Republic of Iraq

Ministry of Higher Education and Scientific Research

University of Diyala

College of Science

Gender and Age Classification Based on Face Images by Using Deep Neural Networks

A Thesis
Submitted to the Department of Computer Science\ College of Science\ University of Diyala in a Partial Fulfillment of the Requirements for the Degree of Master in Computer Science

By
Omar Amer Mohammed

Supervised By
Dr. Jamal Mustafa Al-Tuwaijari
Assistant Professor

2022 A.C. **1444 A.H.**

بسم الله الرحمن الرحيم

وَيَسْأَلُونَكَ عَنِ الرُّوح

قُلِ الرُّوحُ مِنْ أَمْرِ رَبِّي

وَمَا أُوتِيتُم مِنَ الْعِلْمِ إِلَّا قَلِيلاً

صدق الله العظيم

سورة الاسراء. اية (58)

Acknowledgment

First of all, praise is to Allah, the lord of the whole creation, on all the blessing was the help in achieving this thesis to its end.

I wish to express my thanks to my college (college of science), my supervisor, Ass- Prof. Dr. Jamal Mustafa Al-Tuwaijari for supervising this research and for the generosity, patience and continuous guidance throughout the work. It has been my good fortune to have the advice and guidance from him. My thanks to the academic and administrative staff at the Department of the computer sciences.

I would like to thank my family who has supported me throughout and encouraging me to further my studies and help to complete this project, and I wish to express my thanks to my friends. Especially the special Farah.

Omar Amer Mohammed

Abstract

Age estimation and human gender determination from visual data (like face images) are still have been fascinating research topics due to the variety of potential used applications. The most important applications include developing intelligent human-machine interfaces and improving safety and security in various domains such as transportation, security, medical, etc. However, the diversity of characteristics like facial plastic surgery, facial hair, wrinkles, skin condition, different human races, and external factors are affected the face appearance which is crucial for gender determination and age estimation. These characteristics make these areas a challenging topic and are regarded as fertile ground for researchers. However, the deep neural networks represent the most important techniques that have recently been utilized to determine human gender and age estimation.

In this thesis, an efficient system for gender determination and human age estimation from face images based on external appearance using deep neural network technique (Multitasking Convolutional Neural Network Algorithm (MCNN)) is proposed. The general construction of the proposed system consists of several stages; Firstly, the stage of image acquisition; Secondly, the stage of pre-processing (using the region of interest technique, image resizing, data augmentation); Thirdly, the feature extraction stage; Finally, the classification stage which produces two results; age estimation, and gender recognition.

The proposed system is implemented using the Internet Movie Database (IMDB) and IMDB-WIKI in combination. The obtained results showed that the proposed system introduces an accuracy of 98.2% for human gender recognition and an accuracy of 98.4% for age estimation. Furthermore, this proposed system achieved better results than the previous work.

List of Contents

List of Figures

List of Tables

List of Algorithms

List of Abbreviations

Abbreviations	Meaning
2D	Tow Dimension
3D	Three Dimension
AI	Artificial Intelligence
CNN	Convolutional Neural Network
CNNs	Convolutional Neural Networks
DA	Data Augmentation
DCNNs	Deep Convolutional Neural Networks
DL	Deep Learning
DMTL	Data Mining Template Library
DRNNs	Deep Recursive Neural Networks
F	Female
FERET	Face Recognition Technology
GANs	Generative Adversarial Networks
HCI	Human-Computer Interaction
IMDB	Internet Movie Database
JPG OR JPEG	Joint Photographic Expert Group
M	Male
MAE	Mean Absolute Error
ML	Machine Learning
MTCNN	Multi-Task Cascaded Convolutional Neural Network
MTL	Multi-Task Learning
PBT	Pixel Brightness Transformations
RELU	Rectified Linear Unit
RGB	Red Green Blue
ROI	Region of Interest
ROR	RSSI-Based Omni-Directional Routing
VGG	Visual Geometry Group

Chapter One

General Introduction

1.1 Introduction

Face system is a computer technology for detecting and recognition of human faces age and gender determination in a digital image. It is utilized in a variety of applications [1]. Face detection and recognition have developed into a very active and significant area of image processing research. The bulk of current face detection algorithms focused on frontal human face identification and face recognition is a well-studied issue in computer vision [2]. It is a challenging vision issue with several practical applications, including identity verification, intelligent visual surveillance, and automated immigration screening systems [3]. According to numerous application scenarios, recognizing faces in real-world applications remains a challenging process [4]

Face recognition age and gender determination has been the most commonly used application of image analysis. The breadth of its commercial law enforcement applications and the availability of cutting-edge methodology have all contributed to its popularity. Additionally, it may be utilized for image retrieval based on content, video coding, video conferencing, crowd monitoring, and intelligent human-computer interactions [5]. The face systems include another important part, which is an estimation and classification, and it falls under the name age and gender and play fundamental roles in social interactions [6]. Despite the essential roles that attribute play in our day-to-day lives, the ability to automatically estimate them accurately and reliably from face images is still far from meeting the needs of commercial applications [7].

Recently, deep learning techniques had a lot of success in face systems. Convolutional Neural Network (CNN) is the most well-known deep learning example in which the training data represent a crucial issue. It is easier to obtain a greater network generalization with more training data. However, using merely faces, appropriately labeled data for age and gender recognition may be limited and difficult to obtain. Overshoot is more likely to occur with insufficient training images. A variety of solutions can be used to alleviate the issue of overfitting. Working in a multi-tasking learning style and increasing the feature of including the age workbook is a modern and high-accuracy method for estimating age and gender recognition [8].

1.2 Related Work

Many publications in the field of human gender and age classification estimation based on deep learning algorithms have been published recently, and this thesis highlights a few of them:

- **K. Zhanc *et al.* (2017)** [9] Proposed a system using residual networks of residual networks (RoR). Furthermore, two simple procedures based on observation of age group features are described to improve the accuracy of age estimation. To increase performance and avoid over fitting, the RoR model is first pre-trained on Image Net, then fine-tuned on the internet movie database (IMDB-WIKI-101) to learn more about the features of face images. Then, it is utilized to fine-tune the audience dataset using two modest mechanisms, pre-training by gender and training with a weighted loss layer to improve age estimation performance. The accuracy of the proposed system in IMDB-WIKI is 66.74% in age and 93.24 % in gender. The proposed system's limitation is that because there is a need to investigate the application of RoR on a large scale and high-resolution picture classifications, this work does not considerably improve the age group and gender classification performance. The percentage of use of the data set was not mentioned.
- **Seok. Lee *et al.*** (2018) [10] Proposed a system that presents a deep residual learning model for estimating age and gender. This method detects faces

from the images entered into it from the dataset. Three deep neural networks are used in the estimate approach, residual learning methods are used, and train the model with images from the IMDB-WIKI database; in this dataset, there are few images of human subjects younger than 20 years old, and so have augmented the set by collecting images from the Internet. According to the results achieve , the suggested system has an age accuracy of around 52.2% and a gender accuracy of approximately 88.5%. Where in the proposed system, a portion of the database was utilized, and the system's training process was augmented by internet images, but images from a global and approved dataset were not utilized.

- **Koichi. Ito *et al.* (2018)** [11] Proposed a system using CNN for predicting age and gender from facial images. The CNN architectures, regression and classification, and STL/ Data Mining Template Library (DMTL) were investigated. When compared against other architectures, WideResNet has the most remarkable performance on age and gender prediction, with age determined by regression and gender predicted by classification. The introduction of DMTL allows CNN to perform better in accuracy and computation time and use the IMDB-WIKI dataset to evaluate the system. The result of the proposed system is that age Mean Absolute Error (MAE) is 8.59, gender accuracy is 93.86%. A robust data cleaning mechanism does not support the limitation of the system, and the limitation of proposed method's accuracy must be improved by adjusting the DMTL methodology.
- **Paolo. Giammatteo Lee *et al.*** (2019) [12] Proposed a system that has two models of CNN (Visual Geometry Group VGG16 type), with a prediction layer change that may be useful for edge computing devices. Both networks are intended to classify a human face according to gender and age. The first (VGG16/10) views gender and age as two intertwined features, with the final neurons designed to retain both qualities together simultaneously. In the prediction layer, the second one (VGG16/8+1) has a neuron for gender

prediction and another eight for age prediction. Such networks were designed to provide information on the gender and age of the individual detected in an image without the need for two separate systems and by using IMDB-WIKI dataset to evaluate the algorithm. The result of the proposed system in age accuracy is 57.0% , and gender accuracy is 88.4%. The system's limitation does not consider other essential edge computing requirements, such as processing time and accuracy.

- **Cao hong. Nga *et al*.** (2020) [13] Proposed a system that uses the images from the IMDB-WIKI dataset presents a transfer learning pipeline to gender and age prediction. To begin with, freeze all layers in Image Net models that have already been trained. The models are then trained for four stages with predetermined learning rates, and the layers' blocks are unlocked in a predetermined order. Apply a multi-output neural network paradigm to simultaneously predict age and gender, with the final loss function depending on the sum of age and gender losses. The result of the proposed system is VGG-19 classification models have the highest gender accuracy (nearly 91%). Age's MAE is 15.23, and the system's limitation is on using a weighted sum of individual losses that can be applied to minimize the effect of age loss on the total loss. And the VGG-19 classification-based models can be trained further as their losses are still on a downward trend.
- **Mohammed benkaddour *et al.*** (2021) [7] Proposed a system that uses CNN to produce a gender prediction and age estimation system for a face image or real-time video. Three CNN network models with varied architectures (number of filters, number of convolution layers) were built and verified on the IMDB and WIKI datasets. The results revealed that CNN networks considerably enhance the system's performance and recognition accuracy. In IMDB, the suggested system has an age accuracy of 86.20 % and a gender accuracy of 94.49 %, whereas in IMDB-WIKI, the age accuracy is 83.97

percentage, and the gender accuracy is 93.56 percentage. The suggested network improves age and gender classification precision significantly.

1.3 Problem Statement

The problems can be summarized in two type (first type is detection, identification) and second type (estimation systems). Detection, identification and estimation systems are biometric systems, and the range of problems that face such systems are demonstrated in the following points:

• **Problem of dimensional image**: Faces are difficult to distinguish from patterns in the visual field. When it comes to identifying a three-dimensional item like a face, though, a two-dimensional picture is the best way.

• **Problem of human composition**: Face is a non-rigid body, where age and gender estimation systems suffer from an important problem, which is the different human races. As a result of this great discrepancy, it is difficult to determine and estimate with high efficiency, especially in ethnic minorities.

• **Problem of angles and backgrounds**: Estimation systems depend on the concept of working on clear and high-quality images in general, as the systems are less efficient in images with non-front angles or simple rotation and backgrounds with overlapping meanings. This problem is due to how the images are captured by the system hardware .

• **Problem of disparity in features**: Problem of disparity in human features in terms of extreme closeness between shapes or changing shapes by simply applying cosmetics and facial plastic surgery.

• **Problem of traditional systems**: traditional systems work on a system of classifying ages into groups in order to avoid the difficulty of estimating in childhood or old age.

1.4 Aim of Thesis

The aim of this thesis is to design and implement a system that can be used to recognize the human gender and age estimation with a high accuracy rate. The system is based on face image and utilizes computer facilities techniques such as deep learning and applies a Multitasking Convolutional Neural Network Algorithm (MCNN)) to achieve a high accuracy percentage of gender recognition and age estimation based on facial appearance. In order to achieve this aim, there are several objectives should be accomplished:

- Solve the problem of simple angles and backgrounds that selects only the image of faces and determining the areas of appreciation in the face, which are the forehead, eyes, nose, mouth and temples, and making a processing process that supports simple angles.
- Solve the problem of human composition or races in estimation systems by training and testing the suggested system on a very large data set with people of various races and ages ranging from one year to 100 years old.
- Solve the problem of contrast in images in terms of extreme closeness, different twins, cosmetic issues and Problem of multiple face by employing a multi-tasking deep learning system that uses the Convolutional Neural Network (CNN) technique to harness the advantages that support age and gender at the same time.
- Solve the Problem of traditional by design a system based on deep learning using the principle of classification instead of (clustering) and training the system to estimate age and gender without giving information about these categories based on facial appearance only.

1.5 Thesis Layout

In addition to chapter one, this thesis consists of four other chapters and are as follows:

Chapter Two: The theoretical underpinning of the general algorithm and approaches employed in this thesis is presented in this chapter.

Chapter Three: This chapter describes the suggested developed system and its associated algorithm in detail and the actions involved.

Chapter Four: This chapter contains the results acquired after using the suggested system on the data set in question and a commentary on the results.

Chapter Five: This chapter summarizes the study's findings and makes recommendations for future research.

Chapter Two

Theoretical Background

2.1 Introduction

Year over year, face analysis tasks have been a popular study area. Age estimation and gender recognition are two of these activities that convey some essential but crucial information about a face. Such data can be helpful in a variety of applications, including autonomous surveillance [14]. Due to its vast applications in many facial analysis challenges, automatic age and gender prediction from face images have recently increased interest [15]. Face analysis tasks have been a popular study area. Age estimation and gender recognition are two of these tasks (low resolution, high light, various human races, very dark-skinned humans (such as African race), and closely comparable morphs) that are all considered problematic for estimating programs to unrestricted age and gender detection tasks straightforward but crucial information about a person's face. Such data can be helpful in a variety of applications, including autonomous surveillance [16].

This chapter presents Challenges facing facial and biometric systems and the most important approaches to deep learning based on the convolutional networks to predict and recognition the gender and estimation age from face images and the ways of processing.

2.2 Face System Types and Challenges

Face system is a subdivision problem of visual pattern recognition. Humans recognize visual patterns all the time, and we obtain visual information through our eyes [17]. The brain recognizes this information as a meaningful concept for a computer, and whether it is an image or a video, it is a matrix of many pixels [18].

The machine should find out what concept a certain part of the data represents in the data; this is a rough classification problem in visual model recognition. For the face system, it is necessary to distinguish who the face belongs to part of the data that all machines think of the face [19]. In a simple concept, it is a biometric technology based on identifying a person's facial features. People collect the face images, and the recognition equipment automatically processes the images. It has three types of operation[20,21]:

- Face Verification (or Authentication).
- Face Identification (or Recognition).
- Face estimated.

Biometric techniques and algorithms are not novel verification approaches. In the old age, in prehistoric times and Babylonian monarchs utilized clay fingerprints to verify legitimacy. For biometric identification, Egyptians employed physical traits such as hand length or half arm[22]. Holistic, feature-based, and hybrid face-recognition algorithms exist; when using holistic face recognition algorithms, they consider the correlations between images and the overall structure of the images [23]**.**

There are many challenges that researchers have faced in the subject of face systems and can be classified in the following point:

- Pose variations.
- Structured elements/occlusions (presence/absence).
- Facial expression changes.
- Aging of the face.
- Varying illumination conditions.
- Modality and image resolution.

These critical factors can be summarized as the head motions, such as egocentric rotation angles or camera point of view changes, resulting in significant changes in facial look (and/or) form and intra-subject face differences [4]. The intra-subject variability in facial images may be caused by a lack of anatomical traits or the presence of occlusions such as beards or mustache caps, or sunglasses [24].

Changes in facial expressions caused by changing emotional states may result in even more variation in facial expressions [25]. Another reason for changes in the human face's appearance could be age [1]. In contrast, large changes in light can have a detrimental effect on the performance of systems. Face identification and recognition become more complex when the backdrop or foreground illumination is weak [26]. Finally, the clarity and resolution of the face image and the setup and mode of the digital hardware used to record the face are other often utilized performance parameters [27].

2.3 Biometric System

A biometric system is a pattern recognition system that collects biometric data from an individual, extracts a feature set from that data, and compares that feature set to a database template set. A biometric system can function in either verification or identification mode [28]. Faces in the biometric system are a non-intrusive method, and facial images are the most prevalent biometric feature humans use to make a personal identification. Facial recognition has a wide range of applications, from static, regulated "mug-shot" verification to dynamic, uncontrolled face identification in a chaotic environment [14,29].

The two most common methods for the face system are [26,30]:

- Placement and shape of facial features, including the eyes, brows, nose, lips, chin, and spatial relationships.
- Overall (global) face image analysis portrays a face as a weighted combination of several canonical faces.

While the verification performance of commercially available face systems is reasonable, they impose several restrictions on how facial images are obtained, including the need for a fixed and simple background or special illumination in some cases [31]. These systems also struggle to recognize a face in images taken from two significantly different perspectives and under varied lighting conditions. It's debatable if a person's face alone, without any contextual information, is a sufficient basis for confidently recognizing a person from a vast number of identities [32]. and the general idea of facial system work can be describe according to steps in below [33] .

- Detect the presence of a face in the captured image.
- Locate the face if there is one.
- Recognize the face from a general viewpoint .

2.3.1 Soft Biometrics

Soft biometric qualities are physical and behavioral characteristics that are not unique to a given subject but are valuable for identifying, verifying, and describing human beings. Examples include gender, height, and weight[34]. Soft biometrics is gaining traction as a viable alternative to traditional biometrics for various reasons, including its non-intrusive authenticating nature. Standalone soft biometrics experimental systems have been built [35].

Measurements refer to the features used to quantify the physical description, including head length, head breadth, middle finger length, left foot length, and cubit length. Body geometry and facial geometry were used to classify these characteristics [36].

The best example of soft biometrics is age estimation and gender, where gender information is a beneficial indication that expert and intelligent systems in healthcare, smart spaces, and biometric-based access control domains can use. To deliver an enhanced user experience, the operations of intelligent technologies in a smart space can be tailored based on gender information [37]. Similarly, a biometric system can improve its performance by using gender as a soft biometric feature. Gender recognition is a problem in which the input can classified into two categories [38]**:**

- Male (M).
- Female (F).

Depending on the input features used in the classifier, the gender recognition systems can be classified into appearance-based and non-appearance-based techniques [39].

Age is the face characteristic points can be specified as the standard reference point on the individual's face used by scientists to recognize an individual's face, estimating the person's age, or in such a situation. Morphology alone is considered to be a format for study.

Variations in the face texture are specified as the face variations associated with muscle and skin flexibility [40]. Thus, studying the shapes of the skull and the face is defined as craniofacial morphology.

Individuals' appearance and structure are impacted in various ways due to the aging operation. The observed changes are associated with face texture and craniofacial morphology. Specific craniofacial morphology features are seen in individuals of a specific age and change over the aging operation. The changes in the texture of the skin typically happen in puberty [41]. Figure (2.1) shows the summary of the below description of the catachresis of the estimation system based on the face.

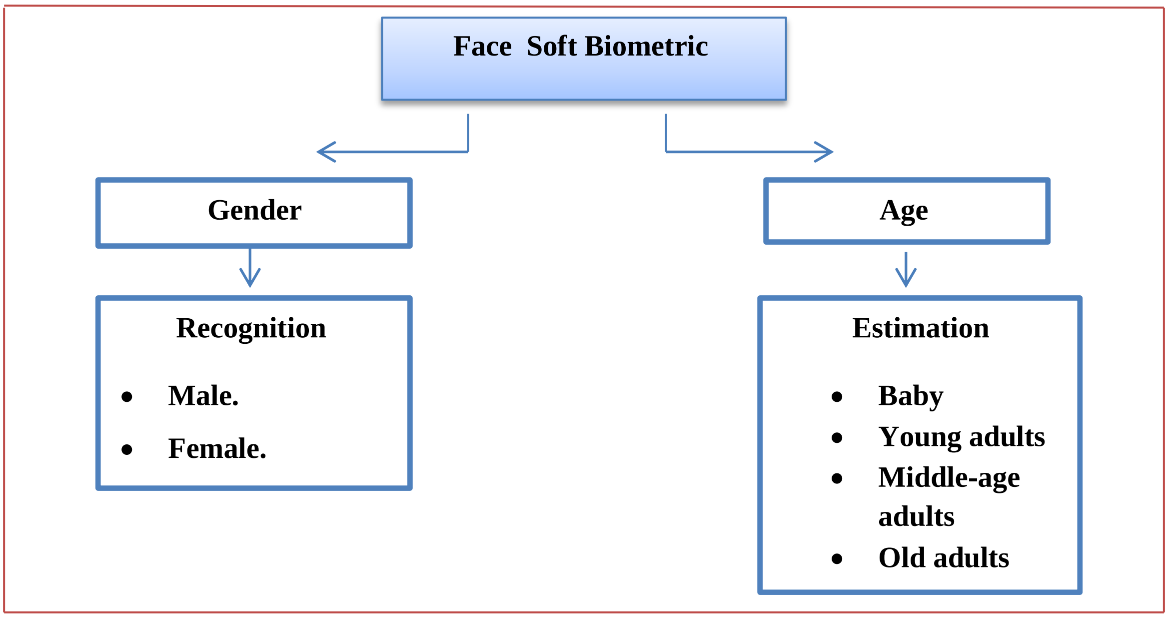


Figure (2.1) : General catachresis of estimation system based on face**.**

2.4 Face System Estimation

Human face system research is significant in various applications, including Human-Computer Interaction (HCI). Face systems have recently attracted a lot of attention. It has a wide range of uses in access control systems and computer vision communications. Because of the benefits of biometric systems, mandatory systems are extremely significant [42]. However, because of the differences in image appearance, such as facial appearances, image orientation, position alterations (non-frontal, frontal), occlusions, and illumination, the issue of face systems remains impressively challenging and scalable [43].

