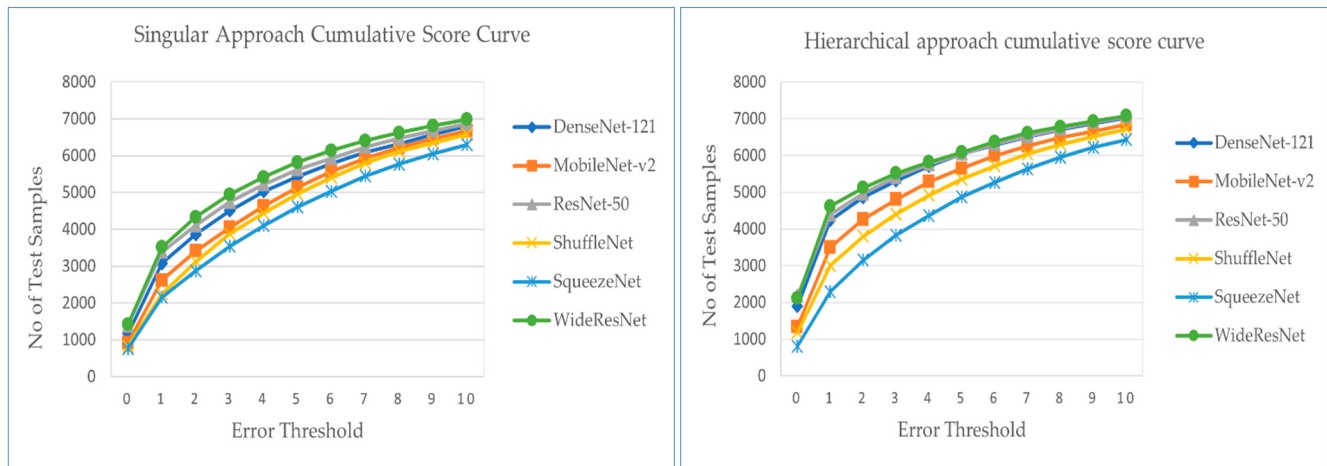


Figure 9. Plotted cumulative score of all CNNs deployed for real age estimation task using singular and hierarchical approach.



Figure 10. Cumulative score comparison among singular versus hierarchical approach designed for the real age estimation task using the WideResNet (WRN) model.

We conducted a statistical test to observe the significant differences between the two approaches (singular and hierarchical), because this allowed us to ensure the robustness of the model performance. We chose the Wilcoxon Signed-Rank test because our experimented observation does not follow the normal distribution. The null hypothesis with the Wilcoxon Signed-Rank test is that the pairwise statistics (e.g., mean) between the observations from both (singular and hierarchical) approaches are equal. The calculated p -values with a confidence level of 5% validate that we would reject the null hypothesis, concluding that the means between the groups are different. Since the p -values contain long fractional parts, we visualize the p -value result in Figure 11, which is similar to work [76], where $-\log_{10}(p - \text{value})$ is presented rather than the actual p -value.

