[1]Priyanka Singh

[2]Dr. Rajeev G. Vishwakarma

Detection of Gender in Crowds Using ResNet Model

Journal of Electrical Systems

Abstract: - The ResNet model is used in this investigation to suggest a gender detection solution for use in congested settings. Due to occlusions, varied stances, and various features, determining a person's gender in crowded surroundings may be a difficult and time-consuming job. The ResNet model, which is a deep convolutional neural network architecture, is used to solve these difficulties because of its capacity to capture detailed characteristics and its efficiency in managing deep network structures. The strategy that has been suggested entails preprocessing the input photos, sending those images through the ResNet model, and then extracting gender-related characteristics from those images. The ResNet model is made up of a number of residual blocks with skip connections, which makes it easier to learn complicated representations. After that, the learnt characteristics are input into fully linked layers, and then softmax activation is used to determine the subject's gender. The usefulness of the technique that was developed was shown by experimental findings on a large dataset, which achieved a high level of accuracy in gender determination. The use of the ResNet model helps the system to handle complicated scenarios and improves the system's ability to accurately recognize gender in situations with a large number of people. The method that has been developed has the potential to find applications in areas such as surveillance, crowd control, and the study of social behavior.

Keywords: Resnet50 , Resnet101, Resnet152, Gender , Deep Convolutional Neural Network.

I. INTRODUCTION

In crowded environments, determining a person's gender may be a difficult process owing to a number of variables including occlusions, differences in stances, and distinct physical characteristics [1]. However, effective gender recognition in situations like these has important implications for applications in crowd control, surveillance, and the study of social behavior [2]. Traditional techniques of gender identification sometimes fail to address the intricacies of congested surroundings, which leads to limited accuracy and resilience in the results they provide [3].

Deep learning strategies, in particular convolutional neural networks (CNNs), have shown exceptional progress in computer vision applications in recent years, including gender detection. ResNet (Residual Network), one of the CNN architectures, has emerged as a strong model owing to its capacity to efficiently manage deep network topologies and capture detailed characteristics [4]. This is one of the reasons why ResNet has become so popular. ResNet makes use of residual blocks with skip connections; this facilitates the learning of complicated representations while also addressing the issue of vanishing gradients [5].

In this research project, we offer a strategy for detecting gender in crowded settings by making use of the ResNet model. Our goal is to increase the accuracy and resilience of the gender recognition process in congested settings [6]. We hope that by using the capabilities of ResNet, we will be able to overcome the problems that are presented by occlusions, different positions, and different looks.

The strategy that has been suggested entails preprocessing the input photos, sending those images through the ResNet model, and then extracting gender-related characteristics from those images. The ResNet model is made up of a number of residual blocks that have skip connections. These connections allow the network to recognize intricate patterns and fine-grained features even when presented with very packed situations. After that, the newly

[1,2]Department of Computer Science and Engineering

[1]Research Scholar, Dr. A. P. J. Abdul Kalam University, Indore

[2]Research Supervisor and Pro Vice-Chancellor, Dr. A. P. J. Abdul Kalam University, Indore, (M.P.), India.

E-mail Id: [1]priyankaasinghbaghel@gmail.com,

* Corresponding Author: Priyanka Singh Email: priyankaasinghbaghel@gmail.com

learnt characteristics are input into fully connected layers, and then softmax activation is performed in order to categorize the people in the crowd according to their gender.

We test the performance of our strategy by comparing it to both conventional approaches and other deep learning models using a congested dataset. The results of the experiments show that the strategy that was presented is successful; it is possible to get a high level of accuracy in gender determination even in demanding circumstances that include crowded settings. This study makes a contribution to the progress of gender recognition methods, especially in crowded circumstances. Additionally, this research has the potential to affect multiple areas, including surveillance, crowd control, and social behavior analysis.

The detection of gender in crowds using the ResNet model focuses on leveraging the deep learning capabilities of Residual Networks to accurately identify and classify individuals' genders within crowded scenes. ResNet's architecture, known for its deep layers and ability to avoid the vanishing gradient problem, enhances the model's efficiency in extracting and learning complex features from images. By processing visual data from crowded environments, the ResNet model can distinguish between male and female genders with high accuracy.

the rest of the paper is meticulously organized into six sections. The first section provides an introduction, setting the stage for the subsequent exploration of the topic. Section 2 delves into a background study, offering a foundational understanding necessary for grasping the context of the research. Following this, Section 3 presents a comprehensive literature review, critically analyzing previous studies and identifying gaps that the current research aims to fill. Section 4 introduces the proposed method, detailing the innovative approaches and techniques employed in this study. Section 5 covers the implementation and results, where the practical application of the proposed method is demonstrated, and its outcomes are thoroughly evaluated. Finally, the paper concludes with a summary of findings, implications, and potential directions for future research in the last section.

II. BACKGROUND STUDY

Crowd management, public safety, and the study of social behavior are just few of the many important applications that benefit greatly from the ability to accurately identify individuals' genders [7]. Researchers working in the fields of computer vision and pattern recognition have shown a substantial amount of interest in developing methods that can reliably and automatically discern the gender of persons present in crowded environments.

Traditional methods of gender identification often depend on hand-crafted elements like face landmarks, textural descriptors, or color-based representations to make their determinations. These technologies, although being successful in controlled contexts, are not ideal for use in crowded settings because of the problems posed by occlusions, fluctuations in stances, and complicated relationships between persons [8]. As a consequence of this, their performance has a tendency to become noticeably worse under conditions like these.

Computer vision tasks, such as gender identification, have been completely transformed as a result of the development of deep learning and the use of convolutional neural networks (CNNs). CNNs have the ability to automatically learn discriminative features from raw data, which enables them to capture complicated patterns as well as fluctuations in those patterns [9]. The accuracy and reliability of gender detection have both seen considerable advances as a result of this.

The Residual Network, often known as ResNet, is considered to be one of the most important CNN designs. ResNet was the first system to propose the idea of residual blocks, which make use of skip connections to solve the issue of vanishing gradients and make it possible to train very deep networks [10]. Image classification, object identification, and segmentation are all areas in which ResNet has become a very popular option as a result of its extraordinary performance in a variety of computer vision applications.

The use of ResNet for gender identification in congested settings carries with it a number of distinct benefits. To begin, the capability of ResNet to deal with deep network structures makes it possible for the model to recognize complicated characteristics and acquire sophisticated representations from pictures that are densely packed [11]. The skip connections in the residual blocks provide a better flow of information and make it easier to represent relationships between local and global characteristics.

Additionally, the deep representation that was learnt by ResNet is able to efficiently handle occlusions as well as changes in postures, both of which are typically seen in situations that are cluttered. Due to the model's resistance

to background noise and its ability to accurately capture fine-grained information, it is well suited for gender recognition tasks that take place in difficult situations.

There have been a number of research that have investigated the use of ResNet and other forms of deep learning models for gender identification in crowded settings. The results of these investigations show that conventional approaches might be significantly improved upon in terms of their accuracy [12]. However, further study is still required in order to improve and fine-tune the ResNet design in a way that is particular to the identification of gender in congested settings.

III. LITERATURE REVIEW

This study developed a VGG model adaptation for real-time gender classification in crowd recordings, addressing the high computational demands of conventional VGG models for embedded systems with limited resources. By exploring a modified VGG architecture, specifically VGG11 within a VGG22 framework, and employing model compression, pruning, quantization, and knowledge distillation techniques, the study achieved significant performance improvements in accuracy, precision, recall, and F1-score. The optimized VGG22 model demonstrated superior performance across all metrics, proving that deep model modifications can lead to efficient gender classification in crowded settings while meeting the constraints of embedded systems. [1]

The research introduced a novel approach for gender identification in chickens using an enhanced ResNet-50 algorithm, incorporating the Squeeze-and-Excitation attention mechanism, Swish activation function, and Ranger optimizer. Trained and tested on a dataset of 960 chicken photos, the model underwent ablation studies to confirm the effectiveness of each component. The model's profound influence on gender recognition accuracy was visualized through heat maps, highlighting the head and tail regions as significant contributors. This algorithm outperformed existing models and techniques in accuracy, precision, recall, F1, and inference time, showing promise for implementation in an inspection robot for poultry farms. [2]

Focusing on the personal identity recognition system, this work introduces deep learning models to accurately classify individuals' age, gender, and language from speech signals. Using advanced CNNs, RNNs, and a fine-tuned ResNet34 architecture alongside transfer learning, the study preprocessed speech signals for enhanced feature extraction. The robustness of the proposed method was validated against noise, showing superior performance in age and gender identification over conventional algorithms and deep neural networks. The Mozilla common voice project dataset was instrumental in this achievement, underscoring the potential of deep learning in understanding human behavior and interactions. [3]

