Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 44 No. 8
August 2023

A MobileNet Model for Identifying Male and Female on Crowd

Priyanka Chauhan and Dr. Rajeev G. Vishwakarma
Department of Computer Science & Engineering,
Dr. A. P. J. Abdul Kalam University, Indore (M. P.)

Abstract -: Efficient real-time models are of paramount importance in several applications within the dynamic domain of computer vision. One such application pertains to the identification of gender in densely populated settings. This study introduces a robust MobileNet model that is especially tailored to accurately classify individuals as either male or female in a highly crowded environment. MobileNet is well recognized for its compact architecture and efficient computational performance, making it highly ideal for on-device applications with constrained resources. The study started with the compilation of an extensive dataset including several crowd scenarios distinguished by diverse levels of density, ethnic compositions, and lighting circumstances. To augment the dataset, several techniques were used, including random cropping, rotation, and horizontal flipping. The augmentation of the model's ability for generalization was accompanied by the mitigation of the problem of overfitting, a phenomenon often seen in specialized domains. The proposed MobileNet model underwent many fine-tuning techniques, focusing particularly on layers that mostly affect spatial features and facial characteristics. By refining these layers, the model demonstrated increased responsiveness to gender-specific cues, such as hair length, facial structure, and attire. The comparative assessment of our MobileNet model with traditional convolutional neural networks revealed comparable levels of accuracy, coupled by a significant decrease in computational burden. The model exhibited a test dataset accuracy of 98.58%, showcasing its exceptional performance relative to other models. Additionally, it showcased a noteworthy decrease of 13% in computational resource use.
Keyword: MobileNet V1 , MobileNet V2, MobileNet V3, Gender , Deep Convolutional Neural Network.

I. Introduction

The exponential growth of visual data in the 21st century may be attributed mostly to the extensive use of digital devices and the surge in online content generation. This trend underscores a conspicuous and undeniable observation: visual media, namely images and videos, are swiftly emerging as the prevailing mode of communication in the digital age. The previously described transformation, although offering many possibilities, also poses a range of complex challenges, notably in the realm of extracting practical knowledge from this abundant flood of visual stimuli [1]. Considerable research focus has been directed on the identification and categorization of gender within densely populated or crowded settings [2]. The intricate task of discerning gender within bustling urban environments, vibrant social gatherings, or densely populated events is not only intellectually intriguing but also holds considerable implications for a range of practical applications, such as targeted marketing, urban planning, safety protocols, and crowd management [3].

Understanding the constraints of traditional methodologies, which mostly concentrate on individual facial identification or isolated subject classification, is crucial in the context of crowded environments. Crowds provide a wide range of problems. The perception of facial features may be impeded to varying degrees by a range of circumstances, including different angles and orientations [4]. As a result, the process of consistently attaining identification becomes challenging. Moreover, variations in lighting conditions might exacerbate the distortion of perceived face features. In addition, the enormous volume of data that has to be examined in real-time makes it challenging to use complex models due to computational limitations, especially in on-field scenarios when there is a lack of sufficient processing resources [5].

MobileNet is an architectural advancement devised by the researchers affiliated with Google. The MobileNet architecture has garnered much recognition as a notable breakthrough in the

domain of on-device visual applications since its first introduction. What sets it apart is its unique amalgamation of a simplistic aesthetic with the guarantee of great accuracy [6]. The architectural design philosophy prioritizes the optimization of the balance between computing resources and performance. This makes it an ideal choice for tasks that need real-time processing capabilities while ensuring the accuracy and reliability of the results.

In light of the given situation, this study aims to adapt the MobileNet model for the specific purpose of identifying individuals' gender inside a crowd. While the main goal is evident, the path is riddled with intricate challenges [7]. The primary obstacle is in the acquisition of data. To ensure the robustness and comprehensiveness of the model, it is essential to use a dataset that spans a broad spectrum of crowd features, such as geographical variety, cultural diversity, and differences in lighting conditions [8]. In addition to the data acquisition procedure, the use of data augmentation strategies is crucial in ensuring the model's ability to generalize well. This enables the model to demonstrate outstanding performance across a diverse array of scenarios, instead than being constrained to certain contexts.

The research delves into the complex process of developing the model, drawing on actual facts to inform each succeeding step. The MobileNet architecture systematically analyzes, assesses, and optimizes every layer, neuron, and parameter with the explicit aim of enhancing its capacity to effectively detect gender cues within intricate and highly filled images [9]. The training of the model encompasses a thorough analysis of many indications that contribute to the discernment of gender. These indicators include a spectrum of factors, including precise facial traits, hair lengths and forms, as well as broader cues offered by clothing choices and posture.

This work will provide readers with a comprehensive examination that spans the initial concept, the successive phases of model development, rigorous testing protocols, and ultimate application in real-world contexts [10]. The primary objective of doing a comparative study is to get valuable insights into the performance of our customized MobileNet model when compared to other advanced architectures. This analysis specifically emphasizes accuracy and computing efficiency as key evaluation metrics.

When outlining this trajectory, our purpose has two distinct components. The primary objective of this work is to provide a thorough and inclusive manual for scholars and professionals who have a keen interest in using MobileNet for similar or equivalent purposes. Moreover, the objective is to promote a broader discourse on the ethical implications, outcomes, and future advancements of gender identification technologies in an increasingly interconnected global landscape [11]. In a contemporary period characterized by the growing convergence of technology in the digital and physical realms, it is of paramount importance to have a comprehensive understanding of the functionalities of models like MobileNet and to use them in an ethical manner. This endeavor transcends ordinary technical endeavors and takes the role of a societal responsibility.