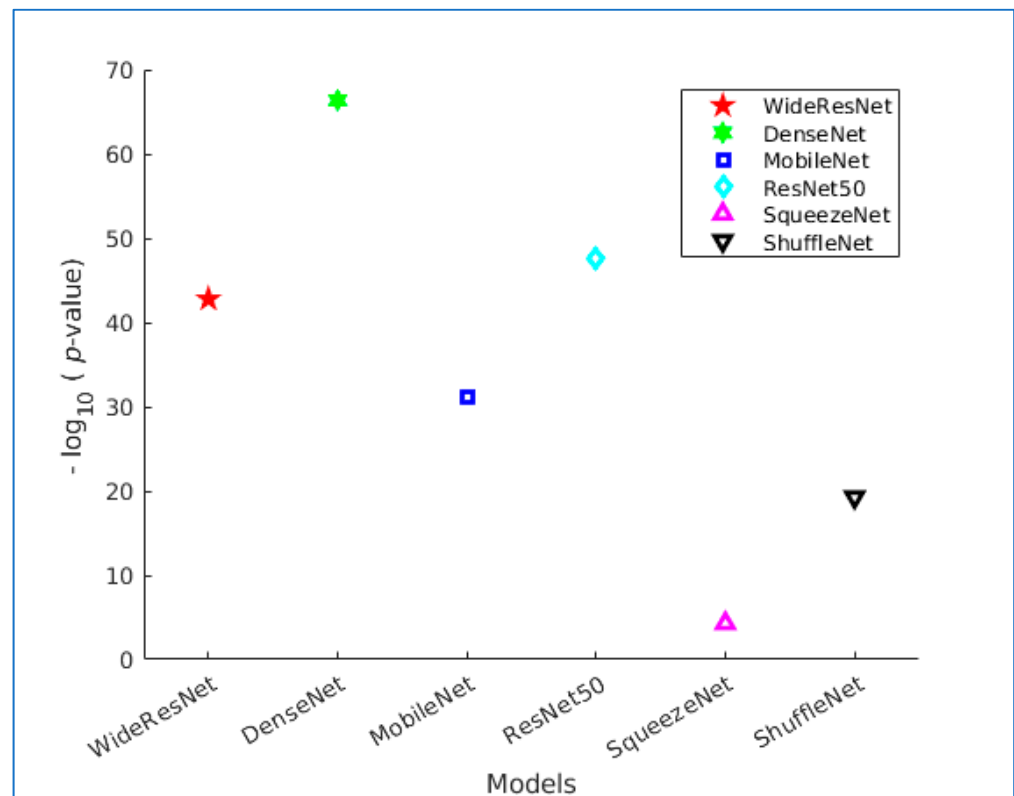


Figure 11. Wilcoxon Signed-Rank statistical significance test analysis over the deployed models between the proposed hierarchical approach and conventional approach for masked face age estimation. In every test, models give a p -value of less than 0.05 which statistically proves that the model's performances are significantly different. The p -values are presented in the form of $-\log_{10}(p\text{-value})$ for clear perception.

5. Discussion

This article proposed a deep-learning-based hierarchical approach for estimating the biological age of a person wearing a face mask in a group-specific manner. In the hierarchical approach, the error spaces were squeezed by introducing the group classification approach prior to the age regression stage of the computation. Firstly, we classified the sample of 101 age classes into three major age groups named child, adult, and elder. Next, the specific network designed for each age group carried out the age estimation process through regression-via-classification strategies described in Section 3.

In the system design, we addressed two crucial types of deep learning model uncertainty: epistemic uncertainty and aleatoric uncertainty. Epistemic uncertainty occurred due to data inadequacy and systematic bias while collecting samples observed partially because some age classes (i.e., 1, 3, 4, or 97) contained very few image samples (<100) for training. Similarly, there was some randomness in the input data due to pixel noise, and the incorrect annotation caused aleatoric uncertainty. We tried to mitigate both the model uncertainties by data augmentation, data supplementation, and manually filtering out the noisy data as much as possible.

The proposed hierarchical approach demonstrates comprehensive results over the synthetic IMDB-WIKI masked face image dataset in anticipation of real age estimation. From the experimental results presented in Figure 7, it is quite evident that the hierarchical approach is marginally better than the singular approach, apart from with regard to deep learning model selection. If we consider the cumulative score graph presented in Figure 10, as the error threshold value is less, the difference between the approaches is quite evident. Furthermore, the performance of the models while experimenting on individual age group

is presented quantitatively and graphically in Section 4. Our models demonstrated enviable performances in the case of the child and adult group, whereas they were still deficient when handling adult faces, since the changes in features among adults are relatively stable and indifferent.

It is worth mentioning that lightweight models demonstrate a slightly higher MAE than deeper models but are competent enough for deployment in low-end devices. Additionally, we performed a statistical significance test to ensure the system's robustness and proved that the performance achieved through hierarchical and singular approaches was significantly different (see Section 4 and Figure 11).

In future research, we will try to build a new face mask dataset by maintaining parity among the age classes since the amount of data in every age is insufficient to train a deep learning model for a complex problem such as real age estimation. We will try to lessen the mean absolute error in the adult group, which was somewhat higher than the other groups formulated in this experiment.