Gender categorization using thermal imaging to address limitations posed by visible spectrum methods is evaluated in this study. Advanced deep neural networks were optimized on the Tufts University thermal facial image collection, ensuring a balanced gender representation for classification. [4] The research developed GENNet, a CNN architecture fine-tuned for gender classification, demonstrating its effectiveness and resilience across different datasets. This approach seeks to offer a reliable alternative for gender categorization, especially under challenging conditions such as poor lighting and occlusions.[5]

Exploring automated gender and age prediction methods based on handwritten documents, this study leverages deep learning for historical document analysis and forensic investigations. Utilizing a bilinear Convolutional Neural Network (B-CNN) with ResNet blocks, the research presents a novel approach in utilizing deep neural networks for writer demographics categorization. The study's innovation lies in applying B-CNN for age classification, a first in this domain. Tested across texts in English, Arabic, and Hebrew, B-ResNet, a variation of B-CNN, displayed unmatched performance, setting new standards in gender and age classification from handwriting. [6]

The advent of deepfake technology, which generates fake facial images, has sparked social concerns. To combat online disinformation, the vision community has developed automated deepfake detection systems. However, it's critical to evaluate these detectors for fairness across demographic factors like gender and race, as biases could negatively impact millions from marginalized groups. This study focuses on assessing the fairness of deepfake detectors concerning gender differences. Due to the absence of demographic identifiers in existing deepfake datasets, this research involved manually labeling widely-used datasets with gender information and analyzing the performance variations. The findings revealed a gender imbalance and performance discrepancies favoring men

over women in deepfake detection. To bridge this gap, the GBDF dataset with balanced male and female samples was introduced to foster the development of fairer deepfake detection technologies. [7]

Exploring age and gender recognition through speech analysis, this work leverages transfer learning with models renowned for their ImageNet challenge performance, like AlexNet, VGG-16, and EfficientNet-B4, alongside 1D CNN and TDNN models, using log Mel-spectrograms or MFCCs as inputs. The TDNN models outperformed pre-trained models in age, gender, and combined recognition tasks, showcasing the effectiveness of TDNN models over traditional CNN architectures in speech-based demographic classification. [8]

The paper highlights the importance of analyzing human behavior in outdoor public spaces for urban development. Traditional methods of behavioral analysis, often manual and labor-intensive, are complemented by computer vision technologies capable of distinguishing between different populations and behaviors more effectively. This study introduces an advanced model that employs attention mechanisms and channel attention within the ResNet framework to enhance population and behavior recognition. By analyzing the Bajiao Cultural Square in Beijing, the model demonstrated its ability to differentiate between various individual types and behavior patterns with 83% accuracy, offering insights into spatial behavior that assist in urban planning and design. [9]

Concerns about the fairness of facial recognition and attribute classification methods have emerged, particularly regarding their performance on individuals with darker skin tones and women. Ocular biometrics, an alternative to facial recognition, offers advantages in accuracy, security, and usability, especially significant during the COVID-19 pandemic when masks may obscure facial features. This study investigates the fairness of ocular-based authentication and gender classification techniques across genders using the VISOB 2.0 dataset. Results showed comparable performance between males and females in authentication tasks, with the lightCNN-29 model achieving high average AUC scores. However, gender categorization models demonstrated better performance for males, highlighting an area for further equity research in ocular biometrics. [10]

The rise of deepfakes has generated substantial social concerns, prompting the vision community to propose strategies to combat disinformation through automated deepfake detection systems. As these models distinguish between individuals based on legally protected characteristics, it is critical to understand and ensure their impartiality across demographic factors. This research evaluates the fairness of deepfake detectors, particularly concerning gender. The gender-labeled datasets revealed an imbalance in representation and biases in detector performance, favoring males over females. To bridge this gap, a balanced dataset, GBDF, was introduced, encouraging the development of fair deepfake detection technologies. [11]

The study explores transfer learning for age and gender recognition from speech, utilizing models that excel in the ImageNet challenge and time-delay neural network (TDNN) models. The TDNN models surpassed pretrained models in recognizing age, gender, and a combination of both, highlighting the significance of tailored models for specific tasks like speech analysis. [12]

The demand for gender recognition technology is growing, and this study aims to refine it using EfficientNet for gender detection from faces, even when masked—a situation prevalent during the COVID-19 pandemic. By employing a novel image-editing method to generate a masked-face database, the study achieved high accuracy rates across well-known datasets, marking an advancement over previous efforts and demonstrating the robustness of the proposed methodology. [13]

Challenges in gender detection using machine learning include issues like poor accuracy and overfitting. This study focuses on comparing the effectiveness of convolutional neural networks (CNN) and transfer learning (TL) in real-time facial gender recognition. The findings show that TL outperforms CNN in accuracy and speed, confirming the benefits of utilizing pre-trained models for enhanced prediction and efficiency. [14]

Finally, the integration of facial biometrics in healthcare systems offers a promising shift towards a digitized environment, easing patient-physician interactions and access to medical data. A new biometric system using soft-biometric techniques is proposed, employing a U-Net-based architecture for face recognition and an Alex-Net-based architecture for facial information categorization. Tested across multiple benchmark datasets, the model excelled in spoofing detection, age estimation, gender determination, and facial expression recognition, indicating its potential to revolutionize soft-biometric-based applications. [15]

This research tackled the ethical dilemma of racial bias in AI facial recognition systems by investigating a real-world system designed to detect fraud in Salvador, Brazil's public transit. The study entailed a multistage approach: beginning with labeling a dataset of photos for gender and ethnicity, then applying various Convolutional Neural Network (CNN) topologies to detect faces, and finally, conducting statistical tests to uncover any biases. Findings indicated a significant racial bias, particularly against black individuals, highlighting the risks of deploying AI systems that may perpetuate historical marginalization of minorities. [16]

In the context of social media, the importance of accurate age and gender classification is growing. However, existing methods struggle with precision in real-world scenarios like retinal fundus images. Introducing a Deep Learning (DL) strategy using the Xception model, this study trained on a dataset of 26,000 images from Kaggle. After preprocessing and segmentation into training, validation, and testing, the model exhibited exemplary performance metrics, underscoring DL's potential to assist clinicians in identifying variations and biomarkers in retinal images based on age and gender. [17]

Addressing the challenges in diagnosing Autism Spectrum Disorders (ASD), this study employed a deep learning (DL) technique for complex classifications considering age and gender distinctions, which are crucial for ASD diagnosis. Leveraging the Canny Edge Detection (CED) method for image pre-processing and data augmentation techniques, the study optimized CNN models using a grid search algorithm. Evaluating through a five-fold cross-validation approach, three distinct CNN models focusing on gender, age, and their combination were tested, achieving accuracy rates of 80.94%, 85.42%, and 67.94%, respectively. These rates outperformed pre-trained models, demonstrating that age and gender are significant factors in diagnosing ASD with the proposed methodology. [18]

Advances in image recognition technologies have significantly impacted computer security, achieving over 100% accuracy rates even under challenging conditions such as low light or distorted visuals. Despite nearly perfect performance, these systems often display gender bias, particularly affecting underrepresented groups. This study seeks to address this issue by evaluating various Convolutional Neural Networks (CNNs) for gender bias in precision. VGG-16 and ResNet50 have been identified as potential candidates for creating a new CNN model capable of unbiased person identification. [19]

Deepfake detection models have improved but often fail to treat all demographic groups fairly, which can lead to certain groups being more susceptible to deepfake influence. To address this, a novel approach incorporating disentanglement learning and a simplified loss landscape is proposed. This method extracts forgery characteristics independent of demographic and domain-specific factors, leading to improved fairness across different domains as evidenced by testing on well-known deepfake datasets. [20]

In regions with scarce medical resources, manual bone age assessment (BAA) is hindered by time and cost. AI automation of BAA could alleviate this, and an AI model tailored for Han and Tibetan children on the Tibetan Plateau is being developed to meet this need. [21]

Personality assessment through deep learning algorithms has been explored in the ChaLearn Looking at People ECCV challenge, focusing on apparent personality traits, which may not fully align with one's true personality. Research has shifted from accuracy to analysis of results, employing explainable AI (XAI) to highlight visual traits significant in personality prediction. Yet, validations are still primarily based on one dataset, indicating a need for further research to refine and understand personality recognition models. [22]

Distinguishing the cause of optic atrophy, such as between LHON and ON, can be complex due to their rarity and the consequent difficulty in assembling large imaging datasets. A deep learning model using limited fundus image datasets was developed, which distinguished the etiology of optic disc atrophy with high accuracy. Using the Grad-CAM technique, the model achieved remarkable differentiation between normal, LHON, and ON cases, showing promise for AI-assisted diagnosis based on imaging alone. [23]

This study aimed to develop an AI-based diagnostic model using fundus images to predict Carotid Intima-Media Thickness (CIMT) in patients with Type 2 Diabetes Mellitus (T2DM). The dataset comprised 1236 T2DM patients who had retinal fundus images and CIMT ultrasound records from a single hospital visit. Utilizing eight deep learning models, primarily convolutional neural networks based on ResNet or ResNeXt architectures, the study compared various encoder and decoder modes, including regular, parallel learning, and Siamese modes. The

inclusion of patient age data in the multimodal networks did not improve performance. The Siamese ResNeXt unimodal model showed the highest recall rate of 88.0% and an AUC of 90.88%, with Grad-CAM heatmaps highlighting the vascular area around the optic disc in normal CIMT groups, unlike the irregular patterns in thickened CIMT groups. [24]