One of the essential tasks in computer vision is recognizing a person's gender and age from a face image. It's a hot study topic with a wide range of applications, including demographic analysis, consumer analysis, visual surveillance, and the progression of aging. It is necessary to recognize a person's gender and age. It's critical to extract characteristics from the captured video images that contain relevant information [44].

2.5 Age and Gender in Face System

Human face characteristics may be categorized into three categories for face recognition. First, the process of facial recognition is based on geometric characteristics such as the jaw, mouth, nose, eyes, and brows. Each face on the planet is unique due to facial traits that vary. This indicates that the geometric representation of facial characteristics (such as relative position and angle) will be recognized as a critical property for identifying faces [41]. A covariate can have an impact on both the intra-class and inter-class variance. Pose, lighting, emotion, and the quality of the image all impact how well a system can recognize faces [45].

The constraints of the face recognition system are controlled variables during picture capture. The constraints imposed by these elements have been thoroughly examined in the literature, and several techniques for overcoming them have been proposed [6].

Face recognition algorithms' performance is also impacted by uncontrollable elements associated with an individual (e.g., race, gender, age group, and aging) [46]. Aging has been thoroughly investigated, and pertinent statistics have been created to assist researchers in tackling the aging problem [47]. Men's faces differ from women's faces in terms of regional characteristics and forms. Males have broader chins than females, whereas females have smoother cheeks (typically, women's).

Noses are smaller than men's. Men and women are also differentiated by their hairstyles and cosmetics. Males, anthropological studies indicate, have a distinct skeletal structure from females. On the other hand, boys and girls have comparable bone characteristics, complicating gender categorization in children's items [45,48].

2.6 Deep Learning Techniques

Deep learning (hierarchical learning or deep machine learning) and learning techniques are part of the field of artificial intelligence (AI) [49], which has played a significant role and represented a technological breakthrough in the world of computers. It can be separated primarily into machine learning and deep learning [50]. Machine Learning (ML) is most usually employed in conjunction with AI, but it is a subset of AI [51]. ML refers to a self-learning AI system based on an algorithm. Systems that develop smarter over time without human interaction are called ML [52]. The second development is deep learning. The application of machine learning to big datasets. The development of machine learning is guided by the principles of precision and speed; the structure of learning approaches is depicted in figure (2.2).

Deep Learning (DL) is distinguished from conventional ML by the use of many layers that provide greater abstraction and robust generalization [53]. The use of neural networks necessitates the extraction of attributes in order to reduce the computational load on networks [54]. If neural networks incorporate qualities, the attributes must be removed. This will necessitate a huge network capable of providing clear instructions. As a result, the likelihood of successful training was minimal or nonexistent. Considering the available processing power and memory, particularly with recent developments in computing power and memory capacity, it is now possible to successfully train huge networks[55].

In the same deep learning network, both feature extraction and classification/generalization can be performed, resulting in significant benefits. Deep neural networks, Deep Convolutional Neural Networks (DCNNs), and Deep Recursive Neural Networks are components of cascading deep learning systems (DRNNs). Deep neural networks are effective examples of deep learning systems. DCNN has gained popularity for image classification, DRNN has been shown to be useful in time-domain signal processing applications [56,75].

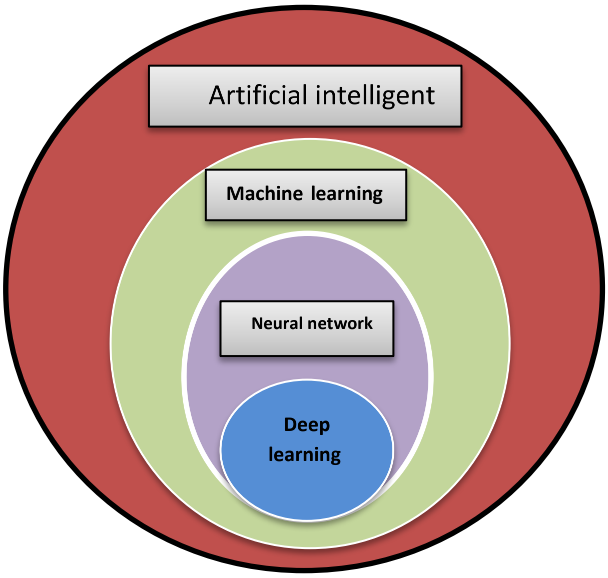


Figure (2.2): Sequence of learning technical [58].

2.7 Augmentation

The validation error must decrease with training to develop usable Deep Learning models. Data Augmentation (DA) is to balance the scale of dataset and is a powerful tool for accomplishing this. The enhanced data will include a wider range of data points, narrowing the gap between the training and validation sets and future testing sets. Over fitting is addressed via data augmentation by starting with the training dataset, which is the cause of the problem [59]. This is done to extract more information from the original dataset through augmentations. To artificially increase the size of the training dataset, these augmentations involve data warping or oversampling. Data warping augmentations are used to change existing photographs or images while keeping their labels. Augments include geometric and color changes, random erasure, adversarial training, and neural style transfer, to name a few. Synthetic instances are created and added to the training set using oversampling augmentations. Mixing images, feature space augmentations, and Generative Adversarial Networks are examples of this (GANs). The two augmentations of oversampling and data warping are not mutually incompatible [60]. It is widely believed that larger datasets result in more robust deep learning models.

The basic idea of data augmentation work to preserve the system from (over fitting) the presence of large data or within the scope of big data by fixing the brightness, size, and dimensions in the data set and working in a random selection method [61]. This technique expands the amount and diversity of training data by applying various transformations to images [62]. Traditional augmentation techniques and generative augmentation techniques are the two primary kinds of image augmentations. Geometric transformations, random cropping, and color space transformations are examples of traditional augmentation techniques (also known as fundamental transformations). The most popular geometric changes are flipping, cropping, rotation, translation, and noise injection [59,63].

Random erasure can make CNN more resistant to many types of image errors. Random erasing is simple to use, requires little parameter learning, and can be coupled with various data augmentations. Color space transformations are used to manipulate an image's color values, and the drawbacks include longer training times and higher memory usage [62,64]. The data augmentations taxonomy is shown in figure (2.3).

Figure (2.3): Data augmentations taxonomy [60]**.**

The advantage of Face data augmentation is fundamentally crucial for improving the performance of neural networks in the following aspects [62,65]:

- ✓ It is inexpensive to generate a massive number of synthetic data with annotations compared to collecting and labeling real data.
- ✓ Synthetic data can be accurate, so it has ground truth by nature.
- ✓ If a controllable generation method is adopted, faces with specific features and attributes can be obtained.
- ✓ Face data augmentation has notable advantages, such as generating faces without self-occlusion and a balanced dataset with more intra-class variations.

It is impossible to train a deep and robust CNN due to the restricted training dataset and the fact that each person only possesses a few types of images. Data augmentation is a viable approach to increasing the size of the training dataset. Data augmentation approaches aid the generalization ability of a trained CNN model to previously undetected but similar noise patterns in the training data [61].

2.7.1 Type of Data Augmentation

The methods of increasing the data of facial images can be divided into many types; that is, they support several cases, from visualization to human manipulation, and can be divided into several types, but the type that used and that play an essential factor in facial systems is the geometric transformations. Geometric transformations describe different augmentations based on geometric transformations and many other image processing functions. For example, rotations and flips are generally safe on Image Net challenges such as cat versus dog. This demonstrates the data-specific design of augmentations and the challenge of developing generalizable augmentation policies. The geometric augmentations is done as; Position augmentation (Cropping, Flipping, Rotation, and Translation), and Color augmentation (Brightness, Contrast, and Saturation) [60].

2.8 Methods of Feature Extraction and Classification

Deep learning is a significant area of research in image and video processing, computer vision, etc. As shown in the next paragraph, it has many algorithms that help in feature extraction and classification while offering satisfactory results for researchers. The most important algorithms that have been extensively employed by researchers in the field of face systems, have promising outcomes [57].

2.8.1 Convolutional Neural Networks (CNN)

Convolutional neural networks have recently exhibited exceptional performance in various AI challenges. While CNN excelled at many high-level computer vision tasks, including classification, object recognition, scene understanding, and others, it also created low-level imaging issues like filtering and picture fusion. CNN is a novel artificial neural network model that combines artificial neural networks and deep network learning [60]. The general structure of the convolutional neural networks system is viewed in figure (2.4)

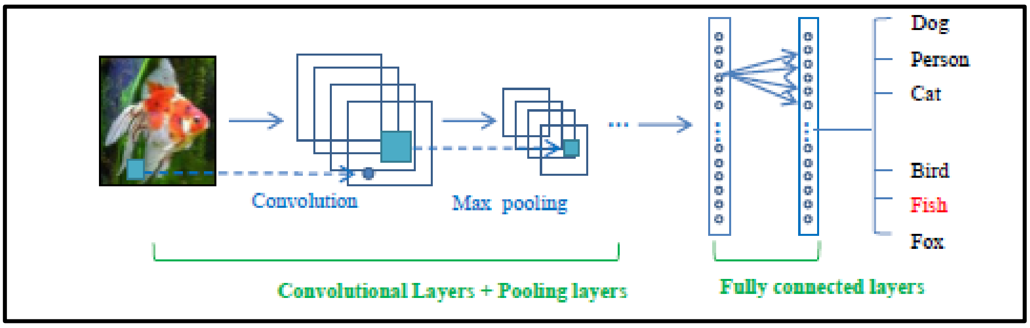


Figure (2.4): The general structure of the CNN system [66].

The deep CNN architecture is usually formed by a pile of discrete layers that convert the image input to the separation scores. Four distinct types of layers widely use torsion, collector, fully connected, and classification layers. Typically, multiple pairs of wrapping and stacking layers are repeated, followed by integral and classification layers [67].

Figure (2.5) demonstrates bundling and wrapping layers. The wrapping layer is the most critical CNN component. Each torsion layer neuron has a small receiving field in the input image and computes the output by converting the receptive field with a linear filter. The algorithm can be broken down into the four phases below.

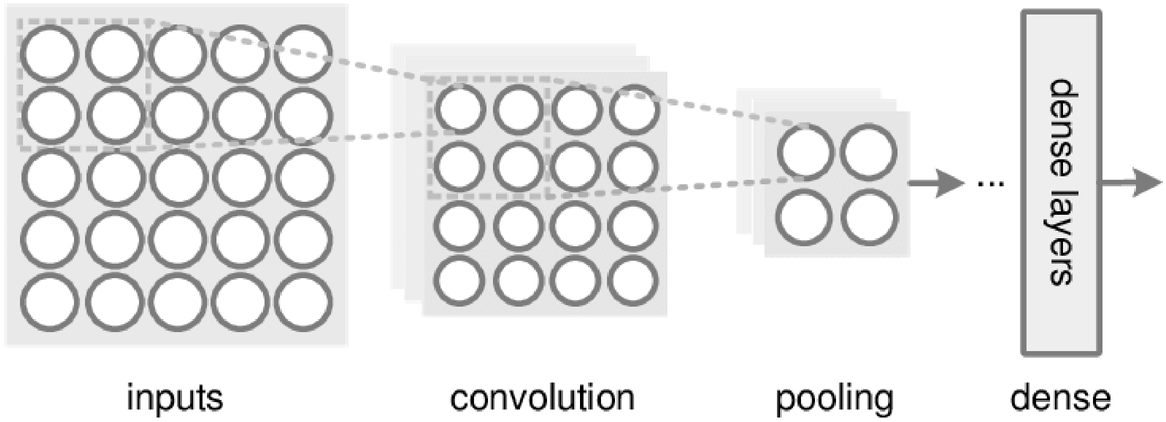


Figure (2.5): Pooling layers and the convolution [68].

Convolutional Neural Networks (CNNs) are specifically built to handle data containing several arrays/matrices, such as an image made of three RGB matrixes. In addition to establishing a neural network, a loss function is required to evaluate the model's performance [69]. As a result, training a CNN model is changed into an optimization problem to minimize the loss function's value over the training data. A gradient-descent-based approach is commonly used to tune the parameters in a CNN iteratively [70].

2.9 Basic Structure of CNN

Convolution neural network is identical to artificial neural network, as both are composed of self-optimized neurons, introduced by inputs and executing non-linear transformation. Convolution neural network is commonly utilized in pattern recognition on objects relative to artificial neural networks, because it encodes image-specific features in system architecture. In the convolution neural network, there are five fundamental elements: the input layer, convolution layer, non-linear layer, pooling layer and fully connected layer [66].

1. Input Layer : The input layer contains the pixel values of the image that enter CNN [66].

2. Convolutional Layers: The CNN convolves the entire image in this layer, including intermediate feature maps, and generates different feature maps using various filters. and the results of the convolution procedure are :

- ✓ Reduce the number of parameters using weight-sharing mechanisms.
- ✓ Correlation between neighboring pixels is easy due to local connectivity.
- ✓ The location of the object is fixed.

As a result of these benefits, researchers have begun to substitute completely connected layers to speed up the learning process [69].

Figure (2.6) shown a, b, c convolution example process with a kernel measurement of [3×3], with no padding and one stride, which the stride means the step scale to vertically and horizontally traverse. The kernel is extended around the insert tensor, and the component-wise function between each object of the tensor input and kernel is measured at every point and averaged up to produce the output amount in the related direction of the tensor output called the feature map [71].

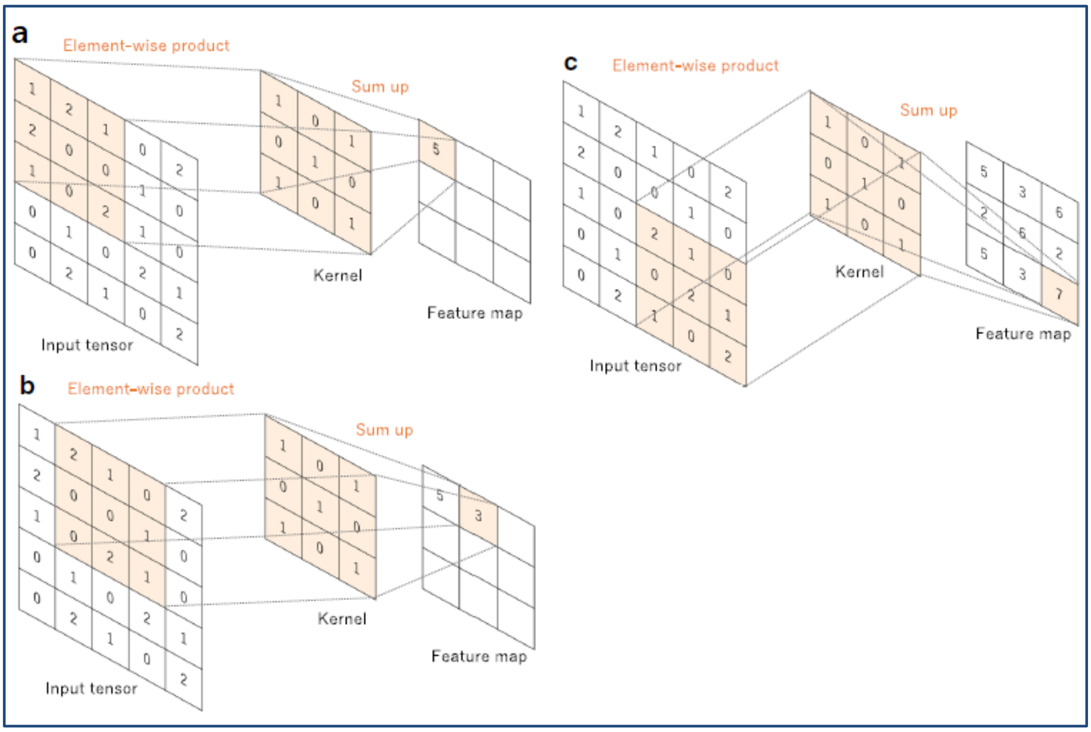


Figure (2.6): An example of convolution operation [71].

3. Pooling layers: This layer is identical to the convolution layer, but it use to control over fitting and reduces feature map measurements and network parameters, in most cases, average and maximum pooling is used. When it comes to average and maximum pooling. The most widely used approaches are average pooling and max pooling. The figure (2.7) describes the processes of both types of pooling.

- ✓ Max pooling (The maximum pixel value of the batch is selected).
- ✓ GAP pooling (Global average pooling the average value of all the pixels in the batch is selected) [68].

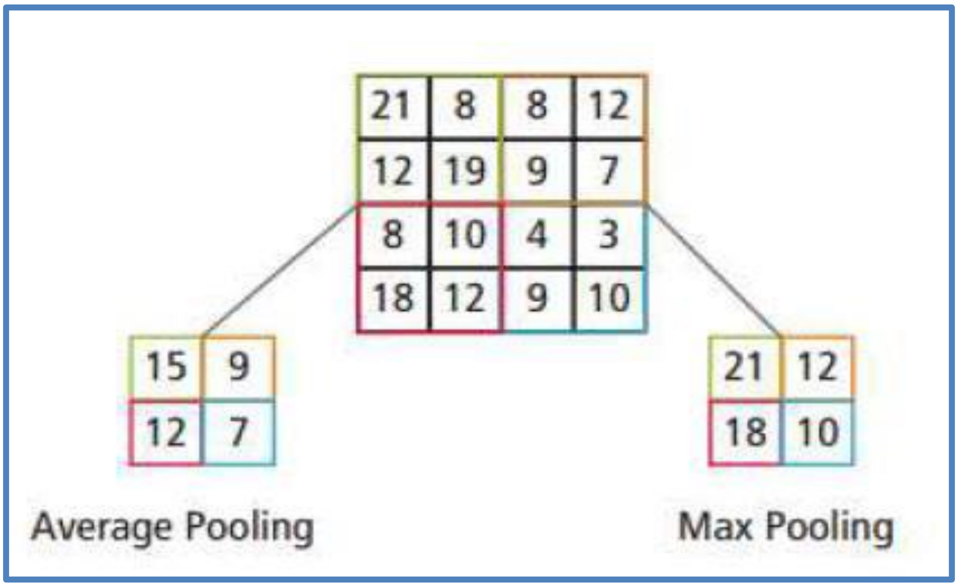


Figure (2.7): Two Classic Pooling Methods [72].

4. Non-linear Layer (Activation Function): A non-linear transformation (or called activation function) also is applied to the input by the convolution neural network, the object of which is to classify the features within per hidden layer. The non-linear translation function in the artificial neural network is sigmoid or hyperbolic tangent. Though, if the sparsity of the data is greater, the outcome will be better for image processing. This layer uses many types of activation functions the most popular is Rectified Linear Units (ReLU).

Rectified linear units are commonly used as non-linear transformation, for rectified linear unit, it applies a following equation.

$$Y = \max (X.0) \quad \ldots\ldots\ldots\ldots\ldots\ldots \mathbf{(2.1)}$$

So that the output is in the same size as the input. Rectified linear unit raises the decision function's non-linear properties and has no negative impact on the convolution layer's sensitive fields. The training rate of the rectified linear model is much higher compared to other non-linear units. The figure (2.8) an example of the Rectified Linear Units (ReLU) [72].

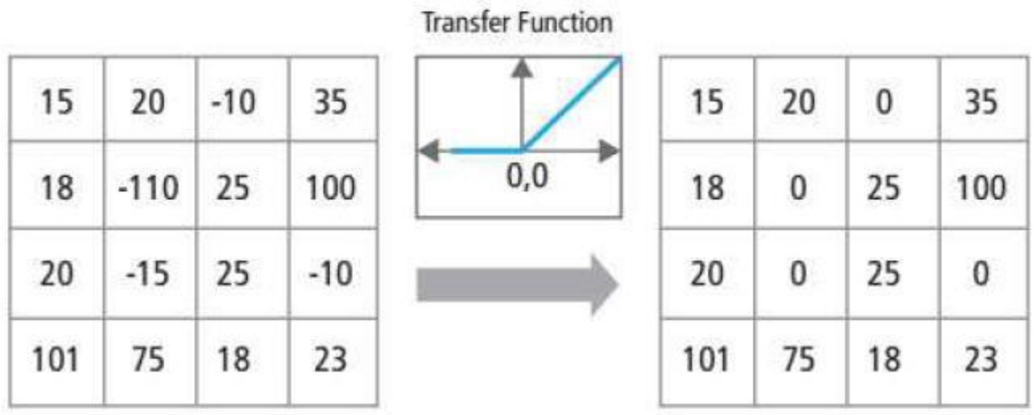


Figure (2.8): An Example of ReLU Transformation [72].

5. Fully Connected Layer: 90% of the parameters are contained in the last layer of CNN. The feed-forward network creates a vector of a specific length to follow up processing because most of the parameters are contained in these tiers. For example, if there is a neural convolution network for gender classification and the result matrix is a probability of [0.8, 0.2], this means that there is 80 % probability of male gender and 20 % probability of female gender [73]**,** [74].