6. Conclusions

In this paper, we proposed a biological age estimation system from masked face images amid the COVID-19 pandemic that has been taking place since 2020. As we know, due to the pandemic, people should wear a face mask, which necessitates the migration of existing face-based real-world applications. Hence, the proposed research could be a promising solution for real-world applications of those services based on the age of a person. Since there is shortage of research regarding demographic estimation from masked face images, the authors were motivated to conduct this research by following research regarding masked face recognition. We tried to tackle the masked face age estimation problem in a group-specific manner, where age group classification is performed prior to age regression via classification. We performed a comparative analysis between the singular and proposed hierarchical approach in terms of the mean absolute error of age. The empirical results proved that the hierarchical approach is marginally better than the singular approach, comprising only one model for masked face real age estimation. It is an insightful observation that the proposed approach performs better considering the low error threshold value shown in the cumulative score comparison curve. As this is perhaps the maiden attempt toward masked face real age estimation, the authors tried to give an empirical outline utilizing multiple accuracies versus memory-efficient convolutional neural networks and prove the superiority of the proposed hierarchical approach.

Author Contributions: Conceptualization, M.M.I. and J.-H.B.; methodology, M.M.I.; software, M.M.I.; validation, M.M.I. and J.-H.B.; formal analysis, M.M.I.; investigation, M.M.I.; data curation, M.M.I.; writing—original draft preparation, M.M.I.; writing—review and editing, J.-H.B. and M.M.I.; visualization, M.M.I.; supervision, J.-H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the GRRC program of Gyeonggi province (GRRC Aviation 2017-B04, Development of Intelligent Interactive Media and Space Convergence Application System).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The authors have used the publicly archived IMDB-WIKI and Adience datasets for the experiments. The IMDB-WIKI dataset is available in [6]. The Adience dataset is available in [65].