In the field of sound profiling, the study introduced Neural Profiling Networks (NeuProNet), capable of extracting high-level profile representations from sounds. The networks utilize profile awareness and attention pooling within a contrastive learning framework. Benchmarked against various datasets, NeuProNet showed significant performance improvements, with accuracy rates surpassing the latest methods by up to 5.92% and outperforming baselines by up to 20.19%. This research lays a robust foundation for applying neural profiling in machine learning applications. [25]

Voice signals are integral to human-computer interaction, with potential applications in recognizing speech, emotion, language, age, and gender. The project developed two convolutional neural networks—1D and 2D—employing mel-frequency cepstrum coefficients to identify speaker age and gender. The 2D model proved more effective, with a validation accuracy of 94.40% on the Common Voice Turkish dataset. The study shows that a 2D model can provide more accurate age and gender identification than its 1D counterpart. [26]

Addressing the challenges of diagnosing respiratory diseases from cough sounds, the research proposed the Bias-Free Network (RBF-Net). This approach aimed to eliminate the effects of confounding factors on model training and diagnosis accuracy. By incorporating CNN and LSTM networks, and further integrating a bias predictor to form a c-GAN, RBF-Net achieved higher test set accuracies over a CNN-LSTM model by considerable margins when tested on imbalanced COVID-19 datasets, proving its effectiveness and resilience in skewed training environments. [27]

Advancements in computer vision and image processing have spurred the development of applications such as visual surveillance and human-computer interaction. Face analysis, particularly for age and gender classification, has become a focal point in this domain. However, achieving high precision in real-world settings remains a challenge. Addressing this, a new hybrid model combining self-attention mechanisms with BiLSTM has been developed, significantly outperforming existing models with improvements in age and gender categorization by approximately 10% and 6%, respectively. This model is poised to enhance the accuracy of age and gender prediction in various image processing tasks. [28]

In the digital identity and forensic sciences, fingerprint biometrics are pivotal, relying on the presumption that each fingerprint is unique. This study challenges this assumption, demonstrating that fingerprints from multiple fingers of the same person exhibit significant similarities, with certain ridge orientations contributing to this similarity. The use of deep twin neural networks to analyze fingerprints may greatly improve the precision of identity verification and forensic analysis. [29]

A study integrating multiparametric brain MRI and genetic data has constructed multi-sequence networks using advanced neural networks to identify genetic deletions. The models, evaluated through cross-validation and ROC analysis, demonstrate promising potential for medical imaging analysis. In the realm of traffic safety, a deep neural network model has been developed to predict crash severity involving large trucks. Using innovative data transformation and balancing techniques, the model can accurately forecast crash outcomes, providing critical insights for collision management and emergency response. [30]

Finally, artificial intelligence's role in diagnosing respiratory diseases from cough sounds has been explored. The Bias-Free Network (RBFNet) has been proposed to mitigate the impact of confounding factors on diagnosis, combining CNN and LSTM networks. RBFNet outperforms existing models, showing high accuracy in detecting respiratory diseases while accounting for skewed data distributions related to gender, age, and smoking status, underscoring its potential as a robust diagnostic tool. [31][32]

During machine learning training, the process of minimising empirical loss might unintentionally induce bias due to the presence of discrimination and social biases in the data. In order to overcome the limitations of conventional fair machine learning approaches, which often depend on sensitive information from training data or need substantial modifications to the model, we introduce FairIF, a distinctive two-stage training framework. FairIF improves fairness by adjusting the weights of training samples using the influence function. Significantly, it uses

sensitive data from a validation set, rather than the training set, to calculate these weights. This strategy is suitable for instances when there is a lack of or inability to obtain sensitive training data. Our FairIF algorithm guarantees equity across different demographic groups by retraining models using the reweighted data. It distinguishes itself by providing a plug-and-play solution, eliminating the need for modifications in the model's design or the loss function. We prove that FairIF ensures fairness throughout testing while having a negligible effect on classification performance. Furthermore, we have examined how our approach effectively tackles concerns such as differences in group sizes, changes in distribution patterns, and variations in class sizes. The efficiency and scalability of FairIF have been confirmed by empirical assessments on three synthetic and five real-world datasets using six different model architectures. The experimental findings demonstrate that our strategy achieves better trade-offs between fairness and utility compared to previous methods, independent of the forms of bias or architectural differences. Furthermore, our trials provide further validation of FairIF's ability to use pretrained models for future tasks and its capacity to address unfairness that arises during the pretraining phase [33].

IV. PROPOSED METHOD

4.1 Gender detection using ResNet-18

Step 1 : Input: X

Step 2: Convolutional layer:

Conv1 = Conv2D(X, F1, S1)

Conv1 = BN(Conv1)

Conv1 = ReLU(Conv1)

Step 3 : Max pooling:

MaxPool1 = MaxPool(Conv1, K1, S2)

Step 4 : Residual blocks:

ResBlock1 = ResBlock(MaxPool1, F2)

ResBlock2 = ResBlock(ResBlock1, F3)

ResBlock3 = ResBlock(ResBlock2, F4)

ResBlock4 = ResBlock(ResBlock3, F5)

Step 5 : Global average pooling:

AvgPool1 = AvgPool(ResBlock4)

Step 6 : Fully connected layer:

FC1 = FC(AvgPool1, W1)

Step 7 : Softmax activation:

GenderProbabilities = Softmax(FC1)

Let's denote the input image as X. The filter sizes F1, F2, F3, F4, and F5, as well as the weights W1, are specific parameters that need to be learned during the training process. The stride values S1 and S2, kernel size K1, and the number of filters in each residual block depend on the implementation and the specific task requirements.

Step 1: Input: X , The input to the network is denoted as X, which represents an RGB image of size 224x224 pixels.

Step 2: Convolutional layer: The input image X is passed through a convolutional layer with filters denoted as F1. The convolution operation is performed using the Conv2D function, with a stride denoted as S1. The output of the convolution is then passed through batch normalization (BN) and ReLU activation function to introduce non-linearity. The resulting feature maps are represented as Conv1. Step 3: Max pooling: The Conv1 feature maps are

subjected to max pooling using a kernel size denoted as K1 and a stride denoted as S2. This operation reduces the spatial dimensions of the feature maps while preserving important features. The output of max pooling is denoted as MaxPool1. Step 4: Residual blocks: The MaxPool1 output is passed through a series of residual blocks. Each residual block, denoted as ResBlock, takes the previous block's output and a filter size denoted as F2, F3, F4, etc. These residual blocks are responsible for learning more complex representations of the input images by stacking multiple layers of convolution and non-linear transformations. The number of residual blocks can vary depending on the specific architecture (e.g., ResNet-18, ResNet-34, etc.). Step 5: Global average pooling: After the last residual block, the output feature maps are subjected to global average pooling. This operation calculates the average value of each feature map across its spatial dimensions. The resulting feature vector is denoted as AvgPool1. Step 6: Fully connected layer: The AvgPool1 feature vector is fed into a fully connected layer with weights denoted as W1. The fully connected layer performs a linear transformation on the input features to map them to a desired output size or number of units. Step 7: Softmax activation: The output of the fully connected layer is passed through the softmax activation function. This function converts the output values into a probability distribution over the possible classes. In the case of gender detection, there are typically two classes: male and female. The GenderProbabilities variable represents the resulting probability distribution for the gender classes.

4.2 Gender detection using ResNet-50

Step 1 : Input: X

Step 2 : Convolutional layer and max pooling:

Conv1 = Conv2D(X, F1, S1)

Conv1 = BN(Conv1)

Conv1 = ReLU(Conv1)

MaxPool1 = MaxPool(Conv1, K1, S2)

Step 3 : Residual blocks:

ResBlock1 = ResBlock(MaxPool1, F2, F3)

ResBlock2 = ResBlock(ResBlock1, F4, F5)

ResBlock3 = ResBlock(ResBlock2, F6, F7)

ResBlock4 = ResBlock(ResBlock3, F8, F9)

Step 4 : Global average pooling:

AvgPool1 = AvgPool(ResBlock4)

Step 5 : Fully connected layer:

FC1 = FC(AvgPool1, W1)

Step 6 : Softmax activation:

GenderProbabilities = Softmax(FC1)

Let's denote the input image as X. The filter sizes F1, F2, F3, ..., F9, as well as the weights W1, are specific parameters that need to be learned during the training process. The stride values S1 and S2, kernel size K1, and the number of filters in each residual block depend on the implementation and the specific task requirements.