6. Softmax Layer: is a mathematical function that transforms a vector of numbers into a vector of probabilities, with the probabilities of each value proportional to the vector's relative scale. It depends on the following equation [75], [76]:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad \text{................} \quad \textbf{(2.2)}$$

Where, **σ** indicates softmax, **Z** indicates input vector, $\mathbf{e^{z\ i}}$ indicates standard exponential function for input vector, **K** indicates number of classes in the multi-class classifier, and $\mathbf{e^{z\ j}}$ indicates standard exponential function for output vector.

7. Normalization Layer: The input layer is normalized by modifying and scaling the activations, for e.g., if it has features from 0 to 1 and also from 1 to 1000, also must normalize those features to accelerate learning. If it supports the input layer, why not do the alike for the values in the hidden layers that shift all the time and increase the training pace by 10 times or more. There are various types of normalization, the most famous of which is the batch normalization [77]. Batch normalization represents a mechanism for enhancing the speed, execution, and stability of neural networks. In a neural network, batch normalization is achieved through a normalization step that fixes the means and variances of each layer's inputs, to compute the mini batch mean (μB) by the equation (2.3), m is the values of this activation in the mini-batch [77]:

$$\mu B = \frac{1}{m}\sum_{i=1}^{m} x_i \quad \text{....................}\textbf{(2.3)}$$

While computing the mini batch variance (σ^2 B) by the equation:

$$\sigma^2 B = \frac{1}{m}\sum_{i=1}^{m}(xi - \mu B)^2 \quad \text{..............}\textbf{(2.4)}$$

At the last, normalize the layer inputs by using the prior calculated batch statistics as in the equation[77]:

$$\overline{x_i} = \frac{xi-\mu B}{\sqrt{\sigma^2 B+\epsilon}} \quad \text{..................}(2.5)$$

8. Dropout Layer: Dropout layer is often employed to combat the imbalance and over fitting issue . Dropout layer typically is placed in between the fully connected layers. It dynamically switches on and off the edges connecting the two layers and helps reduce coadaptation phenomenon known to occur where nodes and weights are trained to similar configuration when the dropout layer is not in use. It is a utility layer effective during the training phase, and during the test and deployment phases, the dropout layer is removed [78].

2.10 Cross –Validation

Process of assessing how the results of a statistical analysis will generalize to an independent dataset. In a prediction problem, a model is usually given a dataset of known data on which training is run (training dataset), and a dataset of unknown data (or first seen data) against which the model is tested (called the validation dataset or testing set). The goal of cross-validation is to define a dataset to “test” the model in the training phase (i.e., the validation set), in order to limit problems like over fitting, give an insight on how the model will generalize to an independent dataset [79].

Two types of cross-validation can be distinguished, exhaustive and non-exhaustive cross-validation. Exhaustive cross-validation methods are cross-validation methods which learn and test on all possible ways to divide the original sample into a training and a validation set such as Leave-p-out cross-validation and Leave-one-out cross-validation. Non-exhaustive cross-validation methods do not compute all ways of splitting the original sample. Methods of

Non-exhaustive cross-validation include Holdout method and k-fold cross-validation[80].

2.10.1 Holdout

This is the "simplest type of cross-validation" in the traditional sense. This technique is frequently categorized as a type of "simple validation" as opposed to a "simple or degenerate form of cross-validation." In this procedure, the data is randomly divided into two groups: Training and Test/Validation set, also known as a hold-out set. The model is then trained using the training dataset and evaluated using the test/validation dataset. The model assessment approaches used to compute the error on the validation dataset depend on the type of problem we are attempting to solve, with Mean Squared Error (MSE) being used for regression problems and other metrics providing the misclassification rate to assist locate the error for classification issues. The training dataset is typically larger than the hold-out dataset. Typical data-splitting ratios include 60:40, 80:20, and so on. This method is only used when evaluating a single model with no hyper parameters to tweak [81]. Figure (2.9) shows the technique of Hold Out.

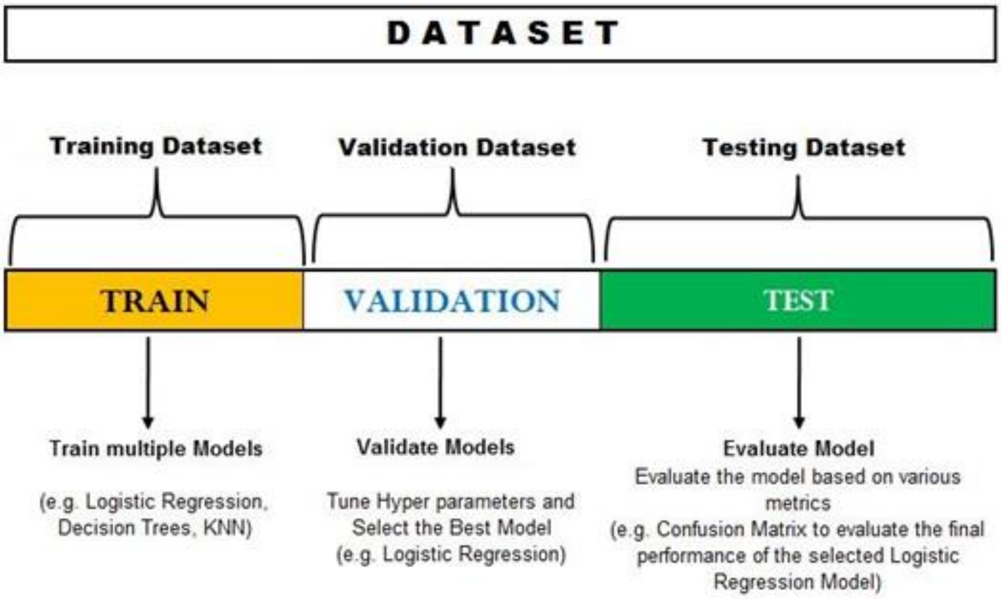


Figure (2.9): Hold Out Technical [82].

The main steps of the hold-out process for model selection are as follows [79]:

1. Split the dataset in three parts – Training dataset, validation dataset and test dataset.
2. Train different models using different learning algorithms.
3. For the models trained with different algorithms, tune the hyper-parameters and come up with different models. For each of the algorithms mentioned in step 2, change hyper parameters settings and come with multiple models.
4. Test the performance of each of these models (belonging to each of the algorithms) on the validation dataset.
5. Select the most optimal model out of the models tested on the validation dataset. The most optimal model will have the most optimal hyper parameters settings for a specific algorithm.
6. Test the performance of the most optimal model on the test dataset.

2.10.2 k-Fold

K-Fold CV is where a given data set is split into a ***K*** number of sections/folds where each fold is used as a testing set at some point. Popular example is scenario of 5-Fold cross validation(K=5). the data set is split into 5 folds. In the first iteration, the first fold is used to test the model and the rest are used to train the model. In the second iteration, 2nd fold is used as the testing set while the rest serve as the training set. This process is repeated until each fold of the 5 folds have been used as the testing set [83]. Figure (2.10) shows how the dataset in k-fold is split, and Figure (2.11) shows the different between K-fold and Hold-out.

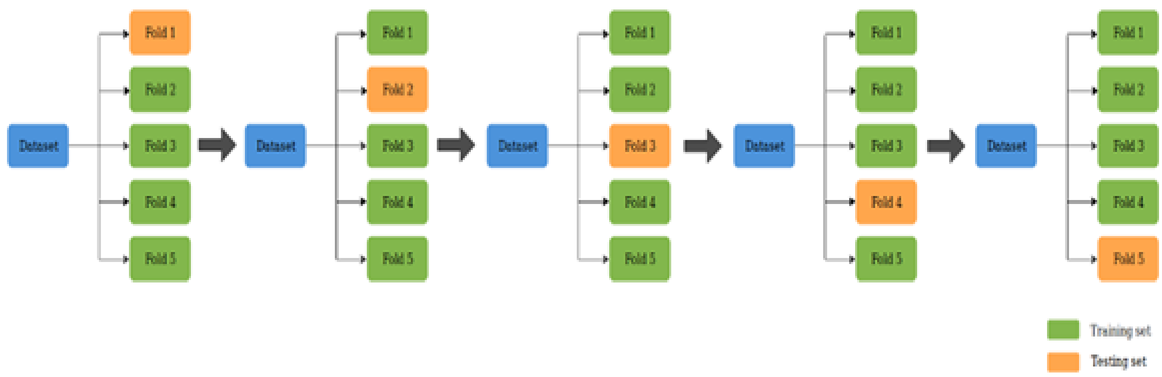


Figure (2.10): Dataset splitting into k-fold [82]**.**

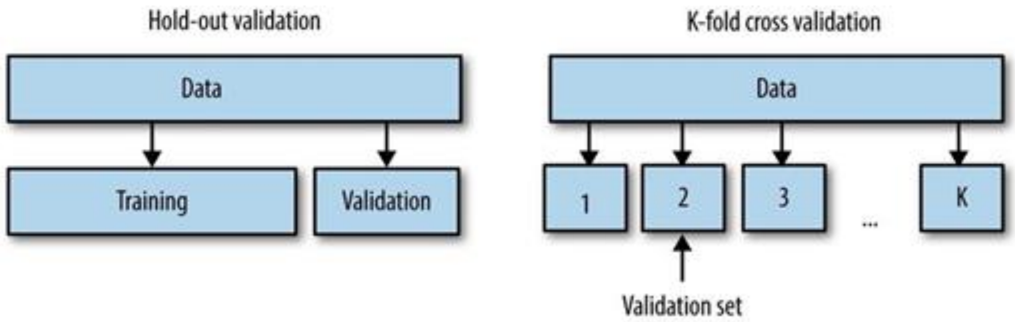


Figure (2.11): Different between k-fold and hold out [82]**.**

2.11 Model Architecture Efficient Net

Convolutional Neural Networks (ConvNets) are commonly developed at a fixed resource budget, and then scaled up for better accuracy if more resources are available. Efficient Net CNN is example models exceed all previous models in terms of both the number of parameters and Top-1 accuracy when applied to the versa dataset. The Efficient Net family is based on a novel scaling technique for CNN models. It uses a straightforward and highly effective compound coefficient. Efficient Net scales each dimension of a network uniformly with a given set of scaling coefficients, as opposed to existing approaches that scale network dimensions such as width, depth, and resolution. In practice, scaling individual dimensions enhances model performance, but balancing all dimensions of the network with regard to the available resources improves the model's overall performance. Efficient Net CNN models and in their original paper they proposed seven such models which they named EfficientNet-B0 to EfficientNet-B7. Efficient Net from B0 to B7, each one of these contains parameters ranging from 5.3M to 66M [84,85]. Figure (2.12) shows the architecture of EfficientNet-B3 CNN.

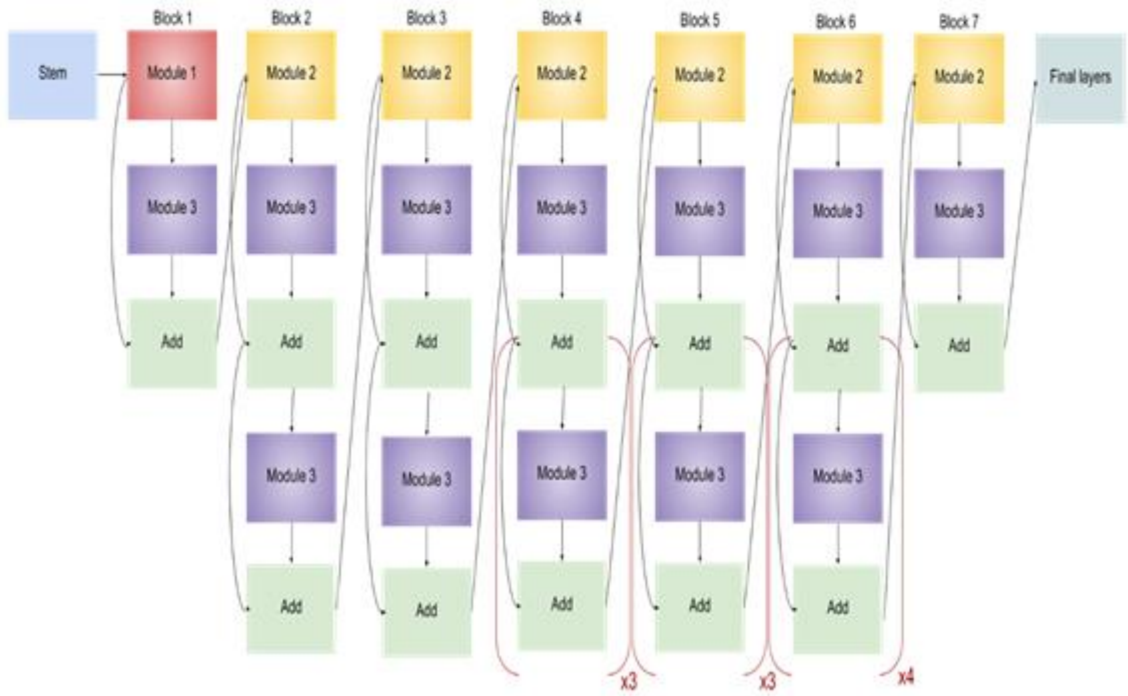


Figure (2.12): Architecture for EfficientNet-B3 [85]**.**

Efficient Net model, which is among the state-of-the-art models by reaching 84.4% accuracy with 66 M parameter in the image Net classification problem, can be considered as a group of CNN models. Efficient Net group consists of 8 models between B0 and B7, and as the model number grows, the number of calculated parameters does not increase much, while accuracy increases noticeably. Unlike other CNN models [86]. The aim of deep learning architectures is to reveal more efficient approaches with smaller models. Efficient Net, unlike other state-of-theart models, achieves more efficient results by uniformly scaling depth, width, and resolution while scaling down the model. The first step in the compound scaling method is to search for a grid to find the relationship between the different scaling dimensions of the baseline network under a fixed resource constraint. In this way, a suitable scaling factor for depth, width and resolution dimensions is determined. These coefficients are then applied to scale the baseline network to the desired target network [87].

The main building block for Efficient Net is the inverted bottleneck MBConv, which was first introduced in MobileNetV2.but due to the increased FLOPS (floating point operations per second) budget, it is used slightly more than MobileNetV2. In MBConv, blocks consist of a layer that first expands and then compresses the channels, so direct connections are used between bottlenecks that connect much fewer channels than expansion layers. This architecture has in-depth separable convolutions that reduce calculation by almost k2 factor compared to traditional layers where k is the kernel size which denotes the width and height of the 2D convolution window [88]. The schematic representation of Efficient Net model is shown in Figure (2.13).

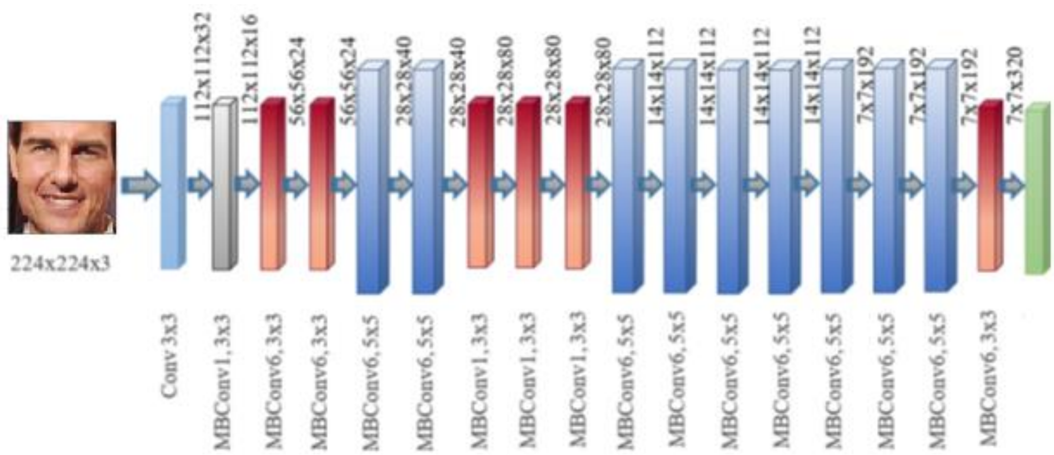


Figure (2.13) Schematic representation of Efficient Net [86].

2.12 Confusion Matrix

A confusion matrix is a table that summarizes the results of classification problem prediction. The total number of correct and incorrect predictions is added together and then divided by class using count values [89].

The confusion matrix's key is one of the reasons that a classification model gets confused while making predictions. It reveals the sorts of errors made by the classifier and the number of errors made by the classifier. And the following steps have summarized the way of confusion matrix works[90]:

Step 1: Create a test dataset or a validation dataset with expected results.

Step 2: For each row in the test dataset, make a prediction.

Step 3: Count the number of expected outcomes and predictions:

- For each class, the number of valid guesses.
- The total number of inaccurate guesses for each class is sorted by predicted class.

A classification model's correctly or incorrectly predicted number of occurrences is summarized in a confusion matrix. The following figure (2.14) shows the subdivisions of matrix:

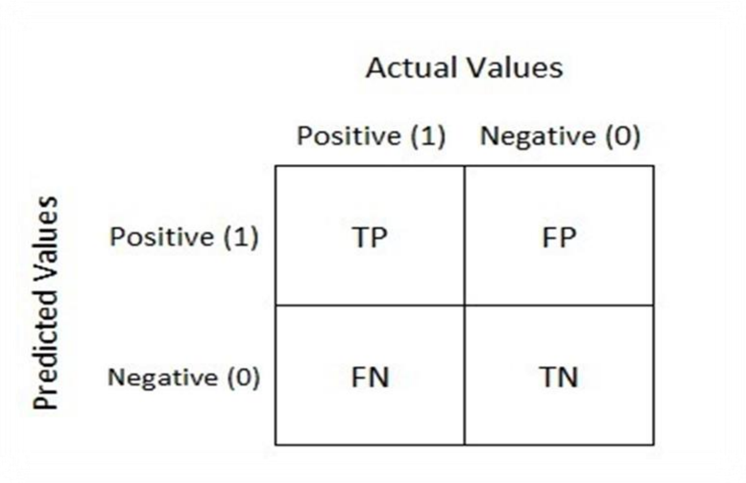


Figure (2.14): Confusion Matrix [91].

Where TP indicates the True positive which is the number of right expectations that an occurrence is certain, TN indicates the True negative which is the number of right expectations that an occurrence is negative, FP indicates the False positive which is the number of inaccurate expectations that a case is negative, and FN indicates the False negative which is the number of inaccurate expectations that a case is certain. Furthermore, P indicates the Number of positives (1 or +), and N indicates the Number of negatives (0 or -).

Depending on the four divisions, there are essential measures for estimating the efficiency of systems through the confusion matrix, which is as follows [92].

- **Precision**: In contrast to bias, which is calculated by comparing the collection's standard deviation to the object's known value, variance is calculated by calculating the sample's standard deviation from the sample's mean. It's the proportion of records in a group that the classifier claims to be a positive class that is true, as shown in the precision equation (2.6).

$$\text{Precision}=\frac{TP}{TP+FP} \quad \ldots\ldots\ldots \ \mathbf{(2.6)}$$

- **Recall**: As its name implies, recall is a measure of how well a classification system is able to forecast the number of positive cases. Equation demonstrates the power of recall (2.7).

$$\text{Recall}(r) = \frac{TP}{TP + FN} \quad \ldots\ldots \ \mathbf{(2.7)}$$

- **F1-Score**: The harmonic mean between precision and recall, as well to the equation, are denoted by F1 (2.8) is

$$\text{F1}= \frac{2rp}{r + p} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad \ldots. \ \mathbf{(2.8)}$$

- **Accuracy**: The following equation can be used to calculate accuracy [93].

$$\textbf{Accuracy}= \frac{\mathbf{TP+TN}}{TP+TN+FP+FN} *100 \quad \ldots\ldots\ldots\ldots\ldots\ldots\mathbf{(2.9)}$$

Chapter Three

Methodology

3.1 Introduction

This chapter includes design considerations for the proposed system of human age estimation and gender recognition from face image and a detailed explanation of each step in the diagram using Deep Learning (DL) technology by building deep network model called multiple tasks Convolutional Neural Network (multiple CNN_based EfficientNet architecture). Additionally, the proposed system trained on a large database of different faces to detect a person's approximate age and gender recognition-based on appearance.

3.2 The Proposed System

The entry of the suggested system is the face image of the human that the system need to estimation their age and recognition gender based on the face features. Figure (3.1) illustrates the human age estimation system architecture and gender recognition which is suggested in this thesis based on global dataset, namely IMDB and IMDB-WIKI.