Acknowledgments: We would like to acknowledge Korea Aerospace University with much appreciation for its ongoing support to our research. We are thankful to the authors who made their datasets publicly available to pave the way for research in this area.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Le, N.Q.K.; Ho, Q.-T. Deep transformers and convolutional neural network in identifying DNA N6-methyladenine sites in cross-species genomes. *Methods*, 2021; *in press*. [\[CrossRef\]](#)
2. Le, N.Q.K.; Ho, Q.-T. Identifying SNAREs by Incorporating Deep Learning Architecture and Amino Acid Embedding Representation. *Front. Physiol.* **2019**, *10*, 1501. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Gawali, A.S.; Deshmukh, R.R. 3d face recognition using geodesic facial curves to handle expression, occlusion and pose variations. *Int. J. Comput. Sci. Inf. Technol.* **2014**, *5*, 4284–4287.
4. Priya, G.N.; Banu, R.W. Occlusion invariant face recognition using mean based weight matrix and support vector machine. *Sadhana* **2014**, *39*, 303–315. [\[CrossRef\]](#)
5. Alyuz, N.; Gokberk, B.; Akarun, L. 3-d face recognition under occlusion using masked projection. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 789–802. [\[CrossRef\]](#)
6. Rothe, R.; Timofte, R.; Gool, L.V. Deep expectation of real and apparent age from a single image without facial landmarks. *Int. J. Comput. Vis.* **2018**, *126*, 144–157. [\[CrossRef\]](#)
7. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [\[CrossRef\]](#)
8. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1701–1708.
9. Sun, Y.; Chen, Y.; Wang, X.; Tang, X. Deep learning face representation by joint identification-verification. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 1988–1996.
10. Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4690–4699.
11. Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Li, Z.; Liu, W. Cosface: Large margin cosine loss for deep face recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5265–5274.
12. Duan, Q.; Zhang, L. Look more into occlusion: Realistic face frontalization and recognition with boostgan. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 214–228. [\[CrossRef\]](#)
13. Li, Y.; Guo, K.; Lu, Y. Cropping and attention based approach for masked face recognition. *Appl. Intell.* **2021**, *51*, 3012–3025. [\[CrossRef\]](#)
14. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
15. Punyani, P.; Gupta, R.; Kumar, A. Neural networks for facial age estimation: A survey on recent advances. *Artif. Intell. Rev.* **2020**, *53*, 3299–3347. [\[CrossRef\]](#)
16. Farkas, L.G. *Anthropometry of the Head and Face*; Raven Press: Ely, MN, USA, 1994. [\[CrossRef\]](#)
17. Cootes, T.F.; Edwards, G.J.; Taylor, C.J. Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 681–685. [\[CrossRef\]](#)
18. Cootes, T.F.; Taylor, C.J.; Cooper, D.H.; Graham, J. Active shape models—Their training and application. *Comput. Vis. Image Underst.* **1995**, *61*, 38–59. [\[CrossRef\]](#)
19. Geng, X.; Zhou, Z.; Smith-Miles, K. Automatic age estimation based on facial aging patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 2234–2240. [\[CrossRef\]](#)
20. Fu, Y.; Xu, Y.; Huang, T.S. Estimating human ages by manifold analysis of face pictures and regression on aging features. In Proceedings of the IEEE Conference Multimedia and Expo, Beijing, China, 2–5 July 2007; pp. 1383–1386.
21. Beymer, D.; Poggio, T. Image representations for visual learning. *Science* **1996**, *272*, 1905–1909. [\[CrossRef\]](#)
22. Dornaika, F.; Bekhouche, S.; Arganda-Carreras, I. Robust regression with deep CNNs for facial age estimation: An empirical study. *Expert Syst. Appl.* **2020**, *141*, 112942. [\[CrossRef\]](#)
23. Thukral, P.; Mitra, K.; Chellappa, R. A hierarchical approach for human age estimation. In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 March 2012; pp. 1529–1532. [\[CrossRef\]](#)
24. Pontes, J.K.; Britto, A.D.; Fookes, C.; Lameiras Koerich, A. A flexible hierarchical approach for facial age estimation based on multiple features. *Pattern Recognit.* **2016**, *54*, 34–51. [\[CrossRef\]](#)
25. Lanitis, A.; Draganova, C.; Christodoulou, C. Comparing different classifiers for automatic age estimation. *IEEE Trans. Syst. Man Cybern.* **2004**, *34*, 621–628. [\[CrossRef\]](#)
26. Ueki, K.; Hayashida, T.; Kobayashi, T. Subspace-based age group classification using facial images under various lighting conditions. In Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition, Southampton, UK, 10–12 April 2006; pp. 43–48.
27. Huerta, I.; Fernandez, C.; Segura, C.; Hernando, J.; Prati, A. A deep analysis on age estimation. *Pattern Recognit. Lett.* **2015**, *68*, 239–249. [\[CrossRef\]](#)
28. Guo, G.; Fu, Y.; Huang, T.S.; Dyer, C. Locally adjusted robust regression for human age estimation. In Proceedings of the IEEE Workshop on Applications of Computer Vision, Lake Tahoe, NV, USA, 12–15 March 2018; pp. 19–21. [\[CrossRef\]](#)