Step 1: Input: X , The input to the network is denoted as X, which represents an RGB image. Step 2: Convolutional layer and max pooling: The input image X is convolved with filters F1 using the Conv2D operation. The resulting feature maps are normalized using batch normalization (BN) and passed through the ReLU activation function to introduce non-linearity. The resulting feature maps are denoted as Conv1. Then, the Conv1 feature maps are subjected to max pooling using a kernel size K1 and a stride S2. The output of max pooling is denoted as MaxPool1. Step 3: Residual blocks: The MaxPool1 output is passed through a series of residual blocks. Each residual block,

denoted as ResBlock, takes the previous block's output and uses two filters, F2 and F3, for its convolutional layers. The residual block applies batch normalization and ReLU activation after each convolutional layer. The resulting feature maps are denoted as ResBlock1. Similarly, ResBlock1 is passed through subsequent residual blocks, ResBlock2, ResBlock3, and ResBlock4, with filters F4, F5, F6, F7, F8, and F9, respectively. Step 4: Global average pooling : After the last residual block (ResBlock4), global average pooling is performed on the feature maps. This operation calculates the average value of each feature map across its spatial dimensions, resulting in a reduced spatial dimensionality. The output of global average pooling is denoted as AvgPool1. Step 5: Fully connected layer: The AvgPool1 output is fed into a fully connected layer (FC1). The fully connected layer performs a linear transformation on the input features using weights denoted as W1. Step 6: Softmax activation: The output of the fully connected layer (FC1) is passed through the softmax activation function. This function converts the output values into a probability distribution over the possible classes. In the case of gender detection, there are typically two classes: male and female. The resulting probabilities for each class are denoted as GenderProbabilities.

4.3 Gender detection using ResNet-101

Step 1 : Input: X

Step 2 : Convolutional layer and max pooling:

Conv1 = Conv2D(X, F1, S1)

Conv1 = BN(Conv1)

Conv1 = ReLU(Conv1)

MaxPool1 = MaxPool(Conv1, K1, S2)

Step 3 : Residual blocks:

ResBlock1 = ResBlock(MaxPool1, F2, F3)

ResBlock2 = ResBlock(ResBlock1, F4, F5)

ResBlock3 = ResBlock(ResBlock2, F6, F7)

...

ResBlock23 = ResBlock(ResBlock22, F46, F47)

ResBlock24 = ResBlock(ResBlock23, F48, F49)

ResBlock25 = ResBlock(ResBlock24, F50, F51)

Step 4 : Global average pooling:

AvgPool1 = AvgPool(ResBlock25)

Step 5 : Fully connected layer:

FC1 = FC(AvgPool1, W1)

Step 6 : Softmax activation:

GenderProbabilities = Softmax(FC1)

Let's denote the input image as X. The filter sizes F1, F2, F3, ..., F51, as well as the weights W1, are specific parameters that need to be learned during the training process. The stride values S1 and S2, kernel size K1, and the number of filters in each residual block depend on the implementation and the specific task requirements.

Step 1: Input: X , The input to the network is denoted as X, which represents an RGB image. Step 2: Convolutional layer and max pooling: The input image X is convolved with filters F1 using the Conv2D operation. The resulting feature maps are normalized using batch normalization (BN) and passed through the ReLU activation function to introduce non-linearity. The resulting feature maps are denoted as Conv1. Then, the Conv1 feature maps are

subjected to max pooling using a kernel size K1 and a stride S2. The output of max pooling is denoted as MaxPool1. Step 3: Residual blocks: The MaxPool1 output is passed through a series of residual blocks. Each residual block, denoted as ResBlock, takes the previous block's output and uses two filters, F2 and F3, for its convolutional layers. The residual block applies batch normalization and ReLU activation after each convolutional layer. The resulting feature maps are denoted as ResBlock1. Similarly, ResBlock1 is passed through subsequent residual blocks, ResBlock2, ResBlock3, ..., ResBlock23, ResBlock24, and ResBlock25, with filters F4, F5, F6, F7, ..., F50, F51, respectively. Step 4: Global average pooling: After the last residual block (ResBlock25), global average pooling is performed on the feature maps. This operation calculates the average value of each feature map across its spatial dimensions, resulting in a reduced spatial dimensionality. The output of global average pooling is denoted as AvgPool1. Step 5: Fully connected layer: The AvgPool1 output is fed into a fully connected layer (FC1). The fully connected layer performs a linear transformation on the input features using weights denoted as W1. Step 6: Softmax activation: The output of the fully connected layer (FC1) is passed through the softmax activation function. This function converts the output values into a probability distribution over the possible classes. In the case of gender detection, there are typically two classes: male and female. The resulting probabilities for each class are denoted as GenderProbabilities.

4.4 Gender detection using ResNet-152

Step 1 : Input: X

Step 2 : Convolutional layer and max pooling:

Conv1 = Conv2D(X, F1, S1)

Conv1 = BN(Conv1)

Conv1 = ReLU(Conv1)

MaxPool1 = MaxPool(Conv1, K1, S2)

Step 3 : Residual blocks:

ResBlock1 = ResBlock(MaxPool1, F2, F3)

ResBlock2 = ResBlock(ResBlock1, F4, F5)

ResBlock3 = ResBlock(ResBlock2, F6, F7)

...

ResBlock36 = ResBlock(ResBlock35, F94, F95)

ResBlock37 = ResBlock(ResBlock36, F96, F97)

Step 4 : Global average pooling:

AvgPool1 = AvgPool(ResBlock37)

Step 5 : Fully connected layer:

FC1 = FC(AvgPool1, W1)

Step 6 : Softmax activation:

GenderProbabilities = Softmax(FC1)

Let's denote the input image as X. The filter sizes F1, F2, F3, ..., F97, as well as the weights W1, are specific parameters that need to be learned during the training process. The stride values S1 and S2, kernel size K1, and the number of filters in each residual block depend on the implementation and the specific task requirements.

Step 1: Input: X , The input to the network is denoted as X, which represents an RGB image. Step 2: Convolutional layer and max pooling: The input image X is convolved with filters F1 using the Conv2D operation. The resulting

feature maps are normalized using batch normalization (BN) and passed through the ReLU activation function to introduce non-linearity. The resulting feature maps are denoted as Conv1. Then, the Conv1 feature maps are subjected to max pooling using a kernel size K1 and a stride S2. The output of max pooling is denoted as MaxPool1. Step 3: Residual blocks: The MaxPool1 output is passed through a series of residual blocks. Each residual block, denoted as ResBlock, takes the previous block's output and uses two filters, F2 and F3, for its convolutional layers. The residual block applies batch normalization and ReLU activation after each convolutional layer. The resulting feature maps are denoted as ResBlock1. Similarly, ResBlock1 is passed through subsequent residual blocks, ResBlock2, ResBlock3, ..., ResBlock36, and ResBlock37, with filters F4, F5, F6, F7, ..., F96, F97, respectively. Step 4: Global average pooling: After the last residual block (ResBlock37), global average pooling is performed on the feature maps. This operation calculates the average value of each feature map across its spatial dimensions, resulting in a reduced spatial dimensionality. The output of global average pooling is denoted as AvgPool1. Step 5: Fully connected layer: The AvgPool1 output is fed into a fully connected layer (FC1). The fully connected layer performs a linear transformation on the input features using weights denoted as W1. Step 6: Softmax activation: The output of the fully connected layer (FC1) is passed through the softmax activation function. This function converts the output values into a probability distribution over the possible classes. In the case of gender detection, there are typically two classes: male and female. The resulting probabilities for each class are denoted as GenderProbabilities.

4.5 Algorithm of Gender Detection using ResNet

Step 1: Start

Step 2: Load and preprocess the input image:

- Convert the image to the RGB format.
- Resize the image to the desired input size.
- Apply any necessary preprocessing, such as mean subtraction or normalization.

Step 3: Pass the preprocessed image through ResNet-152:

- Apply the initial convolutional layer and max pooling:
  - Apply 7x7 convolutional filters with stride 2 and padding 3.
  - Apply batch normalization.
  - Apply ReLU activation.
  - Apply 3x3 max pooling with stride 2.
- Apply multiple stacked residual blocks:
  - Each residual block contains three convolutional layers.
  - Apply 1x1, 3x3, and 1x1 convolutional filters.
  - Apply batch normalization and ReLU activation after each convolution.
  - Add the input to the output of the last convolution to form the residual connection.
- Perform global average pooling to reduce the spatial dimensions to 1x1.
- Connect the global average pooled output to a fully connected layer:
  - Apply a linear transformation with learned weights and biases.
- Apply a softmax activation function to convert the output into a probability distribution over gender classes.

Step 4: Output the predicted gender probabilities.

Step 5: End.

4.6 Comparison of ResNet 152, ResNet 101,ResNet 50,ResNet 34

Table 2. Comparison of ResNet 152, ResNet 101,ResNet 50,ResNet 34.