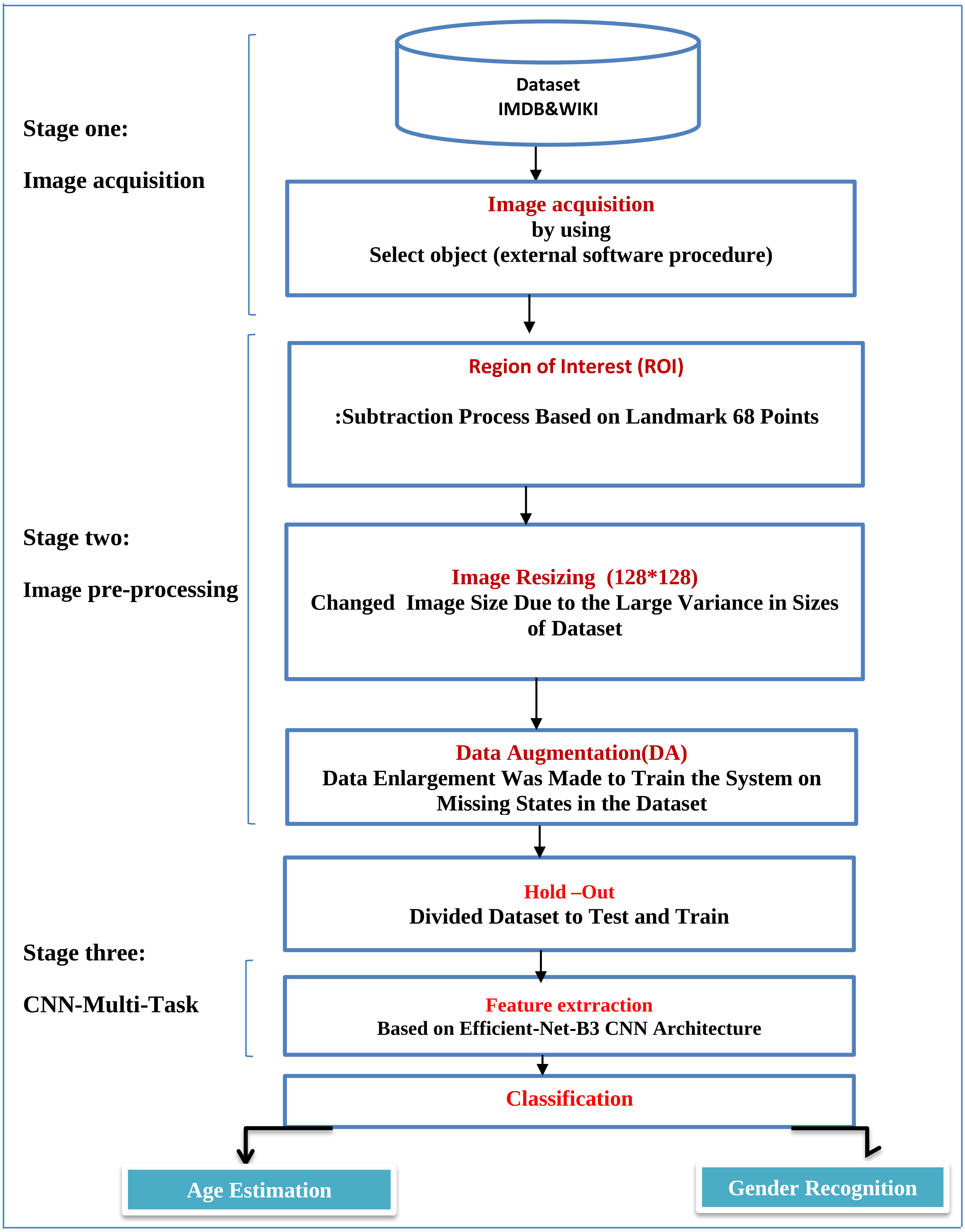


Figure (3.1): System Architecture for Age Estimation and Gender Recognition.

3.2.1 Image Acquisition in the Proposed System

Selection of images is one of the most important stages that trim the dataset and remove unwanted objects or images. In the proposed system, this step was done through an external or separate software procedure, where it worked on selecting images with faces only and images ranging between (1-100) and the following is the algorithm(3.1) which explains the mechanism of action:

Algorithm (3.1): Image selection
Input: Full_ path image
Output: Metadata (age, gender, image paths)

Begin
Step One: For the IMDB dataset (read image file names)
Step Two : Calculating age by image name
1- Month of age < 7 then continue
Age=Age-month
Else
Age=Age+1
Step Three: Ensure that there is a gender category for the person's image in the dataset
1- If it is available, we will continue
Else
Discard
Step Four: Ensure that the age is limited to between 1-100
1- If it is available, it will be accepted
Else
Discard
Step Five : Examination face-score of each person's
If face-score <1 and face-score>0

Accepted

Else

Discard

Step Six: If all the previous conditions are met, this information is stored and added in a row in the metadata file (CSV) that includes age, gender, and image path.

Step Seven: Repeat all previous steps for all dataset images.

Step Eight: We repeat the same conditions and previous steps on the wiki dataset.

End

This algorithm can be summarized through the following points:

- Upload image
- Use the label of dataset facescore
- Determine ages greater than zero by giving an initial value of (Min_score =1). If the value of (face_score) is higher than the default value(Min_score), it works to choose that; else, passes the image and chooses another.
- Determine ages less than or equal to 100 by giving an initial value of (Max_score =100). If the value of (face_score) is less or equal to the default value(Max_score), the works to select it; else, pass the image and select another.
- The result is images of faces ranging in age between (1-100) in an excel file format with the extension comma-separated values (CSV) and stored in a folder called (Meta), which contains inside it all the trimmed images.

3.2.2 Region of Interest (ROI)

The process of extracting an important area within the group, extracting an area within the face based on the use of (landmark) techniques, which consisted of (68) areas (or points) that were identified based on practical experience, where the benefit of this process or technology to limit the so-called deduction area called ROI within the center of the face.

This process is done using the library (dlib) that is and a Python package appropriately named (face_recognition) wraps dlib's face recognition functions into a simple. One of the most widely used face recognition libraries, where this library is the basis of the ROI process, as it works to determine the center of the face depending on the (68) points specified in the proposed system and they are fixed positions in the face of the human being as shown in figure (3.2) to illustrate positions this points around face human . Relying on (landmarks) that localize and represent the prominent areas of the face (eyes, eyebrows, nose, mouth, lip, Jawline, and around the face) in the context of facial features, our goal is to discover critical facial structures on the face using shape prediction methods.

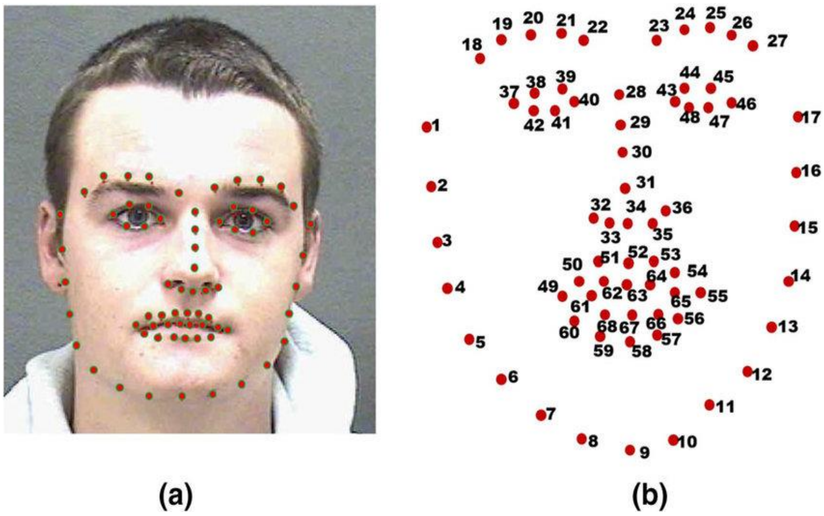


Figure (3.2) Identification of facial landmarks using Dlib. (a) Facial landmarks. (b) The position and order of 68 points on the face[94].

Where this library in nature contains specific coordinates (x, y) for the areas surrounding each face structure, but in the proposed system, a third coordinate was adopted, which is (margin) by 0.4, to address the problem of the simple rotation case in the face, which works to expand the ROI area The facial feature detector pre-trained within the dlib library is used to estimate the position of the 68(x,y)(margin)-formatted that map of the facial structures on the face. Figure (3.3) shows the structure of ROI process.

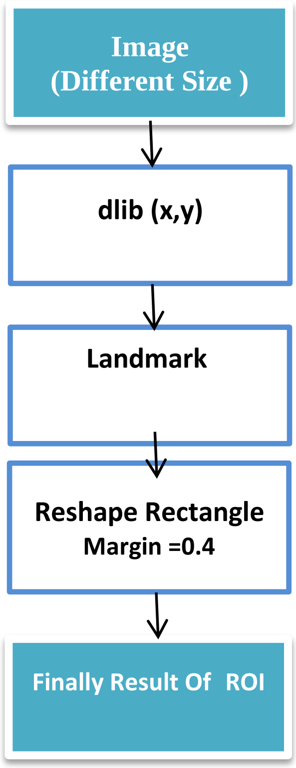


Figure (3.3): The structure of ROI process.

The algorithm of ROI that summarizes all the steps of the process is mentioned in Algorithm (3.2).

Algorithm (3.2): Region of interest (ROI)	
Input: Image	
Output: Crapped the region of interest from the image	
Begin	
Step1:	Load data (image)
Step2:	dlib's Face Detector - Facial Landmark
Step3:	ROI using dlib landmarks
Step4:	Facial points
Step5:	Image display
End	

3.2.3 Resizing the Image

The process of resizing the image is essential because it works to unify the image's dimensions. In the proposed system, there was a problem, which is that the dataset is huge, and most importantly, it varies greatly in size, as its sizes range (the smallest is 71 * 70) and (the largest is 500 * 500) and the image size cannot be changed. Before the ROI process, to implement the process correctly, and as a result, the ROI output will be images of very different sizes. Accordingly, the image size was changed to 128 * 128, and this number was approved based on the experiment on different sizes and their impact on accuracy.

3.2.4 Augmentation

The technique of artificially expanding the dataset for the software model to receive many new states and new image modes. In the proposed system, this method was adopted to train the model on many cases that may not be available in the dataset; it is only training or tesing the model, but it is done. As explained in the second chapter in section (2.7.1). Its exception is in the epoch stage, and it is not included in the evidence set. Any idea of expansion is a dummy training for many cases that do not exist, only to increase the system's efficiency in estimating age and gender in different forms; that is, the artificial training adopted it in the proposed system. Figure(3.4) shows the types of values given based on the experiment. In the proposed system, the following cases were adopted for expansion:

- Rotation
- Contrast
- Brightness(light)
- Flip

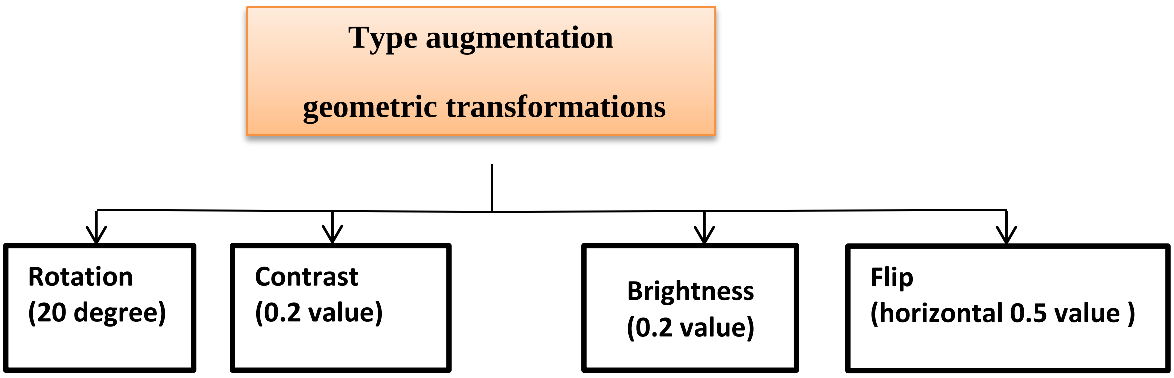


Figure (3.4): Type and value of augmentation**.**

The following is the algorithm (3.3) represents data augmentation in the proposed system:

Algorithm (3.3): Data Augmentation
Input: Image
Output: Image Augment
Begin Step1:Transforms : A.Shift Scale Rotate (Shift_ limit=0.03125, Scale_ limit=0.20, Rotate-limit=20) Step2:Transforms : A.Random Brightness Contrast (Brightness _limit =0.2, Contrast_limit=0.2, A.Horizontal Flip(p=0.5) **End**

3.2.5 Holdout Technique

The holdout technique is an exhaustive cross-validation method, that randomly splits the dataset into two types, test and train data depending on the analysis of the data. For the holdout cross-validation case, the dataset is divided randomly into validation data and training data. Typically, the split of test data is less than training data. The training data is normaly used to train the model, while the validation data is used to evaluate the model's performance. As explained in the second chapter in section (2.10.1) In the proposed system, this type of technique is used in deep learning before the stage of classification by using the Efficent-Net CNN algorithm .

3.2.5.1 Training Set

The data is divided into two main sections, which are training data and testing data. The data is dividing using a function in the python called "sklearn", this function divides the data according to the percentages determined by the user and randomly. It was randomly divided and not chained to obtain a better approach and data is randomly taken from the dataset to make detection later better and stronger. In the proposed approach the data was divided so that the training data would be 80% and the testing data is 20%, after testing all ratios, and this best ratio as it will be shown in the next chapter.

3.2.6 Feature Extraction and Classification

Feature extraction and classification representes the main stages of the age and gender estimation system. CNN based age and gender detection is a popular method for identifying a person utilizing deep analytics patterns and the intrinsic qualities of the human face. The multitasking CNN or efficiency CNN is a deep learning method used as part of the process in this proposed system to make the implementation more efficient because the estimation system requires high classification accuracy.

In the proposed system, the inputs to the CNN algorithm are the results of the re-scaling of the images resulting from the data (Augmentation) operation. A multitasking CNN was designed and trained on the dataset (IMDB& IMDB_WIKI) using different layers which are ((Conv2D)_(top)), ((BatchNormalization)_(top)), (avg_pool) and finally adding two layers (pred_gender Dense) and (pred_age Dense) to get two different outputs.

3.2.6.1 Optimum CNN Parameters

Efficient-Net is a neural network based on the principle that everything is gradually increased. The first convolutional neural network is the (stem), after which all the experimenting with the architecture starts. Each of them contains seven blocks. These blocks further have varying sub-blocks whose number is increased as move from block to other.

Note that two layers have been added to the basic structure, one dedicated to age and the other to gender. The purpose of these two layers is to make an algorithm work in a multitasking manner and give two different results. The condition of the age layer is to display and estimate ages ranging from (1 to 100), and the second layer is its condition. Gender of the person in the image by binary (female(0) or male (1)). A convolutional neural network consists of an input layer, hidden layers, and an output layer. Any middle layers are hidden in any feed-forward neural network because the activation function and final convolution mask their inputs and outputs. It is commonly a feed-forward structure, where each layer receives the output of a preceding layer as its input and produces the results for the next layer. Table (3.1) represents a summary of the neural network algorithm layers, illustrating the way it works.

Table (3.1) Summary of the proposed CNN layers.

Layer name	No.layer	Description
Input layer	1	it is the input of the full CNN. In the neural network of data processing, it often represents the pixel matrix of the data
Conv2D	105	The convolutional layer is used to extract image features And the layers where filters are applied to the original image or other feature maps in a deep CNN.

ReLU	29	Non-linear activation function that is used in multi-layer neural networks (MLN). It is not linear because ReLU is an activation function in Neural Networks. Instead of using the classic Sigmoid function, which attributes probabilities for each neuron (>50% if X>0), attributes the value itself, and 0 if X is negative. The function is not linear since it equals 0 as long as X is less than 0 and X itself if X is positive. Therefore, it is not linear.
Average Pooling	27	Average pooling involves calculating the average for each patch of the feature map. This means that each 2×2 square of the feature map is down sampled to the average value in the square.
Dropout	19	Dropout is a technique where randomly selected neurons are ignored during training. They are “dropped out” randomly. This means that their contribution to the activation of downstream neurons is temporally removed on the forward pass, and any weight updates are not applied to the neuron on the backward pass.
Flatten Layer	2	Inputting data into the next layer by transforming the data into a one-dimensional array. Create a single lengthy feature vector by flattening the output of the convolutional layers. Additionally, a "fully-connected layer" connects it to the final classification model.

Dense	2	Used to classify images based on the output from convolutional layers. Each layer in the neural network contains neurons; the weighted average is passed through a non-linear function, called an "activation function," which computes the weighted average of its input.
Softmax	2	Test the model's reliability using as Loss Function and the Cross-Entropy Function to maximize neural network performance. There are several advantages to using the Cross-Entropy Function.

The following algorithm (3.4) illustrates how CNN's multitasking algorithm operates:

Algorithm (3.4): Efficient-Net-B3 training algorithm to classify image for age estimation and gender recognition.
Input : Image with size 128 * 128 *3 (JPG format)
Output: optimum weight
Begin Step One : Splitting the data into two parts, 80% for training and 20% for the testing (Holdout Techniques). Step Two : It is based on the principle of the backbone, which consists of several blocks (7 main blocks).

Step Three : Start with input layer 128*128*3 size

Step Four : Each block contains :

- Conv2D
- Batch Normalization
- Activation using Rectified linear units (ReLU)
- Global average-pooling
- Dropout for each block

Step Five : Final layers contain :

- Flatten layer
- Two dence layer one for age and other for gender
- Softmax layer

Step Six : Return the optimum weight.

End

Chapter Four

Experimental Results

4.1 Introduction

This chapter presents the results that have been obtained after executing the proposed system and illustrate the work scenario. This chapter deals in depth with the results of the analysis of experiments.

4.2 Software and Hardware

The proposed system is implemented by using the programming language: python v. 3.8.12 (anaconda_spyder). Moreover, the proposed system is applied under the windows ten (Win10 Pro 64 Bit) operating system. The proposed system is implemented on a computer with the following specifications: Intel core i7-3537u CPU due with 8GB RAM Intel® HD Graphics 4000.

4.3 Dataset in the Proposed System

In the system, two types of data that depend on external appearance are combined and described as following:

- **IMDB Dataset**

  This dataset is the biggest one of publicly celebrities' available of face image datasets which is naturally with small size to medium in size, hardly exceeding numbers tens of thousands of images, also occasionally lacks age statistics information, and the total number of images are 460,723 from IMDB . In this proposed system, the website of IMDB is designed and created a roster of the highest 100,000 actors, based on their name, gender, date of birth, and all images connected to that separate (automatically) scraped from their profiles. Also, the dataset skulked entirely profile images from Wikipedia pages of people through similar meta data. The images lacking a timestamp were removed (the date when the image was taken)—assuming that the images with single faces reflect the actor and that the timestamp and birthdate are correct. The total number of images after trimming is 171,852 [95,96].

- **IMDB-WIKI Dataset**

A large dataset of face images with gender and age labels for training, and the total number of images are 62,328 from Wikipedia. In a proposed system using these datasets both because of deep learning, the more it is categorized into a large group, the more good results it gets. A correct system because its foundation is depth, so the proposed system uses nearly 38,138 images of the face from this dataset after trimming. Tables (4.1) and (4.2) show the dataset of proposed system's number before and after trimming.

Table (4.1): Dataset number before trimming.

IMDB	WIKI	Total
460,723	62,328	523,051
Size of dataset	7GB	
Age range	1-100	
Gender	0= Female / 1= Male	

Table (4.2): Dataset number after trimming and using in proposed system.

Dataset name	Total number	Total number (IMDB+WIKI)	Test	Train
IMDB	171,852	209,990	41,998 20%	167,992 80%
WIKI	38,138			

4.4 Image Acquisition

The stage of trimming the dataset from useless images or images that have not been identified due to significant distortion or lack of credibility, such as cartoons or severely cut-out parts of the facial features. Figure (4.1) shows the example to this images.

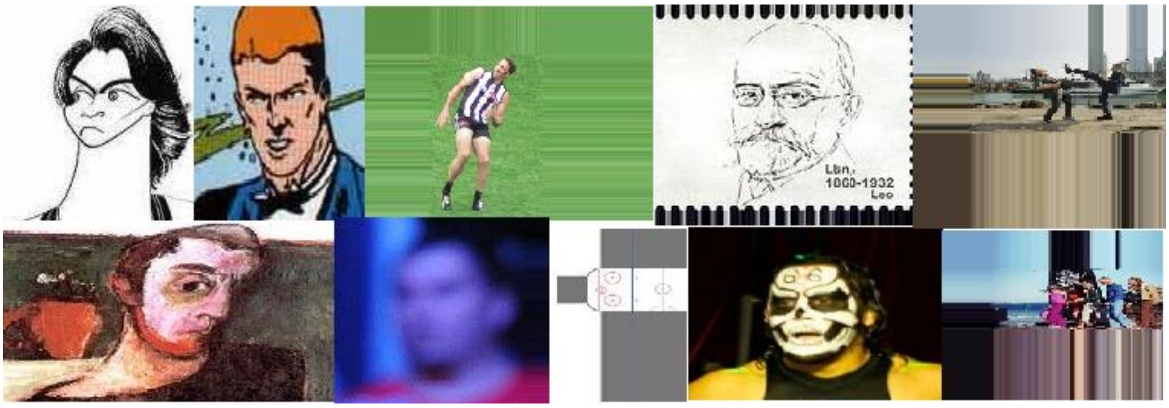


Figure (4.1): Samples of an images trimming from the dataset.