29. Nguyen, D.T.; Cho, S.R.; Park, K.R. Age estimation-based soft biometrics considering optical blurring based on symmetrical sub-blocks for MLBP. *Symmetry* **2015**, *7*, 1882–1913. [[CrossRef](#)]
30. Onifade, O.F.W.; Akinyemi, D.J. A groupwise age ranking framework for human age estimation. *Int. J. Image Graph. Signal Process.* **2015**, *7*, 1–12. [[CrossRef](#)]
31. Guo, G.; Mu, G. Joint estimation of age, gender and ethnicity: CCA vs. PLS. In Proceedings of the IEEE Conference on Face and Gesture Recognition, Shanghai, China, 22–26 April 2013; pp. 1–6.
32. Lu, J.; Tan, Y. Ordinary preserving manifold analysis for human age and head pose estimation. *IEEE Trans. Hum.-Mach. Syst.* **2012**, *43*, 249–258. [[CrossRef](#)]
33. Akinyemi, J.D.; Onifade, O.F.W. An ethnic-specific age group ranking approach to facial age estimation using raw pixel features. In Proceedings of the IEEE Symposium on Technologies for Homeland Security, Waltham, MA, USA, 10–11 May 2016; pp. 1–6.
34. Guo, G.; Fu, Y.; Dyer, C.; Huang, T. Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Trans. Image Process.* **2008**, *17*, 1178–1188. [[PubMed](#)]
35. Guo, G.; Fu, Y.; Huang, T.S.; Dyer, C. A probabilistic fusion approach to human age prediction. In Proceedings of the IEEE in Conference on Computer Vision and Pattern Recognition-Semantic Learning and Applications Multimedia Workshop, Anchorage, AK, USA, 23–28 June 2008; pp. 1–6.
36. Choi, S.E.; Lee, Y.J.; Lee, S.J.; Park, K.R.; Kim, J. Age estimation using hierarchical classifier based on global and local features. *Pattern Recognit.* **2011**, *44*, 1262–1281. [[CrossRef](#)]
37. Han, H.; Charles, O.; Liu, X.; Jain, A.K. Demographic estimation from face images: Human vs. machine performance. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1148–1161. [[CrossRef](#)]
38. Gunay, A.; Nabyev, V.V. Facial age estimation based on decision level fusion of AMM, LBP and Gabor features. *Int. J. Adv. Comput. Sci. Appl.* **2015**, *6*, 19–26.
39. Punyani, P.; Gupta, R.; Kumar, A. Human age-estimation system based on double-level feature fusion of face and gait images. *Int. J. Image Data Fusion* **2018**, *9*, 222–236. [[CrossRef](#)]
40. Yang, M.; Zhu, S.; Lv, F.; Yu, K. Correspondence driven adaptation for human profile recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 505–512.
41. Wang, X.; Guo, R.; Kambhamettu, C. Deeply-learned feature for age estimation. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 5–9 January 2015; pp. 534–541.
42. Levi, G.; Hassner, T. Age and gender classification using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, USA, 7–12 June 2015; pp. 34–42.
43. 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 Workshop, Santiago, Chile, 7–13 December 2015; pp. 258–266.
44. Malli, R.C.; Aygun, M.; Ekenel, H.K. Apparent age estimation using ensemble of deep learning models. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 714–721.
45. Chen, J.-C.; Kumar, A.; Ranjan, R.; Patel, V.M.; Alavi, A.; Chellappa, R. A cascaded convolutional neural network for age estimation of unconstrained faces. In Proceedings of the IEEE Conference on Biometrics, Theory, Applications and Systems, Niagara Falls, NY, USA, 6–9 September 2016. [[CrossRef](#)]
46. Niu, Z.; Zhou, M.; Wang, L.; Gao, X.; Hua, G. Ordinal regression with multiple output CNN for age estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4920–4928.
47. Chen, S.; Zhang, C.; Dong, M.; Lee, J.; Rao, M. Using ranking-CNN for age estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [[CrossRef](#)]
48. Li, K.; Xing, J.; Hu, W.; Maybank, S.J. D2C: Deep cumulatively and comparatively learning for human age estimation. *Pattern Recognit.* **2017**, *66*, 95–105. [[CrossRef](#)]
49. Rodriguez, P.; Cucurull, G.; Gonfaus, J.M.; Roca, F.X.; Gonzalez, J. Age and gender recognition in the wild with deep attention. *Pattern Recognit.* **2017**, *72*, 563–571. [[CrossRef](#)]
50. Duan, M.; Li, K.; Yang, C.; Li, K. A hybrid deep learning CNN-ELM for age and gender classification. *Neurocomputing* **2018**, *275*, 448–461. [[CrossRef](#)]
51. Wan, J.; Tan, Z.; Lei, Z.; Guo, G.; Li, S.Z. Auxiliary demographic information assisted age estimation with cascaded structure. *IEEE Trans. Cybern.* **2018**, *48*, 2531–2541. [[CrossRef](#)] [[PubMed](#)]
52. Yoo, B.; Kwak, Y.; Kim, Y.; Choi, C.; Kim, J. Deep facial age estimation using conditional multitask learning with weak label expansion. *IEEE Signal Process. Lett.* **2018**, *25*, 808–812. [[CrossRef](#)]
53. Rattani, A.; Reddy, N.; Derakhshani, R. Convolutional neural networks for age classification from smart-phone based ocular images. In Proceedings of the IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA, 1–4 October 2018; pp. 756–761.
54. Taheri, S.; Toygar, O. Multi-stage age estimation using two level fusions of handcrafted and learned features on facial images. *IET Biom.* **2018**, *8*, 124–133. [[CrossRef](#)]
55. Angeloni, M.; de Freitas Pereira, R.; Pedrini, H. Age Estimation From Facial Parts Using Compact Multi-Stream Convolutional Neural Networks. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea, 27–28 October 2019; pp. 3039–3045. [[CrossRef](#)]