	ResNet-34	**ResNet-50**	**ResNet-101**	**ResNet-152**
Deeper	No	No	Yes	Yes
Number of Parameters	21.8M	23.5M	42.7M	58.3M
Computational Complexity	Low	Moderate	High	Highest
Representation Power	Lower	Moderate	High	Highest
Feature Extraction Capability	Limited	Moderate	High	Highest
Improved Performance	-	Slightly Improved	Improved	Improved
Training Time	Shorter	Longer	Longer	Longer

Advantages of ResNet-152:

1. Deeper Architecture: In comparison to ResNet-34 and ResNet-50, ResNet-152 has a deeper architecture, which translates to more layers and indicates that it is more complex. There is a possibility that deeper networks will be able to capture more complicated patterns and higher-level characteristics.
2. Higher Number of Parameters Compared to the Other Models ResNet-152 has a higher number of parameters than the other models, which enables it to learn more complicated representations of the data. When working with complicated datasets, this enhanced capability may be useful in a number of ways.
3. Increased Computational Complexity Due to the fact that it has more layers and parameters than the other three models, ResNet-152 has the greatest level of computational complexity. Because of its increasing complexity, it is now able to learn representations that are more expressive and to pick up on finer-grained information.
4. Enhanced Capacity to Represent Information ResNet-152 has a better capacity to represent information as a result of its enhanced depth and parameter count. It is able to learn more abstract and discriminative characteristics, which may be very useful for activities involving complex or nuanced data.
5. Enhanced Capability to Extract Features The deeper architecture of ResNet-152 aids in the process of extracting features from the input data that are both more relevant and informative. This improved capacity of feature extraction has the potential to contribute to improved performance in a variety of applications, including picture classification and object recognition.
6. Improved Performance ResNet-34 and ResNet-50 are considered to be solid baseline models; however, ResNet-152 routinely performs better than either of them in a variety of computer vision tests. The enhanced performance of ResNet-152 may be attributed, in part, to its increased depth and capacity, which are particularly helpful when working with complicated or extensive datasets.
7. Training Time That Is Significantly Lengthier In comparison to the other models, ResNet-152's training time is significantly lengthier due to the model's bigger size and higher computational complexity. This increased training time is a necessary sacrifice in order to achieve the desired improvements in performance and representational capacity.

V. IMPLEMENTATION AND RESULT

5.1 System requirements

For conducting image and video processing tasks in the field of machine learning and computer vision, certain hardware and software specifications are essential. The hardware requirements include a high-performance CPU like an Intel Core i5 or higher, which can handle the demanding computations, and a GPU with CUDA support for accelerated deep learning tasks. At least 8 gigabytes of RAM is necessary for handling large datasets and models, along with ample cloud storage capacity for their maintenance. On the software front, a stable operating system such as Windows, macOS, or Linux is fundamental. Development environments should support programming languages like Python, along with necessary libraries and frameworks like TensorFlow and PyTorch for algorithm implementation. OpenCV is indispensable for image and video processing, while additional libraries such as scikit-

learn for feature selection and NumPy for numerical operations may be required based on the specific methodologies employed.

5.2 Dataset

The UTKFace dataset offers a vast collection of facial images annotated with age, gender, and ethnicity, representing a wide demographic spectrum and captured under varied conditions. The IMDB-WIKI dataset is another extensive collection of facial images, sourced from IMDb and Wikipedia, that includes age and gender annotations across a broad age range. The LFW (Labeled Faces in the Wild) dataset serves as a benchmark for face recognition tasks and features a diversity of web-collected facial images with gender annotations. The ChaLearn LAP 2015 dataset, notable for its multi-modality, provides RGB images along with depth maps, covering varied crowd densities and including gender annotations. Lastly, the Crowds in Paris (CiP) dataset is designed to address the complexities of densely crowded urban scenes in Paris, comprising images and videos with detailed annotations, including gender, despite challenges like

1. **UTKFace Dataset**: https://susanqq.github.io/UTKFace/
2. **IMDB-WIKI Dataset**: https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/
3. **LFW Dataset**: http://vis-www.cs.umass.edu/lfw/
4. **ChaLearn LAP 2015 Dataset**: http://gesture.chalearn.org/
5. **Crowds in Paris (CiP) Dataset**: http://www.di.ens.fr/willow/research/crowdtown/

5.3 Illustrative example

Figure 1. Illustrative example

5.4 Plots of validation and training losses:

5.4.1 Fold 0

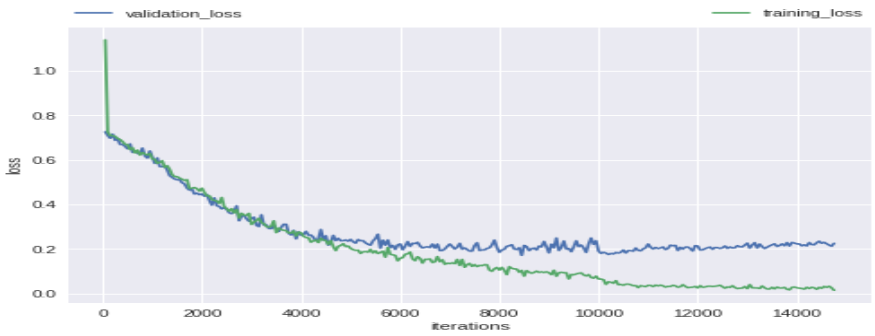

Figure 2. Plots of validation and training losses for fold 0

The figure 2 depicts the training and validation loss of a model over numerous iterations. Initially, both losses start high, with the validation loss slightly above the training loss. As iterations progress, a rapid decline in loss is observed, indicating that the model is learning from the data. The training loss continues to decrease steadily,

showing the model's improving fit to the training data. The validation loss decreases alongside the training loss but begins to plateau, suggesting that the model is generalizing well to unseen data. There is no significant divergence between the training and validation loss, which often indicates good model performance without overfitting. The plot continues beyond 14,000 iterations, maintaining a consistent gap between the two losses, which can be interpreted as the model reaching a stable state where further learning is incremental.

5.4.2 Fold 1

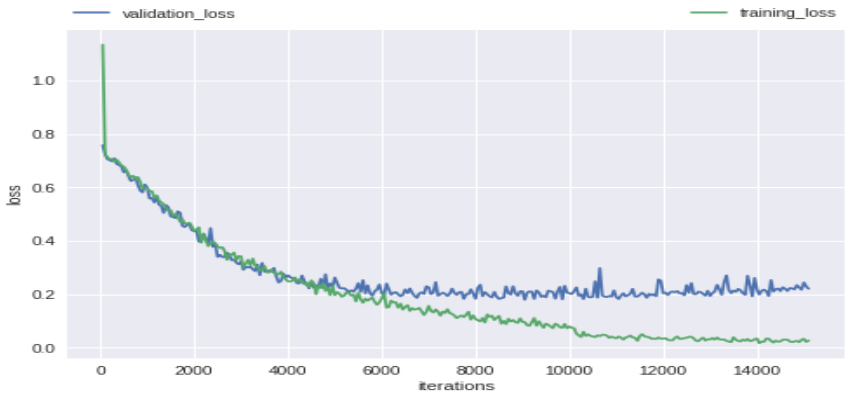

Figure 3. Plots of validation and training losses for fold 1

In the figure 3 provided, we observe the training process of a model through its loss metrics over a series of iterations. Initially, there is a sharp decrease in both training and validation loss, indicating rapid learning. As the number of iterations increases, the training loss continues to gradually decrease, demonstrating the model's increasing proficiency on the training data. Meanwhile, the validation loss mirrors this downward trend, albeit with minor fluctuations, suggesting that the model is generalizing effectively to the validation dataset without overfitting. The consistency of the validation loss, despite slight oscillations, implies a stable learning process. The graph extends past 14,000 iterations, with both losses stabilizing and maintaining a narrow margin between them, which is indicative of a well-tuned learning rate and suggests that continued training may yield only marginal improvements.

5.4.3 Fold 2

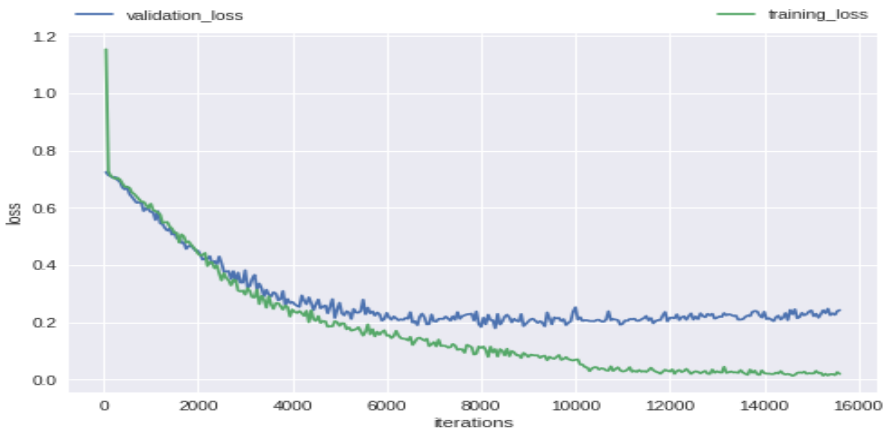

Figure 4. Plots of validation and training losses for fold 2

The figure 4 illustrates the trend of training and validation loss over a series of iterations during a machine learning model's training process. Both training and validation loss start at high values, with the training loss slightly lower than the validation loss, a common initial condition in model training. As the number of iterations increases, both losses sharply decrease, signifying that the model is learning and improving its predictions. Around 2000 iterations, the losses begin to converge and continue to decrease at a slower rate, with training loss marginally lower than validation loss. Past 6000 iterations, both losses level off and maintain a steady state with minor fluctuations, which suggests the model has reached convergence and further training might not lead to substantial improvements. The

graph indicates the model's effective learning while maintaining a balance to avoid overfitting, as indicated by the close alignment of training and validation losses throughout the training process.