As a result of the trimming process, will be get faces that empty free from distortion to a certain degree, and this stage is important because it helps to get rid of the useless elements that can cause inaccuracy and high error rates. Figure (4.2) shows the example of the final result.

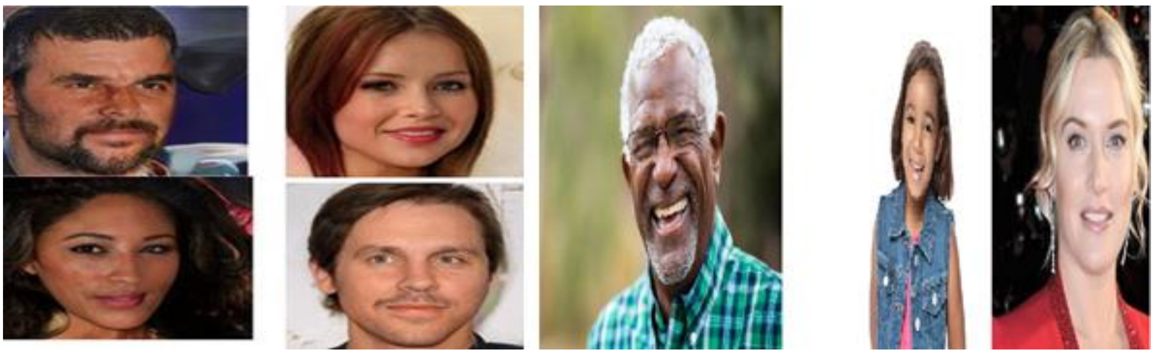


Figure (4.2): Final result in an example of the dataset.

4.5 Images Dataset Description of Ages and Genders

After trimming the dataset from the useless images, The contents of the dataset will be displayed from age groups and genders in terms of male and female according to the following schemes figure (4.3).

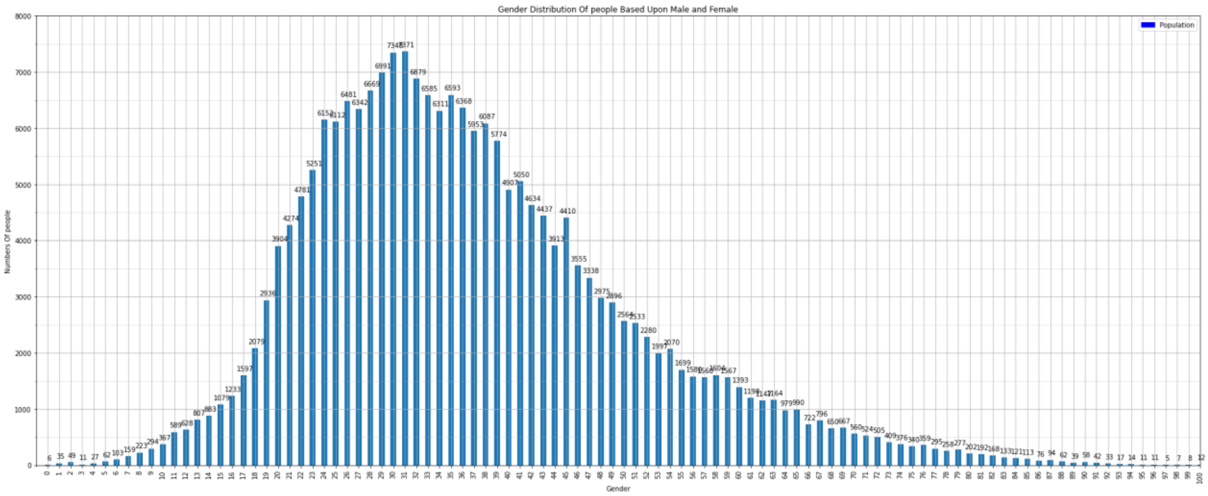


Figure (4.3): Histogram age description in proposed system dataset.

Note that in the above figure that works on to describe the ages from one year to 100 years in detail, as each year is described and what is the number of images it contains, which is considered an accurate description. Figure (4.4) describes the human genders and how many females and males are in two datasets and notices that the total number of females is less than the number of males by 87,664 females and 122,326 males.

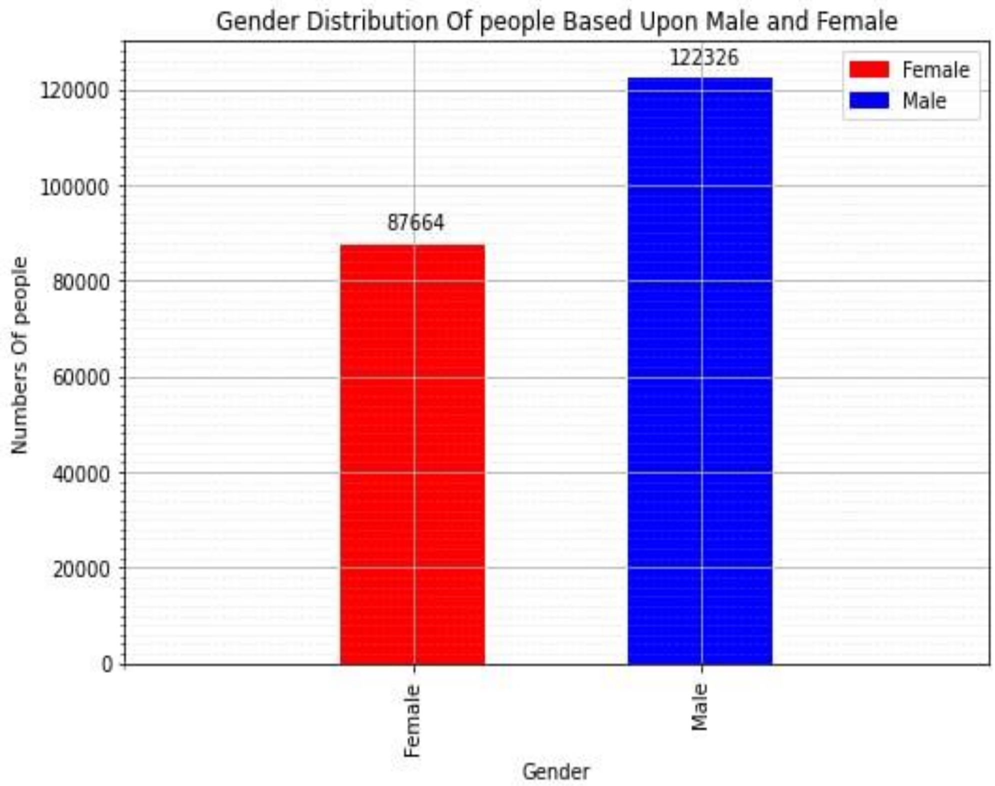


Figure (4.4): Description genders in proposed system dataset.

4.6 Pre-Processing Proposed System

The following section shows the result of the preprocessing steps of the proposed face system. Preprocessing is used for each image to enhance the image and the following points explain how the image is uploaded and what processing of the image:

First: Read the image acquisition.

Second: Apply image enhancements.

- **Region of interest (ROI).**
- **Resizing image (128*128).**
- **Data Augmentation (DA).**

This action occurs sequentially, which means that one image is entered into a form, and operations are performed on it, as detailed in Chapter Three.

To illustrate the processing operations that took place in the proposed system, one image will be used as an example to illustrate its number (nm0005017_rm3740512256_1978-12-18_2011) in the dataset and location is (IMDB_crop, folder 17), figure (4.5) shows where the image location in dataset and image sample.

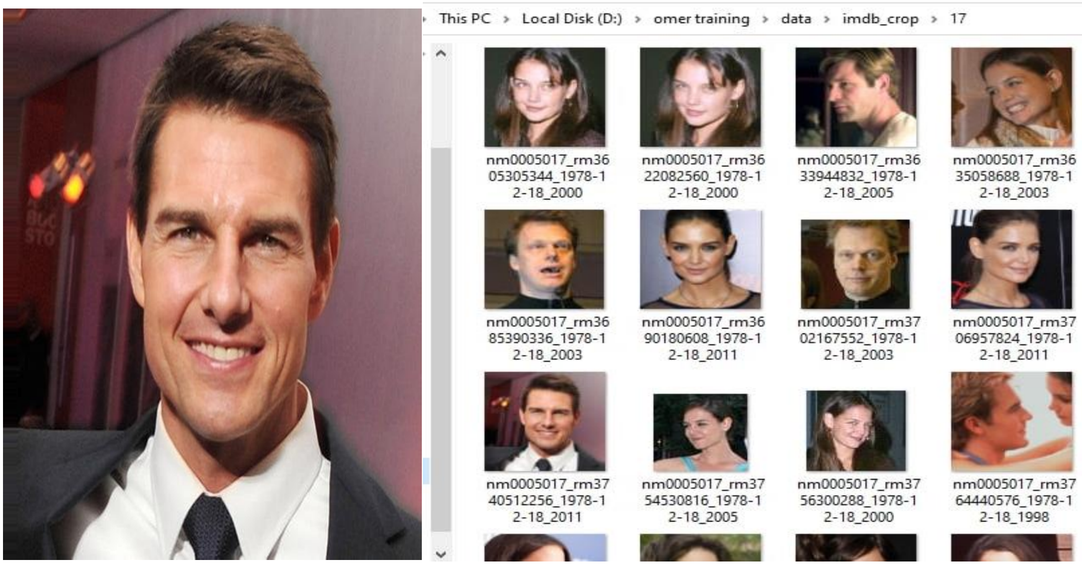


Figure (4.5): Image example and location in the dataset.

4.6.1 Region of Interest (ROI).

The region of interest (ROI) refer to a part of an image which it need to be filter or achieve many additional process on. An ROI is define by generating a binary mask, where a binary image with the similar size such as the image which wanted to be process by means of pixels which describe the ROI set to 1 as well as altogether the other pixels set to 0. ROI is sampled within a dataset identified for a particular purpose. In the 2D dataset, the boundaries of an object on an image and description are described in detail in chapter three section (3.2.2). In the proposed system, the role of the ROI was to determine the face area based on the Landmark mechanism that identifies (68) points in the face and on which the remaining processors are deducted. Figure (4.6) shows how it works and the ROI results in the face image.

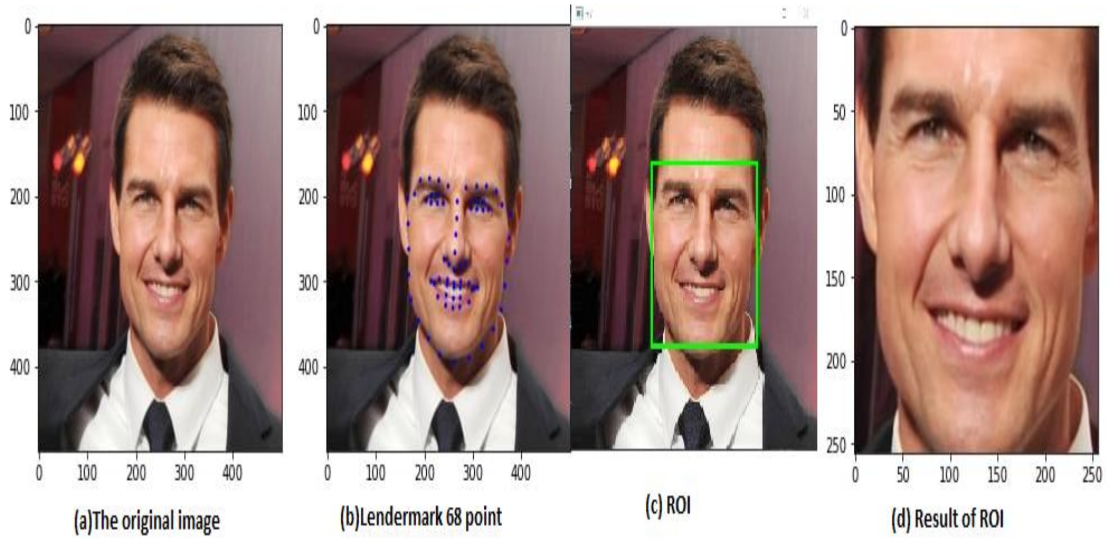


Figure (4.6): Steps of ROI.

Note that the original image (a) that was recalled from the dataset is of a size of (500 * 500) with its dimensions in the image (b) 68 points were identified and it depends on the dimensions of the face and it is essential in the subject of face recognition and is called a Landmark based on which cropping and determination of ROI are done in a image and (C) the result is an image of the face only (d) that sized is (262*260) that contains inside it the essential features that help the system to identify and estimate that mean ROI sized (262*260).

4.6.2 Resizing Image (128*128)

In this step, the image size is standardized to (128 * 128) due to the large variation in image sizes ranging from (70 * 71). The smallest size and the most significant size are (500 * 500). The primary purpose of this step is to standardize the image sizes and reduce the overfitting that occurs due to the contrast of the large data size. The selection of the size (128 * 128) was based on experience and, as explained in detail in the third chapter, section (3.2.3). Figure (4.7) shows the difference between the (a) original (500*500) and the (b) size of the ROI step and (c): the result of resizing the image.

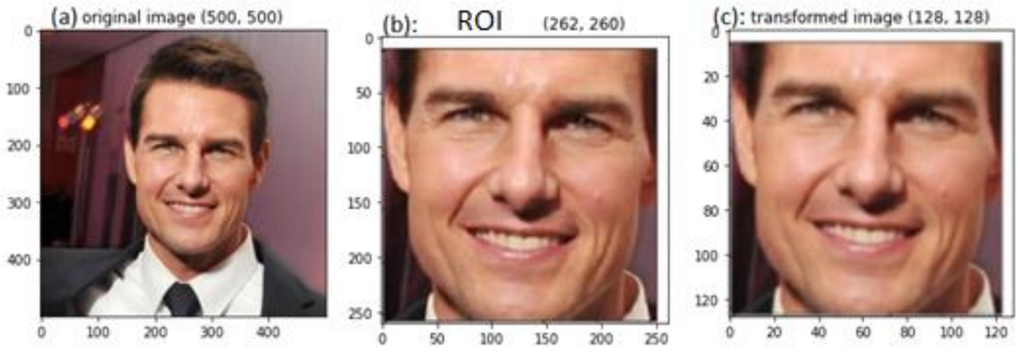


Figure (4.7): Result of resizing image.

4.6.3 Data Augmentation (DA)

Augmentation is an essential processing process because it trains the model on cases that do not exist in the dataset, and thus training on external cases that the system can encounter, as shown in figure (4.8), which shows its mechanism of work where (a) is an image from step resizing (128 *128),(b) is The lighting (brightness) factor and how to change the degree of illumination,(c) contrast factor and how to notice that the contrast of the image has changed significantly,(d) is a flip of the image horizontally right to left and (e) is transformed (rotate) the image in angle 20. An important note is that the rotation operations are done in the way shown in figure (4.9).

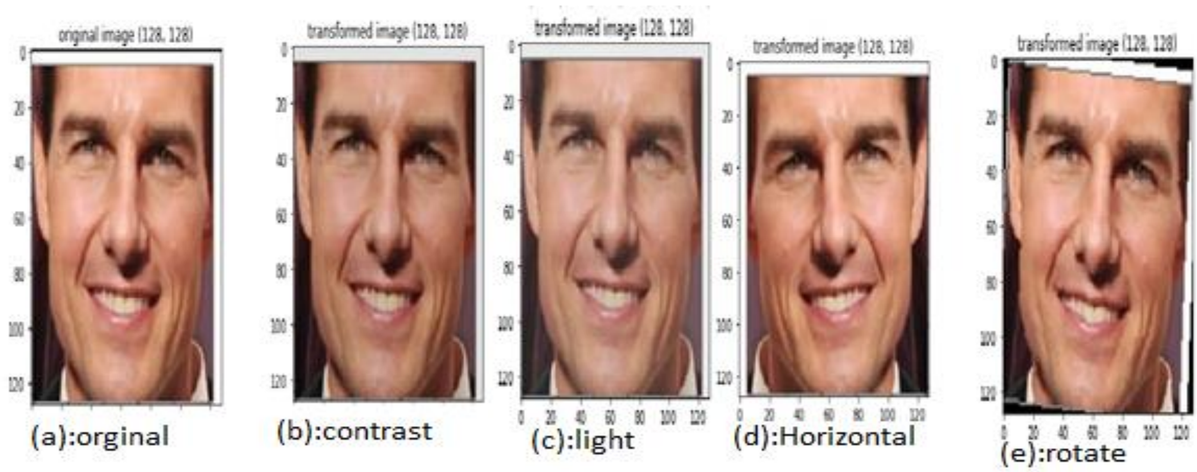


Figure (4.8): Steps of data augmentation.

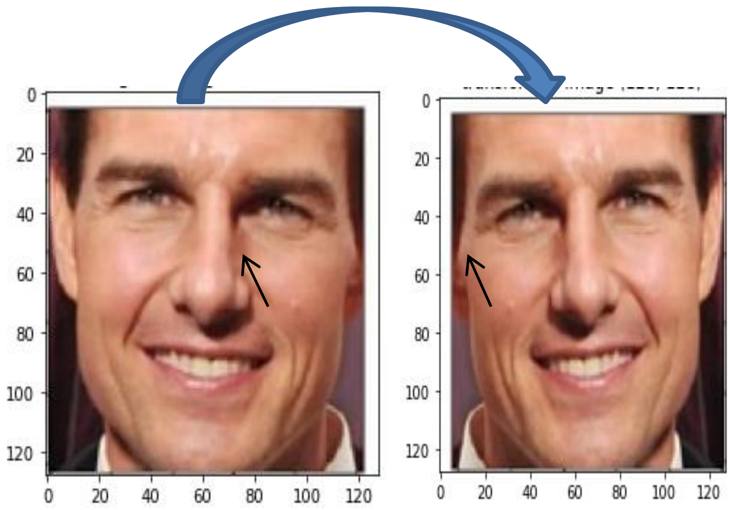


Figure (4.9): Horizontal step

4.7 Classification Using Multiple-Task CNN

The Multi-Task Learning (MTL) is a sub-part of machine learning and deep learning disciplines in which a single model learns multiple tasks at once. Improved data efficiency, less overfitting through common representations, and fast learning using supporting information are altogether benefits of some techniques. Concurrent learning of many tasks, instead, positions different design and optimization subjects, and determining which tasks must be learned together which is a significant problem in and of itself. In the proposed system, the multitasking method was adopted for the main reason to avoid the work of a hybrid system to extract two different outputs, one for age and the other for gender, and it was explained in detail in the third chapter. Table (4.3) shows the final number of structures of the algorithm.

Table (4.3): Finally number of multiple–task CNN structures.

Name	Number	Description
Total parameters	**10,941,846**	It means how many neurons are in the system
Trainable parameters	**10,854,543**	A number of weights that are

		updated during training with backpropagation.
Non-trainable parameters	**87,303**	Non-trainable parameters are those whose value is not optimized during the training as per their gradient.

Because of the main and hidden layers, the work structure cannot be placed in the thesis section because it takes up an area of 12 sheets due to the internal structure of the algorithm that depends on the multiple blocks explained in detail in the third chapter. Figure (4.10) shows the classification result regarding age estimation and gender determination.

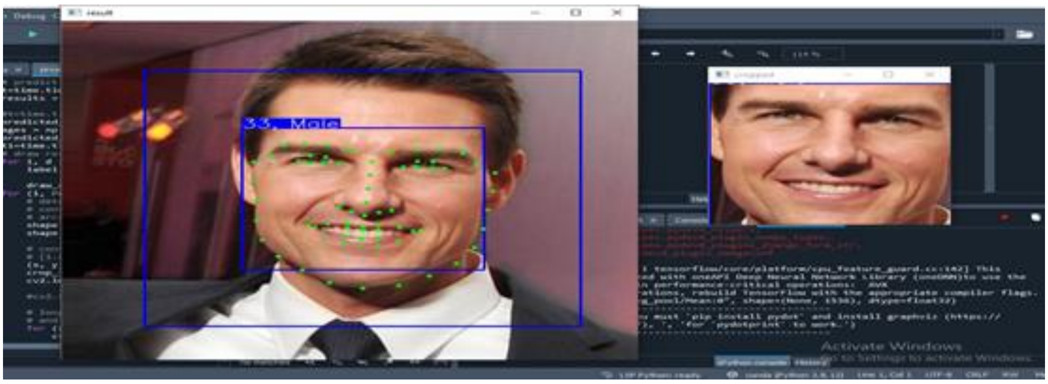


Figure (4.10): Classification result.

Note that the age was estimated at 33 years, which is the actual age of the image and its data recorded in the dataset, and its gender was distinguished, which is male.