56. Hosseini, S.; Lee, S.H.; Kwon, H.J.; Koo, H.I.; Cho, N.I. Age and gender classification using wide convolutional neural network and Gabor filter. In Proceedings of the 2018 International Workshop on Advanced Image Technology (IWAIT), Chiang Mai, Thailand, 7–10 January 2018; pp. 1–3. [\[CrossRef\]](#)
57. Savchenko, A.V. Efficient facial representations for age, gender and identity recognition in organizing photo albums using multi-output ConvNet. *PeerJ Comput. Sci.* **2019**, *5*, e197. [\[CrossRef\]](#)
58. Zhang, K.; Liu, N.; Yuan, X.; Guo, X.; Gao, C.; Zhao, Z.; Ma, Z. Fine-grained age estimation in the wild with attention LSTM networks. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 3140–3152. [\[CrossRef\]](#)
59. Agbo-Ajala, O.; Viriri, S. Deeply learned classifiers for age and gender predictions of unfiltered faces. *Sci. World J.* **2020**, *2020*, 1–12. [\[CrossRef\]](#)
60. Liu, X.; Zou, Y.; Kuang, H.; Ma, X. Face Image Age Estimation Based on Data Augmentation and Lightweight Convolutional Neural Network. *Symmetry* **2020**, *12*, 146. [\[CrossRef\]](#)
61. Liu, N.; Zhang, F.; Duan, F. Facial Age Estimation Using a Multi-Task Network Combining Classification and Regression. *IEEE Access* **2020**, *8*, 92441–92451. [\[CrossRef\]](#)
62. Kim, Y.H.; Nam, S.H.; Park, K.R. Enhanced Cycle Generative Adversarial Network for Generating Face Images of Untrained Races and Ages for Age Estimation. *IEEE Access* **2021**, *9*, 6087–6112. [\[CrossRef\]](#)
63. Islam, M.M.; Baek, J.-H. Deep Learning Based Real Age and Gender Estimation from Unconstrained Face Image towards Smart Store Customer Relationship Management. *Appl. Sci.* **2021**, *11*, 4549. [\[CrossRef\]](#)
64. Eidinger, E.; Enbar, R.; Hassner, T. Age and gender estimation of unfiltered faces. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 2170–2179. [\[CrossRef\]](#)
65. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
66. Zagoruyko, S.; Komodakis, N. Wide Residual Networks. *arXiv* **2016**, arXiv:1605.07146.
67. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [\[CrossRef\]](#)
68. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
69. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856.
70. Iandola, F.N.; Moskewicz, M.W.; Ashraf, K.; Han, S.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size. *arXiv* **2016**, arXiv:1602.07360.
71. Kazemi, V.; Sullivan, J. One millisecond face alignment with an ensemble of regression trees. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1867–1874. [\[CrossRef\]](#)
72. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet Classification with Deep Convolutional Neural Networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS), Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
73. Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; Li, F.-F. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
74. Chang, K.Y.; Chen, C.S.; Hung, Y.P. Ordinal Hyperplanes Ranker with Cost Sensitivities for Age Estimation. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011.
75. Chen, K.; Gong, S.; Xiang, T.; Change Loy, C. Cumulative Attribute Space for Age and Crowd Density Estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 2467–2474.
76. Yu, M.; Duan, Y.; Li, Z.; Zhang, Y. Prediction of Peptide Detectability Based on CapsNet and Convolutional Block Attention Module. *Int. J. Mol. Sci.* **2021**, *22*, 12080. [\[CrossRef\]](#) [\[PubMed\]](#)