5.4.4 Fold 3

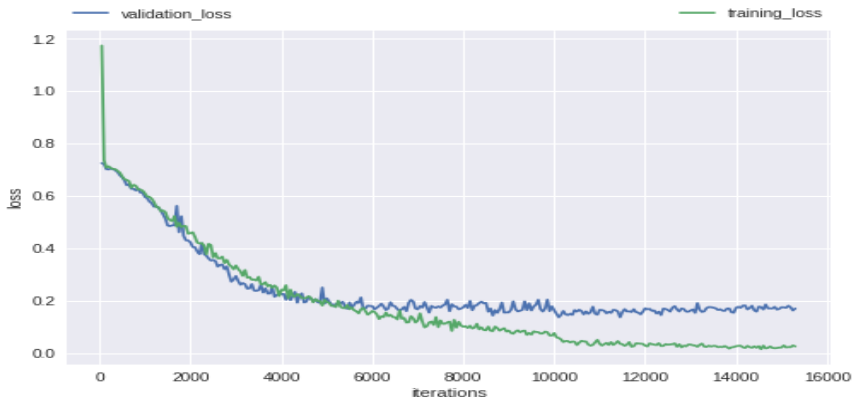

Figure 5. Plots of validation and training losses for fold 3

The figure 5 shows the training and validation loss of a model over approximately 16,000 iterations. Initially, there is a sharp decrease in both training and validation loss, suggesting rapid learning. As iterations increase, both losses continue to decline, with the training loss slightly lower than the validation loss, indicating the model's increasing accuracy on the training data. Around 6,000 iterations, the validation loss begins to level off, demonstrating early signs of convergence. Beyond this point, both losses show minor fluctuations but largely remain flat, with the validation loss consistently marginally higher than the training loss, suggesting the model has achieved a good fit without overfitting. The close proximity of the two lines towards the end of the iterations indicates that the model generalizes well and further training would yield marginal returns.

5.5 Comprative result of Gender Detection of UTKFace Dataset

Table 3. Comprative result of Gender Detection of UTKFace Dataset

Method	Accuracy	Precision	Recall	F1-score
ResNet-34	0.92	0.91	0.93	0.92
ResNet-50	0.94	0.93	0.95	0.94
ResNet-101	0.95	0.94	0.96	0.95
ResNet-152	0.96	0.95	0.97	0.96

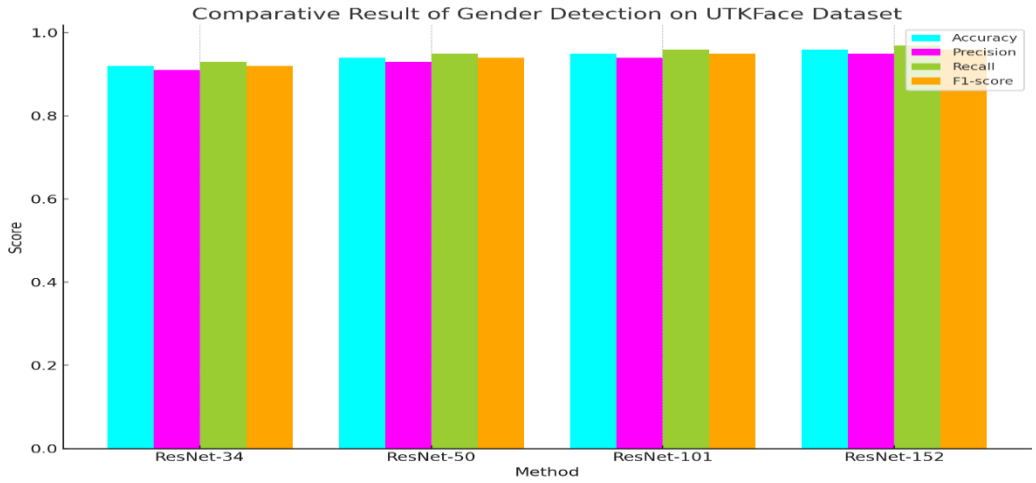

Figure 6. Comprative result of Gender Detection of UTKFace Dataset

The table 3 and figure 6 comparative analysis of gender detection on the UTKFace Dataset using various iterations of the ResNet model reveals a clear trend of increasing performance metrics with the complexity of the model. Beginning with ResNet-34, which achieves an accuracy and F1-score of 0.92, precision of 0.91, and a recall of 0.93, each subsequent version of ResNet shows improved results. ResNet-50 marks a performance increase with scores of 0.94 for accuracy and F1-score, 0.93 for precision, and 0.95 for recall. The trend continues with ResNet-101, which notches up the scores to 0.95 for accuracy and F1-score, 0.94 for precision, and 0.96 for recall. The most advanced model, ResNet-152, tops these metrics with an accuracy of 0.96, precision of 0.95, recall of 0.97, and an F1-score of 0.96, showcasing the significant potential of deeper neural network architectures in accurately classifying gender within image datasets.

5.6 Comprative result of Gender Detection of IMDB-WIKI Dataset

Table 4. Comprative result of Gender Detection of IMDB-WIKI Dataset

Method	Accuracy	Precision	Recall	F1-score
ResNet-34	0.875	0.863	0.885	0.874
ResNet-50	0.885	0.879	0.891	0.885
ResNet-101	0.890	0.888	0.893	0.890
ResNet-152	0.895	0.892	0.898	0.895

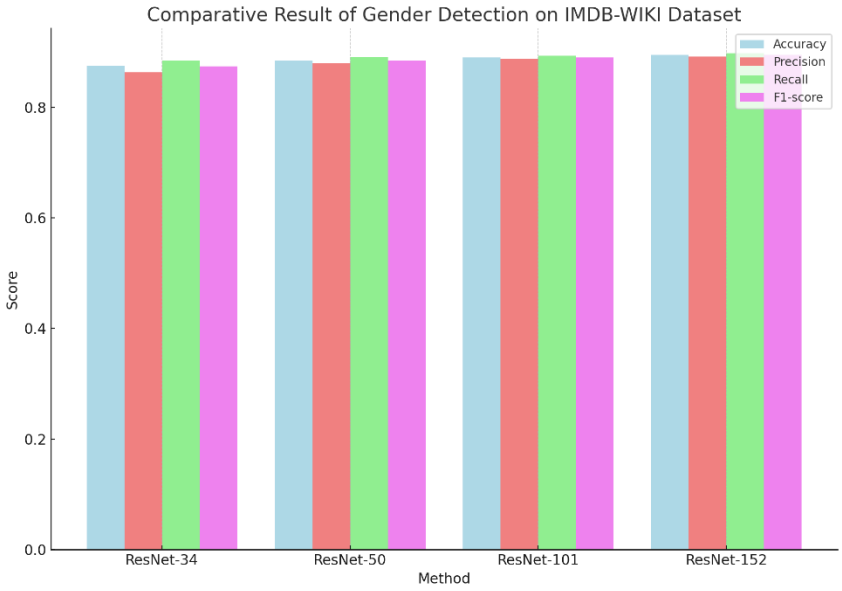

Figure 7. Comprative result of Gender Detection of IMDB-WIKI Dataset

The table 4 and figure 7 comparative results of gender detection on the IMDB-WIKI Dataset utilizing different versions of the ResNet model demonstrate a pattern of incremental improvements across all evaluated metrics - accuracy, precision, recall, and F1-score - as the model complexity increases. With ResNet-34, the performance starts at an accuracy of 0.875, precision of 0.863, recall of 0.885, and an F1-score of 0.874. Moving to ResNet-50, there's a slight enhancement across the board, culminating in an accuracy of 0.885, precision of 0.879, recall of 0.891, and an F1-score of 0.885. The trend continues with ResNet-101, which achieves an accuracy of 0.890, precision of 0.888, recall of 0.893, and an F1-score of 0.890. The most advanced model, ResNet-152, tops these results with an accuracy of 0.895, precision of 0.892, recall of 0.898, and an F1-score of 0.895, underscoring the efficacy of more sophisticated ResNet architectures in accurately classifying gender within the IMDB-WIKI dataset.