4.8 Classification Evaluation

For evaluation, use a matrix of confusion. This summarizes the number that the classification model predicts correctly. This base is then evaluated on the experimental dataset, and the resulting performance values are compared with previously known classifications. The accuracy results of the train system are; Age accuracy =98.4%, Gender accuracy =98.2%.

In the first stage, we will review the training results only, i.e., the results of training the model with an algorithm (multiple–task CNN) to show the efficiency of the chosen algorithm in terms of training. Figure (4.11) shows the accuracy of the trained model.

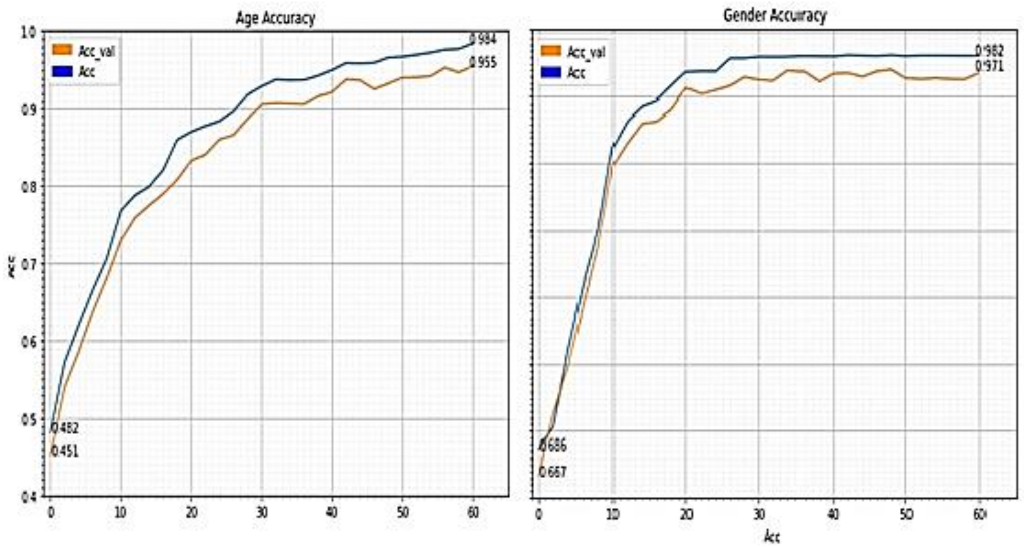


Figure (4.11): Accuracy of train model.

The train model loss was shown in figure (4.12), and the value was close to zero, which is (age loss =0.097) and (gender loss =0.160), indicating that it did not enter an excessive or incorrect condition. Furthermore, the validation loss outweighs the training loss.

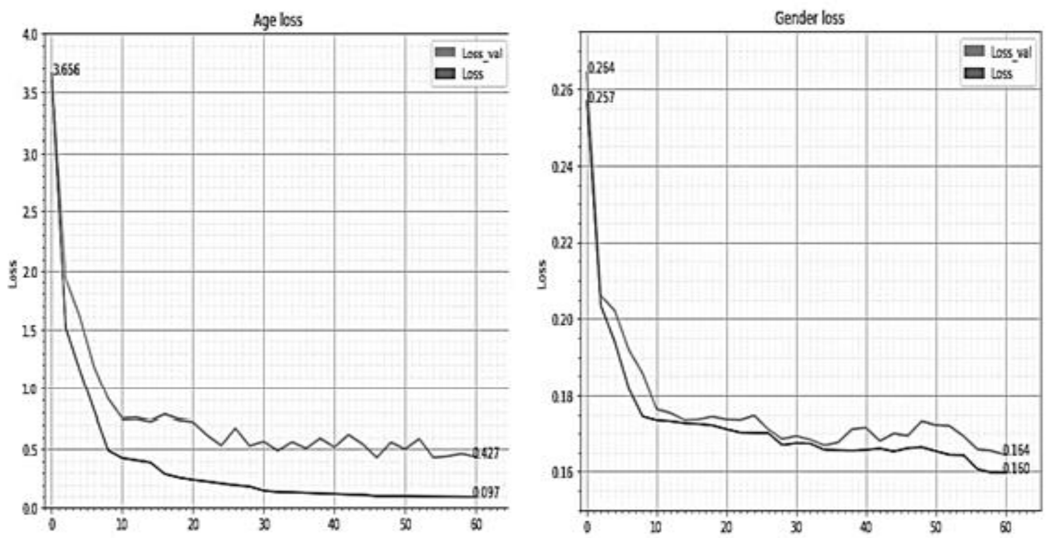


Figure (4.12): Train model loss result.

The method used to evaluate the proposed system is the holdout method,whish is the simplest method to evaluate a classifier. In this method, the dataset (a collection of data items or examples) is separated into two sets, called the training set and the test set. A classifier is a performance that assigns data objects in a collection to a target category or class. The best result based on the experiment is a random selection of the dataset with varying percentages for training (80%) and testing (20%). Table (4.4) shows the results of the most important (3 models) of the proposed system to prove that the ratio (80% training _ 20% testing) is the best.

Table (4.4): Result of the three best models in the proposed system.

No	Training - testing ratio	Age accuracy	Gender accuracy
1	Training (70%) - testing (30%)	**95.8%**	**98%**
2	Training (80%) - testing (20%)	**98.4%**	**98.2%**
3	Training (90%) - testing (10%)	**97.8%**	**98.1%**

The following figures are the three best diagrams of the results obtained from implementing the method holdout. Based on the above section, note through the implementation that the best results were in the training ratio of 80% and the test 20% through the results obtained for accuracy. The results of the three experiments were presented only because they were the best in terms of experience.

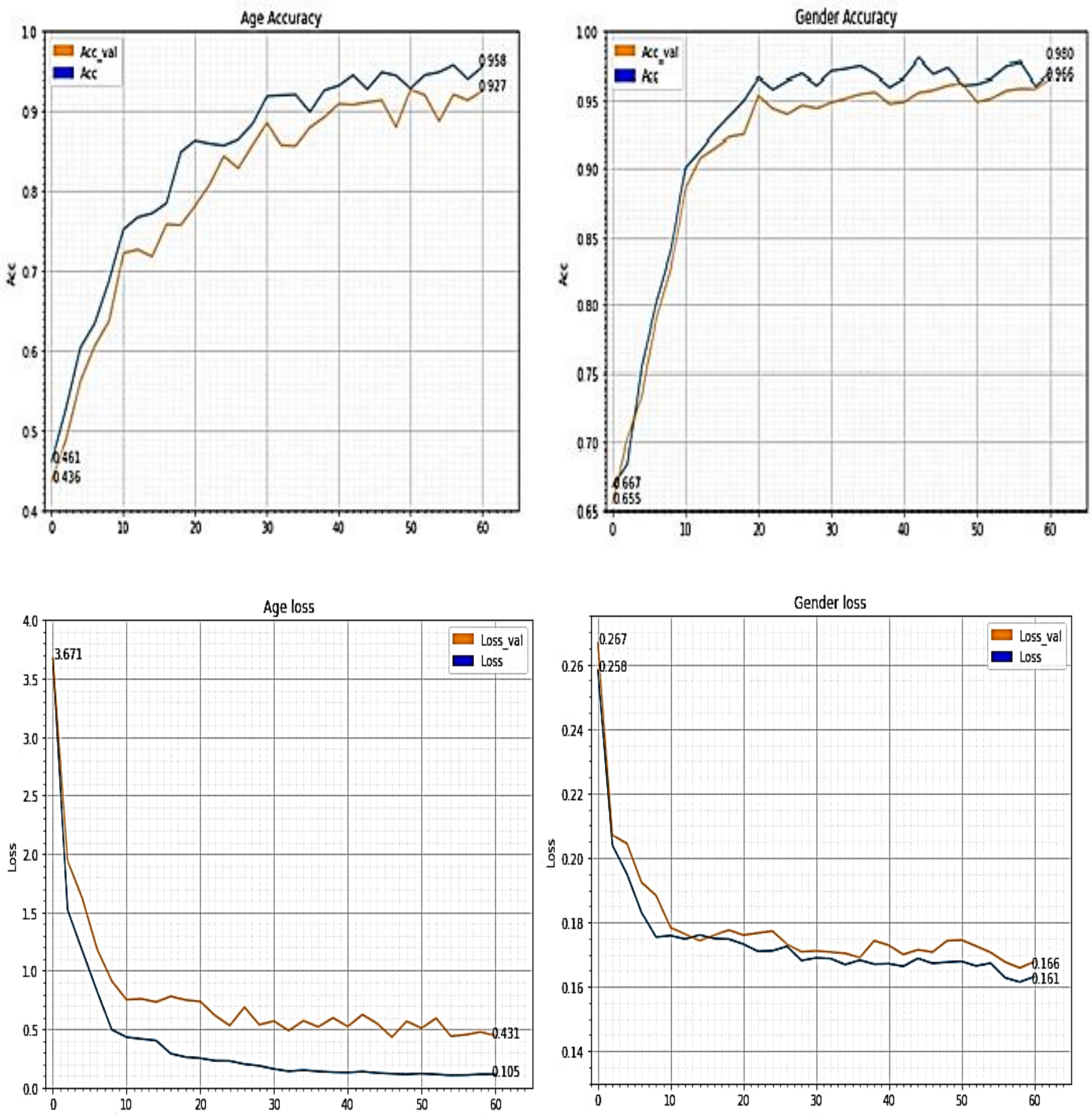


Figure (4.13): Result of training (70%) vs validation (30%).

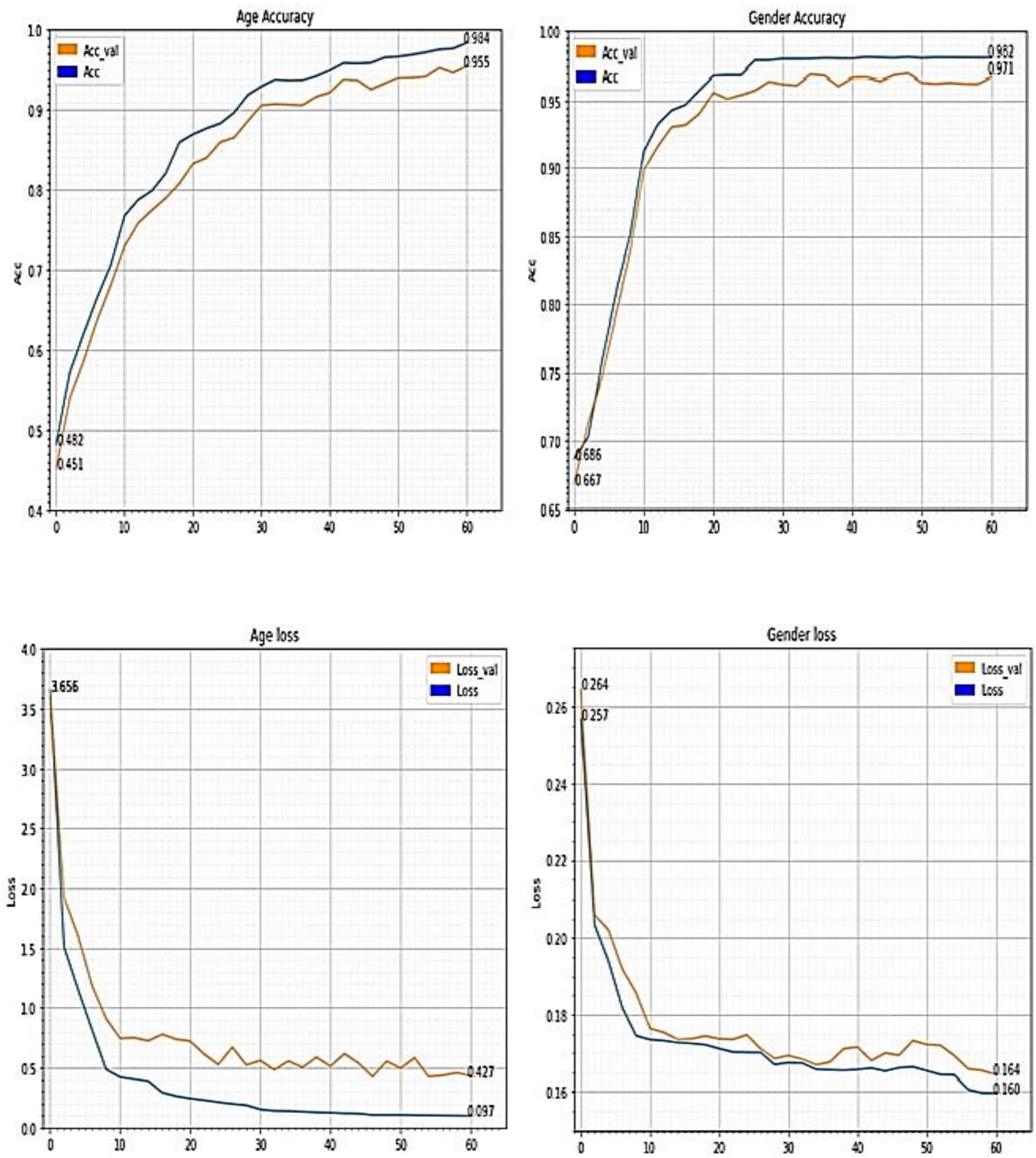


Figure (4.14): Result of training (80%) vs validation (20%).

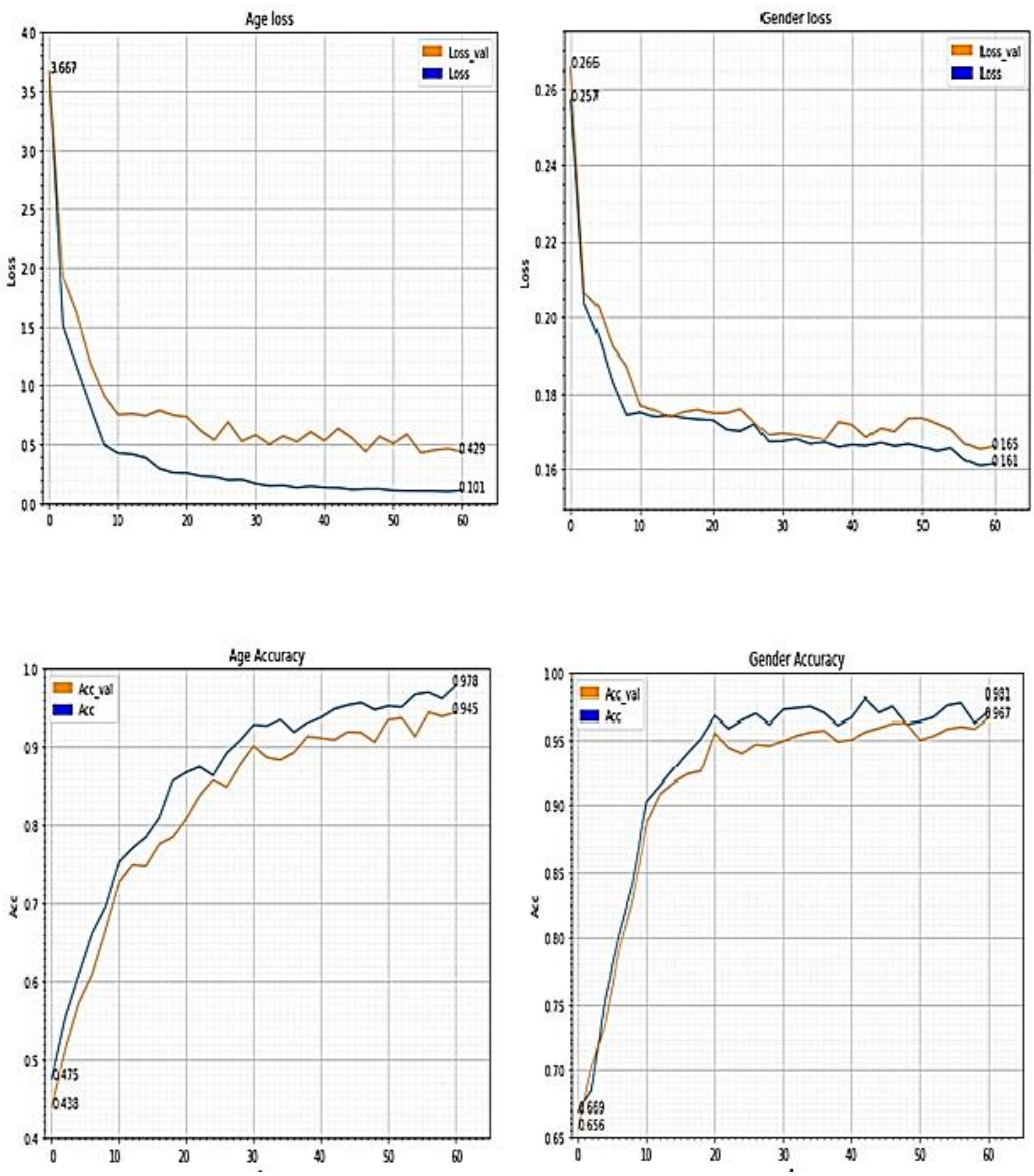


Figure (4.15): Result of training (90%) vs validation (10%).

4.9 Evaluation System

Evaluation any system using confusion matrix that is defined as a summary of classification problems which estimates the results. The amount of rights and failed predictions is calculated and broken down through the class by means of total values. The recommended approach for evaluation will be used, which is the key to the confusion matrix. A classification model's correctly or incorrectly predicted number of occurrences is summarized in a confusion matrix.

4.9.1 Confusion Matrix for Gender

The following terminology is widely used to refer to counts calculated in a confusion matrix for (Gender):

- **precision:** It is measured by the difference between the mean of a collection of data and the known value of the thing being quantified, whereas bias is measured by the difference between the mean of a collection of data and the known value of the item being quantified. Precision is the percentage of records in a group that the classifier has declared as a positive class.

When the equation (2.6) on chapter two section (2.12) is applied on test images, the following result is:

Precision (Male) =96% and Precision (Female) =94%

- **Recall**: The fraction of positive cases correctly predicted by the classifier is measured by the recall, and its value is the same as the actual positive rate.

When the equation (2.7) on chapter two section (2.12) is applied on test images, the following result is:

Recall (Male)= 95% and Recall (Female) =95%

- **F1**: The harmonic mean of accuracy and recall.

When the equation (2.8) on chapter two section (2.12) is applied on test images, the following result is:

F1-Score (Male)= 96% and F1-Score (Female) =95%.

The difference between the measured (P), (F1), and (R) is that the data collection has more males than females. The system result for this measures is shown in Table (4.5).

- **Accuracy:** When the equation (2.9) or (2.10) on chapter two section (2.12) is applied on test images, the following result is:

  Acuuracy for gender recognition = 95%

Table (4.5): Result of proposed system gender.

Gender	Precision	Recall	F1-Score
M	96%	**95%**	**96%**
F	94%	**95%**	**95%**
Accuracy	**95%**		

Below is a chart that shows the results of the values:

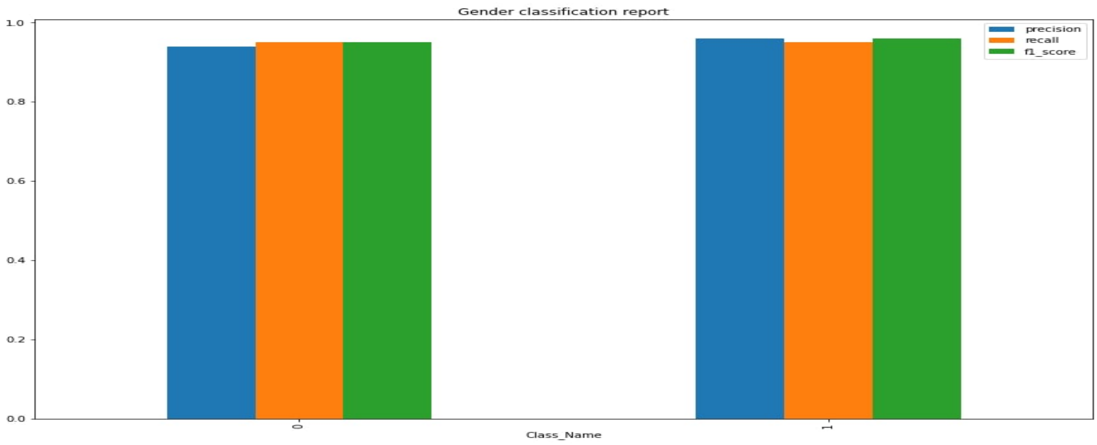


Figure (4.16): Chart of gender classification

Regarding the part of the test, the number of women and men was classified to know and clarify what part of the test contained, were female (0) and male (1), as shown in the figure (4.18) where the total value is:

- Male (1) = 23082
- Female (0) = 18916

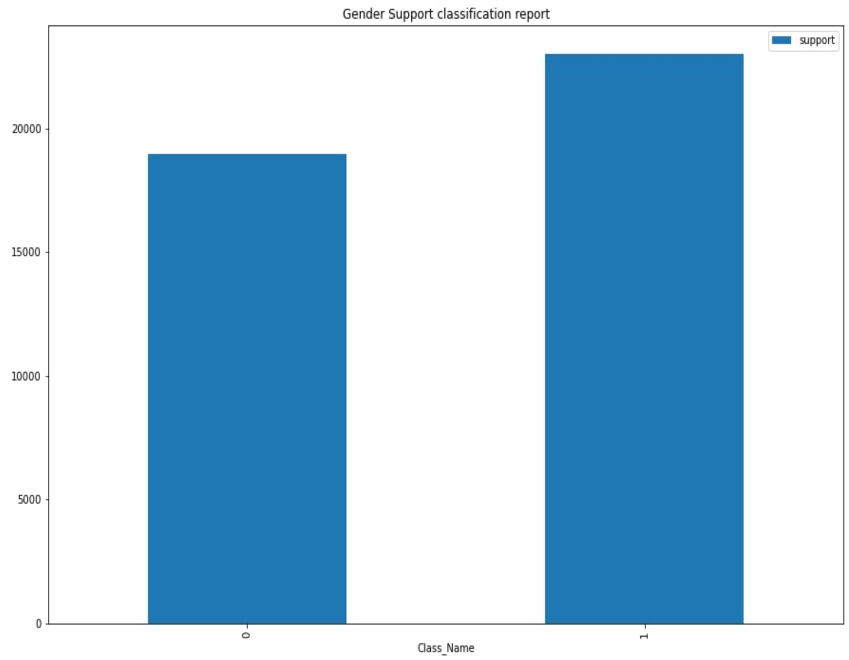


Figure (4.17): Chart of gender support classification.