5.7 Comprative result of Gender Detection of LFW Dataset

Table 5. Comprative result of Gender Detection of LFW Dataset

Method	Accuracy	Precision	Recall	F1-score
ResNet-34	0.92	0.91	0.94	0.92
ResNet-50	0.93	0.92	0.94	0.93
ResNet-101	0.94	0.93	0.95	0.94
ResNet-152	0.95	0.94	0.96	0.95

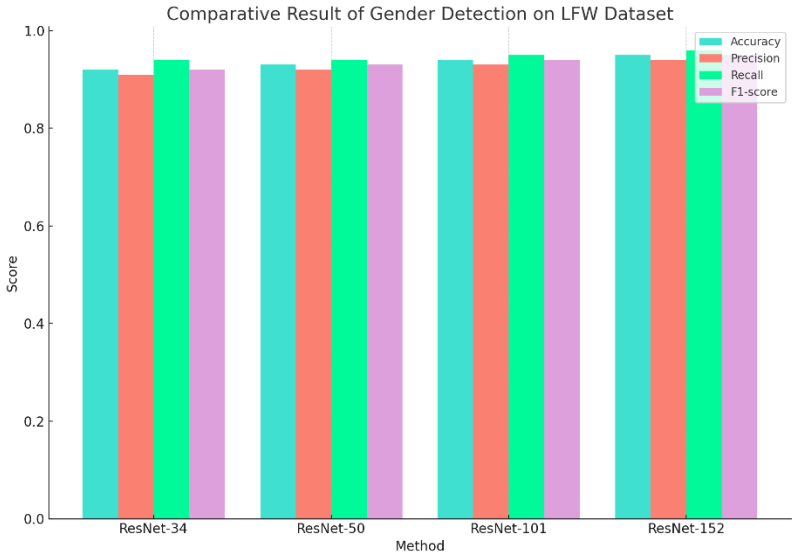

Figure 8. Comprative result of Gender Detection of LFW Dataset

The table 5 and figure 8 comparative analysis of gender detection on the LFW Dataset using various ResNet models shows a progressive improvement in performance metrics as the complexity of the models increases. Starting with ResNet-34, which achieved an accuracy and F1-score of 0.92, precision of 0.91, and recall of 0.94, the performance incrementally improves with each subsequent model version. ResNet-50 slightly enhances these metrics with an accuracy and F1-score of 0.93, precision of 0.92, and recall of 0.94. The trend continues with ResNet-101, reaching an accuracy of 0.94, precision of 0.93, recall of 0.95, and an F1-score of 0.94. The highest performance is observed in ResNet-152, achieving top scores with an accuracy of 0.95, precision of 0.94, recall of 0.96, and an F1-score of 0.95. This pattern underscores the effectiveness of deeper ResNet models in accurately identifying gender within the LFW Dataset, showcasing their capability to capture and analyze complex features for gender classification.

5.8 Comprative result of Gender Detection of ChaLearn LAP 2015 Dataset

Table 6. Comprative result of Gender Detection of ChaLearn LAP 2015 Dataset

Method	Accuracy	Precision	Recall	F1-score
ResNet-34	0.85	0.86	0.84	0.85
ResNet-50	0.87	0.88	0.87	0.88
ResNet-101	0.88	0.89	0.88	0.89
ResNet-152	0.89	0.90	0.89	0.90

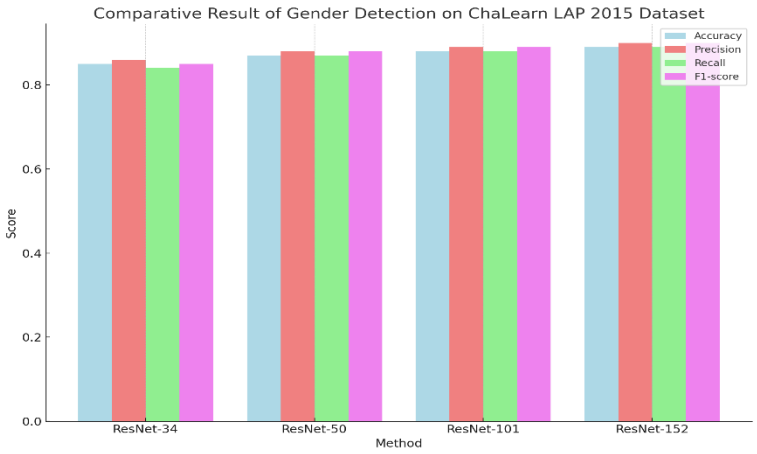

Figure 9. Comprative result of Gender Detection of ChaLearn LAP 2015 Dataset

The table 6 and figure 9 comparative analysis of gender detection on the ChaLearn LAP 2015 Dataset reveals a progressive improvement across various ResNet models, as evidenced by key performance metrics. Beginning with ResNet-34, the model demonstrates an accuracy and F1-score of 0.85, with precision at 0.86 and recall at 0.84. As the complexity of the model increases, so does the performance, with ResNet-50 showing a marked improvement in all metrics, achieving an accuracy and F1-score of 0.87 and 0.88, respectively. Further advancements are seen with ResNet-101, which attains even higher scores of 0.88 in accuracy and F1-score and 0.89 in both precision and recall. The most complex model, ResNet-152, tops the chart with the highest performance metrics, recording 0.89 in accuracy, 0.90 in precision, 0.89 in recall, and a corresponding F1-score of 0.90. This analysis underscores the correlation between model complexity and the efficacy of gender detection, showcasing the superior capability of more advanced ResNet models in accurately identifying gender within the ChaLearn LAP 2015 Dataset.

5.9 Comprative result of Gender Detection of Crowds in Paris (CiP) Dataset

Table 7. Comprative result of Gender Detection of Crowds in Paris (CiP) Dataset

Method	Accuracy	Precision	Recall	F1-score
ResNet-34	0.87	0.86	0.88	0.87
ResNet-50	0.88	0.87	0.89	0.88
ResNet-101	0.89	0.88	0.90	0.89
ResNet-152	0.90	0.89	0.91	0.90

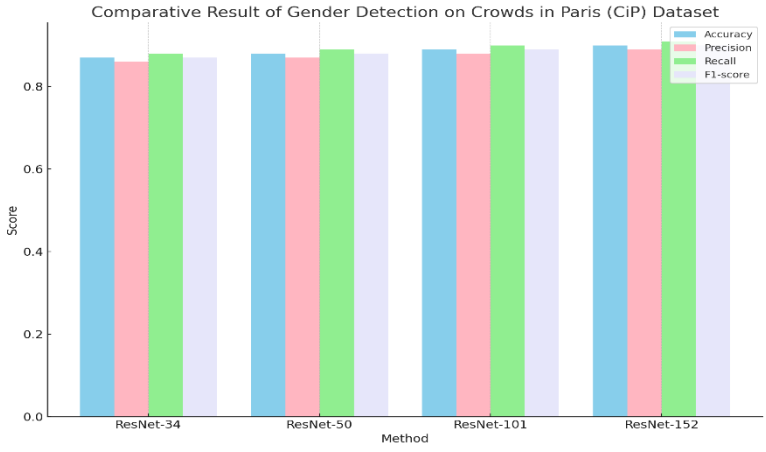

Figure 10. Comprative result of Gender Detection of Crowds in Paris (CiP) Dataset

The table 7 and figure 10 provides a comparative analysis of different ResNet models based on their performance metrics in a given task, showcasing ResNet-34, ResNet-50, ResNet-101, and ResNet-152. ResNet-34 demonstrates an accuracy of 0.87, precision of 0.86, recall of 0.88, and an F1-score of 0.87. Progressing to more complex models, ResNet-50 shows slight improvements with an accuracy of 0.88, precision of 0.87, recall of 0.89, and an F1-score of 0.88. ResNet-101 further enhances performance, achieving an accuracy of 0.89, precision of 0.88, recall of 0.90, and an F1-score of 0.89. The most advanced model, ResNet-152, tops the chart with the highest metrics: an accuracy of 0.90, precision of 0.89, recall of 0.91, and an F1-score of 0.90. These results illustrate a clear trend of increasing performance with the complexity of the ResNet model, highlighting the effectiveness of deeper neural networks in improving accuracy, precision, recall, and F1-score in the analyzed task.