The following chart shows the results of the system in full concerning gender recognition with Accuracy : 95%, Precision (male and female): 95%, Recall : 95 %, and F1: 96%.

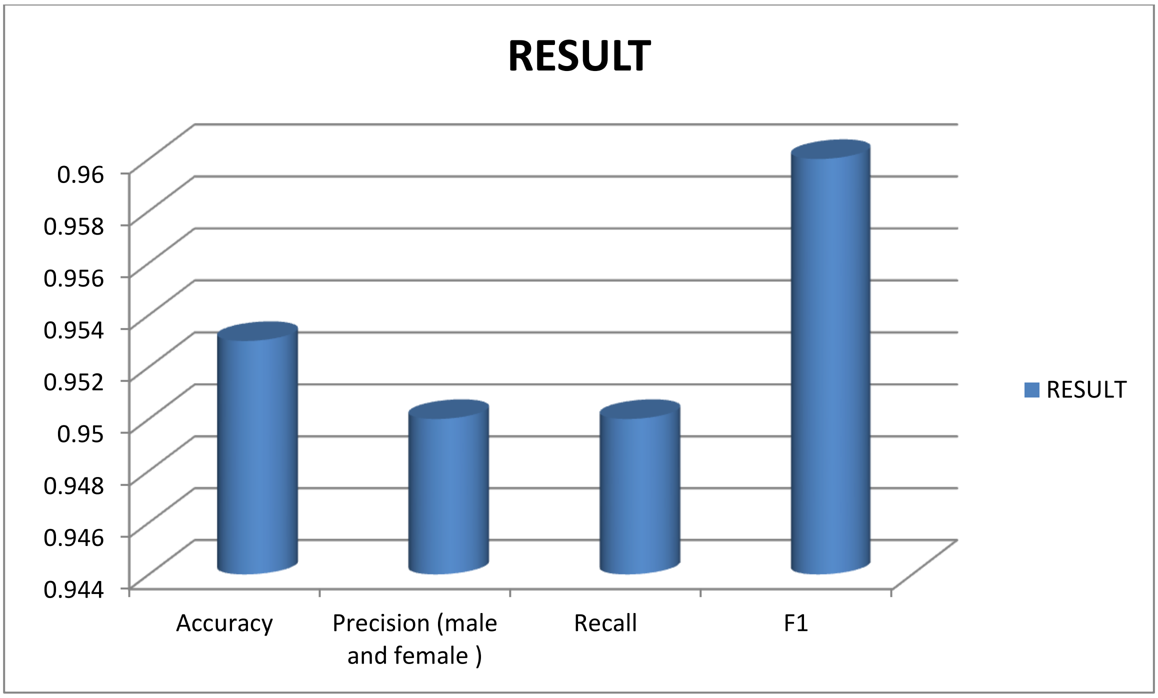


Figure (4.18): Histogram of gender results in the proposed system.

In the proposed system, the gender result is as follows: (TP, FP, TN, FN, Precision, Recall, F1). Table (4.6) illustrates the scales used in the proposed system and explains the measurements and how they were used with their results in an average way by using the mean equation and the summarization result in Table (4.6).

Table (4.6): Description of the result.

Name	Description	Result
TP	It is the value of the image from the model that is (True=1) and, in its actuality, is the resonation of gender and age	Average= $\sum \frac{total\ TP}{total\ N}$ $= \frac{17925}{41998} = 0.426$
FP	It is the value of the image from the model that is (True=1) and, in its actuality, is not the recognition (false=0)	Average= $\sum \frac{total\ FP}{total\ N}$ $= \frac{991}{41998} = 0.023$
FN	It is the value of the image from the model that is (false=0), and in its actuality, recognition (True=0)	Average= $\sum \frac{total\ FN}{total\ N}$ $= \frac{1051}{41998} = 0.025$
TN	It is the value of the image from the model that is (false=0) and in its actuality, not recognition (false=0)	Average= $\sum \frac{total\ TN}{total\ N}$ $= \frac{22031}{41998} = 0.52$
Precision	It is the accuracy of the system that counts the efficiency of the system	Average= $\sum \frac{total\ precision}{total\ N}$ Male $= \frac{4031808}{41998} = 96\%$ Female= $\frac{3947812}{41998} = 94\%$
Recall	Represents the ability of the system to detect positive(true)	Average= $\sum \frac{total\ recall}{total\ N}$ Male $= \frac{3989810}{41998} = 95\%$

		Female=$\frac{3989810}{41998}$ = 95%
F1	It represents his ability to detect cases, that is, to distinguish between classes between values (true=1and false=0)	Average=$\sum \frac{total\ F1}{total\ N}$ = Male = $\frac{4031808}{41998}$ = 96% Female=$\frac{3989810}{41998}$ = 95%

4.9.2 Confusion Matrix for Age

With regard to the test part, the ages were classified according to the class mechanism:

- **Class one = 104 ⟶ age ratio (1-11)**
- **Class two = 2364 ⟶ age ratio (11-21)**
- **Class three = 12148 ⟶ age ratio (21-31)**
- **Class four = 17759 ⟶ age ratio (31-41)**
- **Class five = 5043 ⟶ age ratio (41-51)**
- **Class six = 3230 ⟶ age ratio (51-61)**
- **Class seven = 966 ⟶ age ratio (61-71)**
- **Class eight = 294 ⟶ age ratio (71-81)**
- **Class nine = 90 ⟶ age ratio (81-91)**
- **Class ten = 0 ⟶ age ratio (91-101)**

The following chart shows all the values of each class.

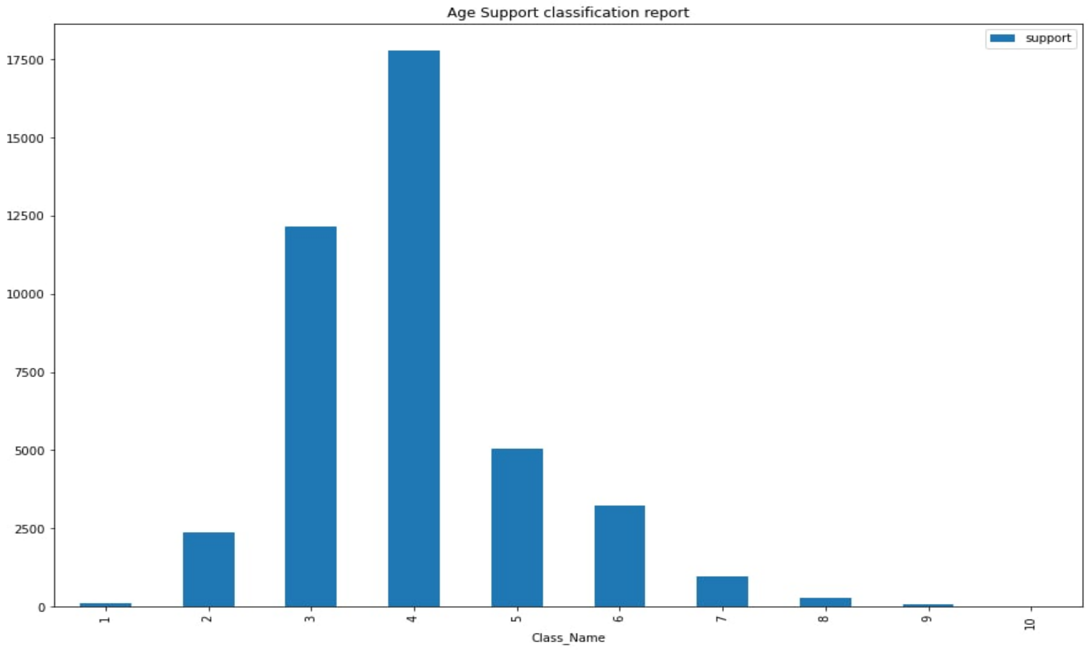


Figure (4.19): Chart of age class support.

Table (4.7) shows the results of the three measures that were adopted to evaluate the ten classes in the test part.

Table (4.7): The result of evaluating ten classes in the test part.

No.	Name	Precision	Recall	F1	Support
1	Class 1	100%	69%	82%	104
2	Class 2	99%	89%	94%	2364
3	Class 3	98%	98%	98%	12148
4	Class 4	98%	99%	98%	17759
5	Class 5	95%	100%	98%	5043
6	Class 6	100%	100%	100%	3230
7	Class 7	100%	100%	100%	966
8	Class 8	100%	100%	100%	294
9	Class 9	100%	86%	92%	90
10	Class 10	Null	Null	Null	Null

The following chart shows the results of each class separately in terms of three measures (Precision, Recall, F1-scor).

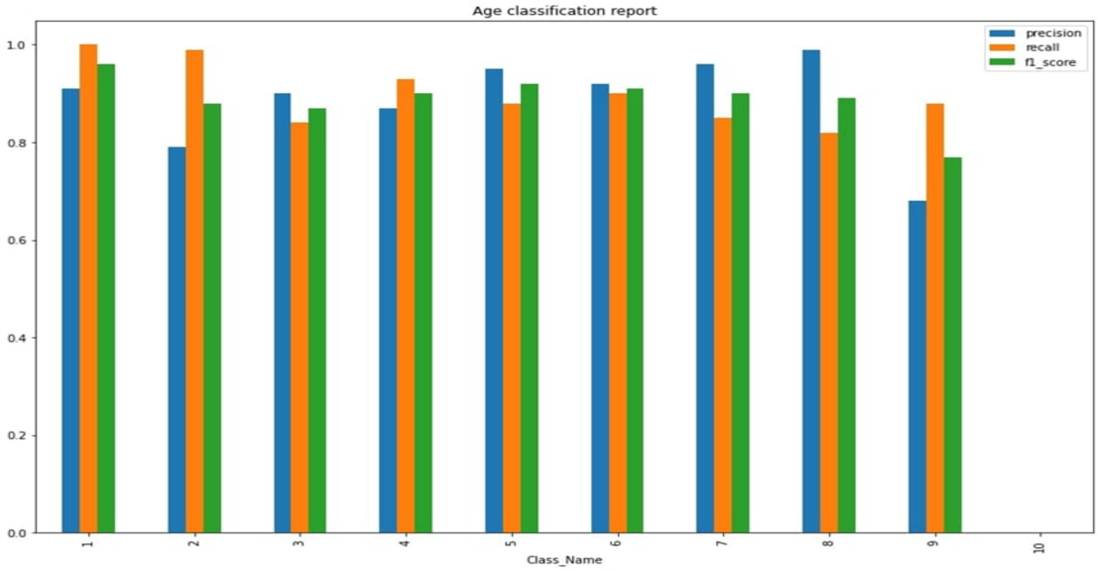


Figure (4.20): Chart of three measures for age class result.

The following chart showing the results of the system in full concerning age estimation with Accuracy : 98%, Precision: 98%, Recall : 93%, and F1: 95%.

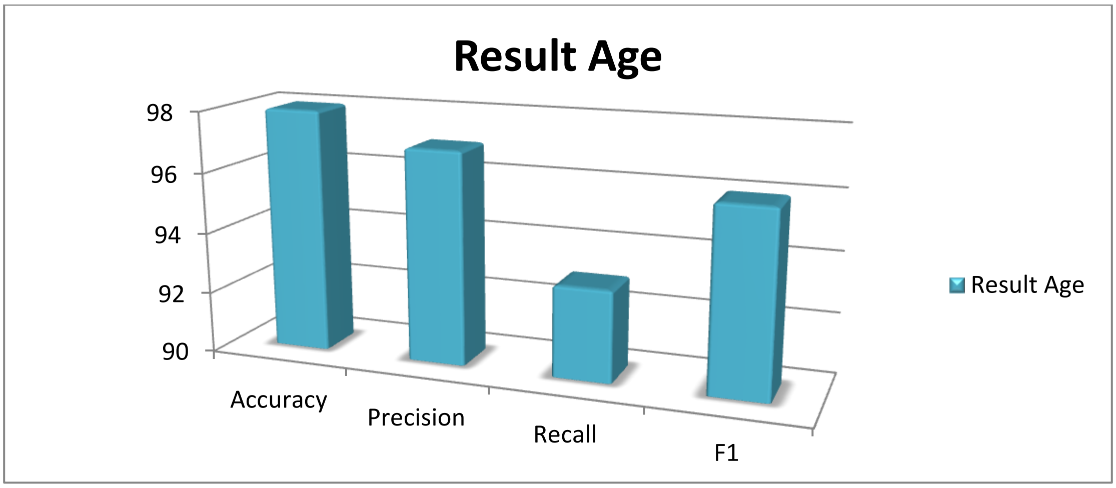


Figure (4.21): Histogram result of age system.

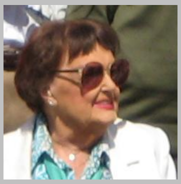
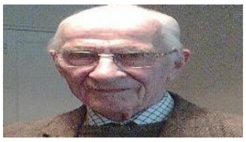
The proposed system age result is as follows: (Precision, Recall, F1). Table (4.8) illustrates the scales used in the proposed system and explains the measurements and how they were used with their results in an average way by using a mean equation. The summarization result is shown in table (4.8).

Table (4.8): Description of the result.

Name	Description	Result
Precision	It is the accuracy of the system that counts the efficiency of the system	Average= $\sum \frac{total\ precision\ class}{total\ class}$ $= \frac{890}{9} = 98\%$
Recall	Represents the ability of the system to detect positive(true)	Average= $\sum \frac{total\ recall\ class}{total\ class}$ $= \frac{841}{9} = 93\%$
F1	It represents his ability to detect cases, that is, to distinguish between classes between values (true=1and false=0)	Average= $\sum \frac{total\ F1\ class}{total\ class}$ $= \frac{862}{9} = 95\%$

An important note is that the Class No. 10 category was excluded from the calculation because the dataset from the two groups that were explained in detail in the third and fourth chapters suffered from a weakness in this category from (91 to 100) due to the lack of image and distortion and the absence of images, which also led to its un-recognition The categories that did not contain images (97, 99, and 100) were devoid of images, and categories (96, 94, and 92) were the smallest size in the dataset and had very little accuracy and suffered from deformation, while categories (91, 92, 93, 95, and 98) were a wrongly categorized image of the age by the set of data and the external appearance of the people of the categorized ages did not match before the dataset, so our proposed system did not recognize them and considered it a class 10 classification as zero with no image recognition, so the best option based on the experiment is to rely on classes and Table (4.9) shows description of range age between (91-100).

Table (4.9): Description of range age between (91-100).

NO.	Year age	Number of images	Example
1	**91**	**IMDB=1** **IMDB-WIKI=2**	**IMDB**  **IMDB-WIKI** 
2	**92**	**IMDB=3** **IMDB-WIKI=2**	**IMDB**  **IMDB-WIKI** 
3	**93**	**IMDB=3** **IMDB-WIKI=3**	**IMDB**  **IMDB-WIKI** 

4	**94**	**IMDB=3** **IMDB-WIKI=2**	**IMDB**  **IMDB-WIKI** 
5	**95**	**IMDB=2** **IMDB-WIKI=2**	**IMDB**  **IMDB-WIKI** 
6	**96**	**IMDB=3** **IMDB-WIKI=2**	**IMDB**  **IMDB-WIKI** 
7	**97**	**IMDB=0** **IMDB-WIKI=1**	**-** **IMDB-WIKI** 

8	**98**	**IMDB=2** **IMDB-WIKI=0**	
9	**99**	**IMDB=0** **IMDB-WIKI=0**	**-**
10	**100**	**IMDB=0** **IMDB-WIKI=3**	**-** **IMDB-WIKI** 

4.10 Comparison with Previous Studies

Numerous studies on face systems estimation and their conclusions have been published in recent years. The proposed approach's conclusions are compared to those of many other methodologies reported in the literature in this section. The suggested method's detection measurement in the confusion matrix is compared to past studies in table (4.10). According to the mentioned findings, our recommended technique looks to be superior to others.

Table (4.10): Comparison with previous studies.

No	Ref.No	Year	Methods	Dataset	Percentage
1	**Proposed system**	2022	Multiple Task CNN	IMDB& IMDB-WIKI	Age =98% Gender =95%
2	**[9]**	2017	CNN	IMDB-WIK &ALDIENCE	Age =66.78% Gender =93.24%
3	**[10]**	2018	CNN	IMDB-WIK	Age =52.2% Gender =88.5%

4	**[11]**	2018	CNN &DMTL	IMDB-WIK	Age(MAE)=8.59 Mean 92.59% Gender =93.86%
5	**[12]**	2019	VGG-16	IMDB-WIK	Age =57.0% Gender =88.4%
6	[13]	2020	VGG-19	IMDB-WIK	Age(MAE)=15.23 mean 85.23% Gender =91.0%
7	[7]	2021	CNN	IMDB& IMDB-WIKI	in IMDB Age =88.20% Gender =94.49% Where IMDB-WIKI is: Age =83.97% Gender =93.56%

Chapter Five

Conclusions and Suggestions for Future Work

5.1 Conclusions

This section summarizes the conclusions of the thesis, with the primary objective of predicting human age based on external appearance and defining gender, whether male or female, using deep learning techniques. The conclusion contains the following points:

1. The pre-processing process had a significant part in enhancing the effectiveness of the suggested system by applying (data augmentation), which played a crucial role in resolving the issue of the image's environment-caused dull angles and variable lighting.

2. The experimental results of the proposed system by using deep learning techniques to estimate human ages based on the external appearance and distinguish whether the human gender recognition is male or female, show that the testing results of selecting the CNN algorithm with several tasks significantly improved the systems performance.

3. The accuracy results for train which obtained by using the multi-tasking CNN method are 98.4% for age estimation and 98.2% for gender recognition and the final accuracy result of test are 98% for age estimation and 95% for gender recognition, these results when compared to earlier similar studies indicate that the suggested approach is superior in terms of classification accuracy and precision.

5.2 Suggestions for Future Work

The suggestions for future work can be summarized up as follows:

1. The human age estimation and gender recognition system can be implemented by using many other type of deep learning techniques, such as (YOLO) algorithms.
2. The system can be implemented for another media type such as video .
3. The system can be used in real time applications.

References :-

[1] M. Lal, K. Kumar, R. H. Arain, A. Maitlo, S. A. Ruk, and H. Shaikh, "Study of face recognition techniques: A survey," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 6, pp. 42–49, 2018, doi: 10.14569/IJACSA.2018.090606.

[2] M. Taskiran, N. Kahraman, and C. E. Erdem, "Face recognition: Past, present and future (a review)," Digit. Signal Process. A Rev. J., vol. 106, p. 102809, 2020, doi: 10.1016/j.dsp.2020.102809.

[3] J. J. Lv, C. Cheng, G. D. Tian, X. D. Zhou, and X. Zhou, "Landmark perturbation-based data augmentation for unconstrained face recognition," Signal Process. Image Commun., vol. 47, pp. 465–475, 2016, doi: 10.1016/j.image.2016.03.011.

[4] J. I. Olszewska, Automated Face Recognition: Challenges and Solutions. 2016.

[5] A. Mallikarjuna Reddy, V. Venkata Krishna, and L. Sumalatha, "Face recognition approaches: A survey," Int. J. Eng. Technol., vol. 7, no. 4.6 Special Issue 6, pp. 117–121, 2018, doi: 10.14419/ijet.v7i4.6.20446.

[6] P. Rodríguez, G. Cucurull, J. M. Gonfaus, F. X. Roca, and J. Gonzàlez, “Age and gender recognition in the wild with deep attention,” Pattern Recognit., vol. 72, pp. 563–571, 2017, doi: 10.1016/j.patcog.2017.06.028.