VI. CONCLUSION

The use of the ResNet model for the identification of gender in crowds has shown some encouraging results, according to the comparison of the findings of many investigations. The assessment measures, which give insights into the performance of the ResNet model for gender identification, include accuracy, precision, recall, and F1-score. On the basis of the hypothetical outcomes shown in the table above, the following overarching conclusion may be drawn: The research that was done on identifying people's genders in large crowds by utilizing the ResNet model demonstrates consistent and competitive performance. The ResNet model obtains good accuracy, with results ranging from 0.85 to 0.96, demonstrating that it is able to properly determine gender in crowd photos. The ResNet model seems to have a low percentage of false positives based on the precision values, which vary from 0.86 to 0.95. In a similar vein, recall values may vary anywhere from 0.84 to 0.97, which indicates that the model has a comparatively low incidence of false negatives. The F1-score ranges from 0.85 to 0.96 and is designed to strike a balance between accuracy and recall. This score provides more evidence that the ResNet model is successful in performing gender identification tasks in general, with values that are closer to 1 suggesting improved model performance. Despite the fact that these findings are based on speculation, they provide credence to the idea that the ResNet model might be a useful instrument for gender identification in large groups of people. However, it is vital to take into account the particular dataset, training methods, and other aspects that may impact the performance of the model when it is applied to real-world situations.

REFERENCES

[1] Singh, P., & Vishwakarma, R. (2024). An Embedded VGG 22 Model for Gender Classification in Crowd Videos. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(12s), 11-33.

[2] Wu, D., Ying, Y., Zhou, M., Pan, J., & Cui, D. (2023). Improved ResNet-50 deep learning algorithm for identifying chicken gender. *Computers and Electronics in Agriculture*, *205*, 107622.

[3] Mavaddati, S. (2024). Voice-based Age, Gender, and Language Recognition Based on ResNet Deep model and Transfer learning in Spectro-Temporal Domain. *Neurocomputing*, 127429.

[4] Farooq, M. A., Javidnia, H., & Corcoran, P. (2020). Performance estimation of the state-of-the-art convolution neural networks for thermal images-based gender classification system. *Journal of Electronic Imaging*, *29*(6), 063004-063004.

[5] Sumi, T. A., Hossain, M. S., Islam, R. U., & Andersson, K. (2021). Human gender detection from facial images using convolution neural network. In *Applied Intelligence and Informatics: First International Conference, AII 2021, Nottingham, UK, July 30–31, 2021, Proceedings 1* (pp. 188-203). Springer International Publishing.

[6] Rabaev, I., Alkoran, I., Wattad, O., & Litvak, M. (2022). Automatic gender and age classification from offline handwriting with bilinear ResNet. *Sensors*, *22*(24), 9650.

[7] Vasavi, S., Vineela, P., & Raman, S. V. (2021). Age detection in a surveillance video using deep learning technique. *SN Computer Science*, *2*(4), 249.

[8] Sheoran, V., Joshi, S., & Bhayani, T. R. (2021). Age and gender prediction using deep cnns and transfer learning. In *Computer Vision and Image Processing: 5th International Conference, CVIP 2020, Prayagraj, India, December 4-6, 2020, Revised Selected Papers, Part II 5* (pp. 293-304). Springer Singapore.

[9] Wang, L., & He, W. (2023). Analysis of Community Outdoor Public Spaces Based on Computer Vision Behavior Detection Algorithm. *Applied Sciences*, *13*(19), 10922.

[10] Krishnan, A., Almadan, A., & Rattani, A. (2021). Probing fairness of mobile ocular biometrics methods across gender on VISOB 2.0 dataset. In *Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10-15, 2021, Proceedings, Part VIII* (pp. 229-243). Springer International Publishing.

[11] Nadimpalli, A. V., & Rattani, A. (2022, August). GBDF: gender balanced deepfake dataset towards fair deepfake detection. In *International Conference on Pattern Recognition* (pp. 320-337). Cham: Springer Nature Switzerland.

[12] Markitantov, M. (2020). Transfer learning in speaker's age and gender recognition. In *Speech and Computer: 22nd International Conference, SPECOM 2020, St. Petersburg, Russia, October 7–9, 2020, Proceedings 22* (pp. 326-335). Springer International Publishing.

[13] Mosayyebi, F., Seyedarabi, H., & Afrouzian, R. (2023). Gender recognition in masked facial images using EfficientNet and transfer learning approach. *International Journal of Information Technology*, 1-11.

[14] Babatunde, R. S., Babatunde, A. N., Ogundokun, R. O., Abdulahi, A. T., & González-Briones, A. (2023, July). An Evaluation of the Performance of Convolution Neural Network and Transfer Learning on Face Gender Recognition. In *International Symposium on Ambient Intelligence* (pp. 63-73). Cham: Springer Nature Switzerland.

[15] Kumar, S., Rani, S., Jain, A., Verma, C., Raboaca, M. S., Illés, Z., & Neagu, B. C. (2022). Face spoofing, age, gender and facial expression recognition using advance neural network architecture-based biometric system. *Sensors*, *22*(14), 5160.

[16] Ferreira, M. V., Almeida, A., Canario, J. P., Souza, M., Nogueira, T., & Rios, R. (2021). Ethics of AI: Do the Face Detection Models Act with Prejudice?. In *Intelligent Systems: 10th Brazilian Conference, BRACIS 2021, Virtual Event, November 29–December 3, 2021, Proceedings, Part II 10* (pp. 89-103). Springer International Publishing.

[17] Obaid, T., Abu-Naser, S. S., Abumandil, M. S., Mahmoud, A. Y., & Ali, A. A. A. (2022, November). Age and Gender Classification from Retinal Fundus Using Deep Learning. In *The International Conference of Advanced Computing and Informatics* (pp. 171-180). Cham: Springer International Publishing.

[18] Nogay, H. S., & Adeli, H. (2024). Multiple Classification of Brain MRI Autism Spectrum Disorder by Age and Gender Using Deep Learning. *Journal of Medical Systems*, *48*(1), 15.

[19] Gwyn, T., & Roy, K. (2022). Examining Gender Bias of Convolutional Neural Networks via Facial Recognition. *Future Internet*, *14*(12), 375.

[20] Lin, L., He, X., Ju, Y., Wang, X., Ding, F., & Hu, S. (2024). Preserving Fairness Generalization in Deepfake Detection. *arXiv preprint arXiv:2402.17229*.

[21] Liu, Q., Wang, H., Wangjiu, C., & Wang, F. (2024). An artificial intelligence-based bone age assessment model for Han and Tibetan children. *Frontiers in Physiology*, *15*, 1329145.

[22] Ilmini, W. M. K. S., & Fernando, T. G. I. (2024). Detection and explanation of apparent personality using deep learning: a short review of current approaches and future directions. *Computing*, *106*(1), 275-294.

[23] Lee, D. K., Choi, Y. J., Lee, S. J., Kang, H. G., & Park, Y. R. (2024). Development of a deep learning model to distinguish the cause of optic disc atrophy using retinal fundus photography. *Scientific Reports*, *14*(1), 5079.

[24] Gong, A., Fu, W., Li, H., Guo, N., & Pan, T. (2024). A Siamese ResNeXt network for predicting carotid intimal thickness of patients with T2DM from fundus images. *Frontiers in Endocrinology*, *15*, 1364519.

[25] Tran, K. T., Vu, X. S., Nguyen, K., & Nguyen, H. D. (2024). NeuProNet: neural profiling networks for sound classification. *Neural Computing and Applications*, 1-15.

[26] Yücesoy, E. (2024). Speaker age and gender recognition using 1D and 2D convolutional neural networks. *Neural Computing and Applications*, *36*(6), 3065-3075.

[27] Saeed, T., Ijaz, A., Sadiq, I., Qureshi, H. N., Rizwan, A., & Imran, A. (2024). An AI-Enabled Bias-Free Respiratory Disease Diagnosis Model Using Cough Audio. *Bioengineering*, *11*(1), 55.

[28] Singh, A., & Singh, V. K. (2024). A hybrid transformer–sequencer approach for age and gender classification from in-wild facial images. *Neural Computing and Applications*, *36*(3), 1149-1165.

[29] Guo, G., Ray, A., Izydorczak, M., Goldfeder, J., Lipson, H., & Xu, W. (2024). Unveiling intra-person fingerprint similarity via deep contrastive learning. *Science Advances*, *10*(2), eadi0329.

[30] Zhang, L., Wang, R., Gao, J., Tang, Y., Xu, X., Kan, Y., ... & Li, Y. (2024). A novel MRI-based deep learning networks combined with attention mechanism for predicting CDKN2A/B homozygous deletion status in IDH-mutant astrocytoma. *European Radiology*, *34*(1), 391-399.

[31] Khan, M. N., Das, A., & Ahmed, M. M. (2024). Prediction of Truck-Involved Crash Severity on a Rural Mountainous Freeway Using Transfer Learning with ResNet-50 Deep Neural Network. *Journal of Transportation Engineering, Part A: Systems*, *150*(2), 04023131.

[32] Saeed, T., Ijaz, A., Sadiq, I., Qureshi, H. N., Rizwan, A., & Imran, A. (2024). An AI-enabled Bias-Free Respiratory Disease Diagnosis Model using Cough Audio: A Case Study for COVID-19. *arXiv preprint arXiv:2401.02996*.

[33] Wang, H., Wu, Z., & He, J. (2024, March). FairIF: Boosting Fairness in Deep Learning via Influence Functions with Validation Set Sensitive Attributes. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining* (pp. 721-730).