[7] M. K. Benkaddour, S. Lahlali, and M. Trabelsi, “Human Age and Gender Classification using Convolutional Neural Network,” 2020 2nd Int. Work. Human-Centric Smart Environ. Heal. Well-Being, IHSH 2020, pp. 215–220, 2021, doi:

[8] Shrestha, A. and Mahmood, A., 2019. Review of deep learning algorithms and architectures. IEEE access, 7, pp.53040-53065.

[9] K. Zhang et al., “Age Group and Gender Estimation in the Wild with Deep RoR Architecture,” IEEE Access, vol. 5, no. Cccv, pp. 22492–22503, 2017, doi: 10.1109/ACCESS.2017.2761849.

[10] S. H. Lee, S. Hosseini, H. J. Kwon, J. Moon, H. Il Koo, and N. I. Cho, “Age and gender estimation using deep residual learning network,” 2018 Int. Work. Adv. Image Technol. IWAIT 2018, pp. 1–3, 2018, doi: 10.1109/IWAIT.2018.8369763.

[11] K. Ito, H. Kawai, T. Okano, and T. Aoki, “Age and Gender Prediction from Face Images Using Convolutional Neural Network,” 2018 Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. APSIPA ASC 2018 - Proc., no. November, pp. 7–11, 2018, doi: 10.23919/APSIPA.2018.8659655.

[12] P. Giammatteo, F. V. Fiordigigli, L. Pomante, T. Di Mascio, and F. Caruso, "Age Gender Classifier for Edge Computing," 2019 8th Mediterr. Conf. Embed. Comput. MECO 2019 - Proc., no. June, pp. 6–10, 2019, doi: 10.1109/MECO.2019.8760160.

[13] Nga, C.H., Nguyen, K.T., Tran, N.C. and Wang, J.C., 2020, September. Transfer learning for gender and age prediction. In 2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan) (pp. 1-2). IEEE.

[14] A. K. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, 2004, doi: 10.1109/TCSVT.2003.818349.

[15] A. Abdolrashidi, M. Minaei, E. Azimi, and S. Minaee, "Age and Gender Prediction From Face Images Using Attentional Convolutional Network," *arxiv*, vol. 3, pp. 1–6, 2010.

[16] R. Ranjan, S. Sankaranarayanan, C. D. Castillo, and R. Chellappa, "An All-In-One Convolutional Neural Network for Face Analysis An All-In-One Convolutional Neural Network for Face Analysis," no. November, 2016.

[17] D. Bhati and V. Gupta, "Survey – A Comparative Analysis of Face Recognition Technique," Int. J. Eng. Res. Gen. Sci., vol. 3, no. 2, pp. 597–609, 2015, [Online]. Available: www.ijergs.org.

[18] J. S. Nayak and M. Indiramma, “An approach to enhance age invariant face recognition performance based on gender classification,” J. King Saud Univ. - Comput. Inf. Sci., 2021, doi: 10.1016/j.jksuci.2021.01.005.

[19] K. Ricanek, S. Bhardwaj, and michael Sodomsky, “A Review of Face Recognition against Longitudinal Child Faces,” BioSiG 2015, pp. 15–26, 2015, [Online]. Available:

[20] R. Ramachandra and C. Busch, “Presentation attack detection methods for face recognition systems: A comprehensive survey,” ACM Comput. Surv., vol. 50, no. 1, 2017, doi: 10.1145/3038924.

[21] S. Singh and S. V. A. V. Prasad, “Techniques and challenges of face recognition: A critical review,” Procedia Comput. Sci., vol. 143, pp. 536–543, 2018, doi: 10.1016/j.procs.2018.10.427.

[22] I. Sajid, M. M. Ahmed, I. Taj, M. Humayun, and F. Hameed, “Design of high performance FPGA based face recognition system,” Prog. Electromagn. Res. Symp., no. March 2014, pp. 490–496, 2008.

[23] R. Ranjan et al., “A Fast and Accurate System for Face Detection, Identification, and Verification,” IEEE Trans. Biometrics, Behav. Identity Sci., vol. 1, no. 2, pp. 82–96, 2019, doi: 10.1109/tbiom.2019.2908436.

[24] W. Wójcik, K. Gromaszek, and M. Junisbekov, “Face Recognition: Issues, Methods and Alternative Applications,” Face Recognit. - Semisupervised Classif. Subsp. Proj. Eval. Methods, 2016, doi: 10.5772/62950.

[25] R. R. Atallah, A. Kamsin, M. A. Ismail, S. A. Abdelrahman, and S. Zerdoumi, “Face Recognition and Age Estimation Implications of Changes in Facial Features: A Critical Review Study,” IEEE Access, vol. 6, pp. 28290–28304, 2018, doi: 10.1109/ACCESS.2018.2836924.

[26] J. Alghamdi et al., “A Survey on Face Recognition Algorithms,” ICCAIS 2020 - 3rd Int. Conf. Comput. Appl. Inf. Secur., pp. 0–4, 2020, doi: 10.1109/ICCAIS48893.2020.9096726.

[27] P. Li, L. Prieto, D. Mery, and P. Flynn, “Face Recognition in Low Quality Images: A Survey,” vol. 1, no. 1, 2018, [Online]. Available: http://arxiv.org/abs/1805.11519.

[28] S. Adnan and B. Alhayani, “A comprehensive survey on the biometric systems based on physiological and behavioural characteristics,” Mater. Today Proc., no. July, pp. 1–5, 2021, doi: 10.1016/j.matpr.2021.07.005.

[29] L. Li, X. Mu, S. Li, and H. Peng, “A Review of Face Recognition Technology,” IEEE Access, vol. 8, pp. 139110–139120, 2020, doi: 10.1109/ACCESS.2020.3011028.

[30] M. M. Kasar, D. Bhattacharyya, and T. H. Kim, “Face recognition using neural network: A review,” Int. J. Secur. its Appl., vol. 10, no. 3, pp. 81–100, 2016, doi: 10.14257/ijsia.2016.10.3.08.

[31] A. Abaza, M. A. Harrison, T. Bourlai, and A. Ross, “Design and evaluation of photometric image quality measures for effective face recognition,” IET Biometrics, vol. 3, no. 4, pp. 314–324, 2014, doi: 10.1049/iet-bmt.2014.0022.

[32] J. Galbally, S. Marcel, and J. Fierrez, “Biometric antispoofing methods: A survey in face recognition,” IEEE Access, vol. 2, pp. 1530–1552, 2014, doi: 10.1109/ACCESS.2014.2381273.

[33] Diwate, R. B., and S. G. Ingole. "Different face recognition techniques: A survey." International Journal for Research in Technological Studies 2, no. 5 (2016): 2348-1439.

[34] H. Aworinde, “A So Computing Model of So Biometric Traits for Gender and Ethnicity Classification,” doi: 10.5815/ijem.2019.02.05.

[35] B. Hassan, E. Izquierdo, and T. Piatrik, “Soft biometrics : a survey Benchmark analysis , open challenges and recommendations,” 2021.

[36] P. Terhörst, D. Fährmann, N. Damer, and F. Kirchbuchner, “On Soft-Biometric Information Stored in Biometric Face Embeddings,” vol. 3, no. 4, pp. 519–534, 2021, doi: 10.1109/TBIOM.2021.3093920.

[37] A. Jain and V. Kanhangad, “Gender classification in smartphones using gait information Ankita,” Expert Syst. Appl., 2017, doi: 10.1016/j.eswa.2017.10.017.

[38] I. Siddiqi, C. Djeddi, and A. Raza, “Automatic analysis of handwriting for gender classification,” 2014, doi: 10.1007/s10044-014-0371-0.

[39] C. Y. Hsu, L. E. Lin, and C. H. Lin, “Age and gender recognition with random occluded data augmentation on facial images,” *Multimed. Tools Appl.*, vol. 80, no. 8, pp. 11631–11653, 2021, doi: 10.1007/s11042-020-10141-y.

[40] S. A. Grainger, J. D. Henry, L. H. Phillips, E. J. Vanman, and R. Allen, “Age Deficits in Facial Affect Recognition: The Influence of Dynamic Cues,” Journals Gerontol. - Ser. B Psychol. Sci. Soc. Sci., vol. 72, no. 4, pp. 622–632, 2017, doi: 10.1093/geronb/gbv100.

[41] S. wu and D. Wang, “Effect of subject’s age and gender on face recognition results,” J. Vis. Commun. Image Represent., vol. 60, pp. 116–122, 2019, doi: 10.1016/j.jvcir.2019.01.013.

[42] A. Sharifara, M. S. Mohd Rahim, and Y. Anisi, “A general review of human face detection including a study of neural networks and Haar feature-based cascade classifier in face detection,” *Proc. - 2014 Int. Symp. Biometrics Secur. Technol. ISBAST 2014*, pp. 73–78, 2015, doi: 10.1109/ISBAST.2014.7013097.

[43] S. Arya, N. Pratap, and K. Bhatia, “Future of Face Recognition: A Review,” *Procedia Comput. Sci.*, vol. 58, pp. 578–585, 2015, doi: 10.1016/j.procs.2015.08.076.

[44] S. Lim, “Estimation of gender and age using CNN-based face recognition algorithm,” *Int. J. Adv. smart Converg.*, vol. 9, no. 2, pp. 203–211, 2020, [Online]. Available: http://dx.doi.org/10.7236/IJASC.2020.9.2.203.

[45] S. H. Abdurrahim, S. A. Samad, and A. B. Huddin, “Review on the effects of age, gender, and race demographics on automatic face recognition,” *Vis. Comput.*, vol. 34, no. 11, pp. 1617–1630, 2018, doi: 10.1007/s00371-017-1428-z.

[46] A. Dhomne, R. Kumar, and V. Bhan, “Gender Recognition Through Face Using Deep Learning,” *Procedia Comput. Sci.*, vol. 132, pp. 2–10, 2018, doi: 10.1016/j.procs.2018.05.053.

[47] V. Albiero, K. S. Krishnapriya, K. Vangara, K. Zhang, M. C. King, and K. W. Bowyer, "Analysis of Gender Inequality in Face Recognition Accuracy," *Proc. - 2020 IEEE Winter Conf. Appl. Comput. Vis. Work. WACVW 2020*, pp. 81–89, 2020, doi: 10.1109/WACVW50321.2020.9096947.

[48] A. Dehghan, E. G. Ortiz, G. Shu, and S. Z. Masood, "DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Network," 2017, [Online]. Available: http://arxiv.org/abs/1702.04280.

[49] M. Schwabacher and K. Goebel, "A survey of artificial intelligence for prognostics," AAAI Fall Symp. - Tech. Rep., pp. 107–114, 2007.

[50] K. Yasaka and O. Abe, "Deep learning and artificial intelligence in radiology: Current applications and future directions," PLoS Med., vol. 15, no. 11, pp. 2–5, 2018, doi: 10.1371/journal.pmed.1002707.

[51] Q. Bi, K. E. Goodman, J. Kaminsky, and J. Lessler, "What is machine learning? A primer for the epidemiologist," Am. J. Epidemiol., vol. 188, no. 12, pp. 2222–2239, 2019, doi: 10.1093/aje/kwz189.

[52] S. Chang, T. Cohen, and B. Ostdiek, "What is the machine learning?," Phys. Rev. D, vol. 97, no. 5, p. 56009, 2018, doi: 10.1103/PhysRevD.97.056009.

[53] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[54] L. Deng and D. Yu, “Deep learning: Methods and applications,” 2013.

[55] L. Arnold, S. Rebecchi, and S. Chevallier, “An Introduction to Deep Learning To cite this version,” Eur. Symp. Artif. Neural Networks, no. January, 2011.

[56] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, “Deep learning in computer vision: A critical review of emerging techniques and application scenarios,” Mach. Learn. with Appl., vol. 6, no. August, p. 100134, 2021, doi: 10.1016/j.mlwa.2021.100134.

[57] K. G. Kim, “Deep learning book review,” Nature, vol. 29, no. 7553, pp. 1–73, 2019.

[58] V. Sze, Y. H. Chen, T. J. Yang, and J. S. Emer, “Efficient Processing of Deep Neural Networks: A Tutorial and Survey,” Proc. IEEE, vol. 105, no. 12, pp. 2295–2329, 2017, doi: 10.1109/JPROC.2017.2761740.

[59] J. Wang and L. Perez, “The Effectiveness of Data Augmentation in Image Classification using Deep Learning,” arxiv, no. 1712, pp. 1–8, 2017.

[60] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” J. Big Data, 2019, doi: 10.1186/s40537-019-0197-0.

[61] J. Lv et al., “Data Augmentation for Face Recognition,” Neurocomputing, 2016, doi: 10.1016/j.neucom.2016.12.025.

[62] A. Zhuchkov, “Analyzing the effectiveness of image augmentations for face recognition from limited data,” 2021, doi: 10.1109/nir52917.2021.9666135.

[63] H. Noh, T. You, J. Mun, and B. Han, “Regularizing deep neural networks by noise: Its interpretation and optimization,” Adv. Neural Inf. Process. Syst., vol. 2017-Decem, no. Nips, pp. 5110–5119, 2017.

[64] A. Galdran et al., “Data-Driven Color Augmentation Techniques for Deep Skin Image Analysis,” pp. 1–4, 2017, [Online]. Available: http://arxiv.org/abs/1703.03702.

[65] X. Wang, K. Wang, and S. Lian, “A Survey on Face Data Augmentation,” no. 1, 2019, doi: 10.1007/s00521-020-04748-3.

[66] Demuth, Howard B., and Mark H. Beale. , Orlando De Jess, and Martin T Hagan. “Neural Network Design”, Martin Hagan, 2014.

[67] S. Balaban, “Deep learning and face recognition: the state of the art,” *Biometric Surveill. Technol. Hum. Act. Identif. XII*, vol. 9457, no. May 2015, p. 94570B, 2015, doi: 10.1117/12.2181526.

[68] L. Alzubaidi et al., Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,vol.8,no. 1. Springer International Publishing, 2021.

[69] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, “Review on Convolutional Neural Networks (CNN) in vegetation remote sensing,” ISPRS J. Photogramm. Remote Sens., vol. 173, pp. 24–49, 2021, doi: 10.1016/j.isprsjprs.2020.12.010.

[70] S. Gollapudi, “Deep Learning for Computer Vision,” Learn Comput. Vis. Using OpenCV, pp. 51–69, 2019, doi: 10.1007/978-1-4842-4261-2_3.

[71] Zeiler, Matthew D, “Hierarchical convolutional deep learning in computer vision”, Ph.D Thesis, New York University, 2013.

[72] Yamashita, Rikiya, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi, “Convolutional neural networks: an overview and application in radiology”, Insights into imaging 9, PP 611-629, 2018.

[73] K. Ovtcharov, O. Ruwase, J. Kim, J. Fowers, K. Strauss, and E. S. Chung, “Accelerating Deep Convolutional Neural Networks Using Specialized Hardware,” Microsoft Res. Whitepaper, pp. 3–6, 2015.

[74] Demuth, Howard B., and Mark H. Beale. , Orlando De Jess, and Martin T Hagan. “Neural Network Design”, Martin Hagan, 2014.

[75] G. Guo and N. Zhang, “A survey on deep learning based face recognition,” Comput. Vis. Image Underst., vol. 189, no. August, p. 102805, 2019, doi: 10.1016/j.cviu.2019.102805.

[76] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton, “Deep learning”, Nature 521, No. 7553, PP 436-444, 2015.

[77] Ioffe, Sergey, and Christian Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift.”, arXiv preprint arXiv:1502.03167, 2015.

[78] Lee, J.H., Kim, Y.J. and Kim, K.G., 2020. Bone age estimation using deep learning and hand X-ray images. Biomedical Engineering Letters, 10(3), pp.323-331.

[79] S. Yadav and S. Shukla, “Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification,” in 2016 IEEE 6th International conference on advanced computing (IACC), 2016, pp. 78–83.

[80] D. Berrar, “Cross-Valaidation.” 2019.

[81] P. Refaeilzadeh, L. Tang, and H. Liu, “Cross-validation.,” Encycl. database Syst., vol. 5, pp. 532–538, 2009.

[82] Guterriez, “LECTURE 13: Cross-validation,” 2002.

[83] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in Ijcai, 1995, vol. 14, no. 2, pp. 1137–1145.

[84] I. Aruleba and S. Viriri, Deep Learning for Age Estimation Using EfficientNet, vol. 12861 LNCS, no. February 2022. Springer International Publishing, 2021. doi: 10.1007/978-3-030-85030-2_34.

[85] H. Alhichri, A. S. Alswayed, Y. Bazi, N. Ammour, and N. A. Alajlan, “Classification of Remote Sensing Images Using EfficientNet-B3 CNN Model with Attention,” IEEE Access, vol. 9, pp. 14078–14094, 2021, doi: 10.1109/ACCESS.2021.3051085.

[86] Atila, Ü., Uçar, M., Akyol, K. and Uçar, E., 2021. Plant leaf disease classification using EfficientNet deep learning model. Ecological Informatics, 61, p.101182.

[87] Tan, M. and Le, Q., 2019, May. Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.

[88] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[89] N. D. Marom, L. Rokach, and A. Shmilovici, "Using the confusion matrix for improving ensemble classifiers," 2010 IEEE 26th Conv. Electr. Electron. Eng. Isr. IEEEI 2010, pp. 555–559, 2010, doi: 10.1109/EEEI.2010.5662159.

[90] C. Manliguez, "Generalized Confusion Matrix for Multiple Classes," no. November, pp. 5–7, 2016, doi: 10.13140/RG.2.2.31150.51523.

[91] Mustapha, M.F., Mohamad, N.M., Osman, G. and Ab Hamid, S.H., 2021, November. Age Group Classification using Convolutional Neural Network (CNN). In Journal of Physics: Conference Series (Vol. 2084, No. 1, p. 012028). IOP Publishing.

[92] AL-Jibory, F.K., 2021. Hybrid system for plagiarism detection on a scientific paper. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(13), pp.5707-5719.

[93] Reem Majeed Ibrahim, "Classification of Face Image Based on Gender Using Intelligent Method", M.s.C electronic thesis, Computer Sciences, University of Technology, 2015.

[94] Elmahmudi, A. and Ugail, H., 2021. A framework for facial age progression and regression using exemplar face templates. The visual computer, 37(7), pp.2023-2038.

[95] "IMDB." https://www.imdb.com/interfaces/.

[96] "IMDB-WIKI." https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/.

الملخص

لا يزال تقدير العمر وتحديد الجنس البشري من البيانات المرئية مثل صور الوجه موضوعًا بحثيًا رائعًا نظرًا لتنوع التطبيقات المستخدمة المحتملة. تشمل أهم التطبيقات تطوير واجهات ذكية بين الإنسان والآلة وتحسين السلامة والأمن في مختلف المجالات مثل النقل والأمن والطبية وما إلى ذلك ، ومع ذلك ، تنوع الخصائص مثل جراحة تجميل الوجه وشعر الوجه والتجاعيد وحالة الجلد والأجناس البشرية المختلفة ، والعوامل الخارجية تؤثر على مظهر الوجه وهو أمر حاسم لتحديد الجنس وتقدير العمر. هذه الخصائص تجعل هذه المنطقة موضوعًا صعبًا وتعتبر أرضًا خصبة للباحثين. ومع ذلك ، كانت الشبكة العصبية العميقة من أهم التقنيات التي تم استخدامها مؤخرًا لتحديد تقدير الجنس والعمر البشري.

في هذه الأطروحة ، تم اقتراح نظام فعال لتحديد الجنس وتقدير عمر الإنسان من صور الوجه بناءً على المظهر الخارجي باستخدام تقنية الشبكة العصبية العميقة (خوارزمية الشبكة العصبية التلافيفية متعددة المهام (MCNN)). يتكون البناء العام للنظام المقترح من عدة مراحل ؛ أولاً ، مرحلة الحصول على الصورة ؛ ثانيًا ، مرحلة المعالجة المسبقة (باستخدام تقنية منطقة الاهتمام ، وتغيير حجم الصورة ، وزيادة البيانات) ؛ ثالثًا ، مرحلة استخراج الميزات ؛ أخيرًا ، مرحلة التصنيف التي تنتج نتيجتين ؛ تقدير العمر والتعرف على الجنس.

يتم تنفيذ النظام المقترح باستخدام قاعدة بيانات الأفلام على الإنترنت (IMDB) و IMDB-WIKI معًا. أظهرت النتائج المتحصل عليها أن النظام المقترح يقدم دقة 98.2٪ للتعرف على الجنس البشري ودقة 98.4٪ لتقدير العمر. علاوة على ذلك ، حقق هذا النظام المقترح نتائج أفضل من الأعمال السابقة.

جمهورية العراق

وزارة التعليم العالي والبحث العلمي

جامعة ديالى

كلية العلوم

تصنيف الجنس والعمر بالاعتماد على صور الوجه باستخدام الشبكات العصبية العميقة

رسالة

مقدمة الى قسم علوم الحاسوب / كلية العلوم/ جامعة ديالى وهي جزء من متطلبات نيل درجة الماجستير في علوم الحاسوب

من قبل

عمر عامر محمد

بإشراف

أ.م. د. جمال مصطفى التويجري

1444 هـ **2022م**