Ii. Background Study

2.1 Historical Context: The historical development of the problem of distinguishing between male and female people within a group may be traced back to the early advancements in computer vision. Initially, researchers mostly focused their efforts on the detection of faces in controlled settings, using manually designed features. Nevertheless, the level of complexity rose considerably when the process of identification shifted from individual individuals to densely crowded scenarios [12]. The existence of large gatherings offered significant obstacles, including occlusions, diverse orientations, and intricate interconnections, which proved to be problematic for standard methods of interpretation.

2.2 Early Detection Mechanisms: In the period before the advent of deep learning, scholars used approaches such as Haar Cascades and Histogram of Oriented Gradients (HOG) to conduct their researches. While Haar Cascades have shown impressive speed and accuracy in recognising faces in controlled environments, their effectiveness tends to decline in complicated and densely populated settings [13]. The use of Histogram of Oriented Gradients (HOG) in combination with classifiers like Support Vector Machines (SVM) has

shown advancements in the domain of gender identification. Nevertheless, these methodologies proved to be inadequate in terms of the requisite degree of robustness and precision required for pragmatic use in real-life situations.

2.3 The Revolution of CNN: The emergence and extensive use of Convolutional Neural Networks (CNNs) marked a pivotal milestone. Deep learning models, such as LeNet, AlexNet, and VGG, has the capability to independently acquire hierarchical features from data [14]. Due to their superior performance relative to traditional methodologies, these techniques were further expanded in their scope to include challenges such as the identification of gender throughout extensive populations. Significant improvements have been made in the performance of attention mechanisms and area proposal operations, allowing efficient processing of several topics inside densely crowded frames.

2.4 Dataset Development: The efficacy of models is considerably impacted by the caliber and diversity of datasets. The first datasets, such as FERET or Yale Face Database, mostly consisted of carefully controlled frontal photographs. However, these datasets have been followed by more complex collections that specifically emphasize situations involving crowds [15]. The recently obtained datasets included the complexities of real-world scenarios, including a wide range of lighting conditions, viewpoints, and interactions. Consequently, the models that underwent training demonstrated improved ability in generalization.

2.5 The Emergence of Mobile Architectures: The growing need for processing tasks to be performed on-device and in real-time has resulted in the heightened acknowledgment and adoption of architectural designs like MobileNet. The algorithms were devised with the objective of striking a balance between computational efficiency and accuracy, making them attractive choices for the task of discerning male and female persons in a crowd [16]. This feature offers notable benefits for applications that need timely answers or are restricted by hardware limitations.

5.6 Ethical and Societal Implications: Alongside the technical challenges, the endeavor of discerning gender among large gatherings has given rise to ethical concerns. The primary focus of attention revolved on issues pertaining to personal privacy, potential biases stemming from unevenly distributed training data in model forecasts, and the broader consequences associated with surveillance and targeted profiling [17]. The need of ensuring fair and responsible use of this technology has become just as crucial as the progress made in its technical elements. 2.7 Ethical and Societal Implications: The domain of gender identification among large groups has significant promise for future breakthroughs. The opportunity for driving innovations and enhancing accuracy lies in integrated models that possess the power to simultaneously detect many demographics or include other modalities like infrared or depth sensing [18]. The present investigation also indicates the potential of few-shot learning, transfer learning, and generative models in enhancing the robustness and generalization capabilities of detection systems.

Iii. Literature Review

The primary objective of this research was to identify and assess gender differences in algorithm performance for the diagnosis of anorexia nervosa on social media postings. We used a set of automated predictors trained on a Spanish data set consisting of 177 users exhibiting signs of anorexia (471,262 tweets) and 326 control cases (910,967 tweets). Initially, we looked at how the algorithms' predictive performance differed for male and female users. After biases were identified, we used a feature-level bias characterisation to determine their origin and compared those aspects to those that are important to physicians. Last but not least, we demonstrated various bias mitigation strategies for creating more equitable automated classifiers, especially for risk assessment in sensitive domains. Our findings revealed worrying predictive performance differences, with significantly higher false negative rates (FNRs) for female samples (FNR=0.082) compared to male samples (FNR=0.005). Male cases were better classified by biological processes and suicide risk factors, whereas female cases were better classified by age, emotions, and personal concerns. Although we were able to demonstrate that inequalities can

be reduced, they cannot be eradicated, we also provided methods for reducing prejudice [19].

In order to better understand the state of Diversity, Equity, and Inclusion (DEI), this study used an online survey directed at members of the Generation Z demographic. We analyzed 675 survey responses and utilized the data to develop an AI-based framework that can identify and evaluate biases based on demographic factors such as ethnicity, socioeconomic status, and gender. The overarching goal of this initiative is to create a digital environment that is more welcoming to people of all backgrounds. In this research, we used a risk assessment model built on top of Natural Language Processing. Word2Vec and a dimensionality reduction technique were used throughout its creation. Our survey results provided the model's training data on the unique linguistic patterns of Generation Z. Importantly, our study demonstrates a paradigm shift in which adjectives formerly associated with one gender are now more often associated with the other. There has been an increase in the use of neutral phrases, such as "strong," and a decrease in the use of gender-specific terms, such as "doll," which may indicate a movement away from old gender norms. In addition, by employing the K-means clustering technique, we have identified a constant presence of social justice issues in debates pertaining to gender, ethnicity, and wealth, demonstrating that members of Generation Z are aware of contemporary racial issues. In light of these results, it is clear that members of Generation Z are acutely aware of inequalities in many areas of society [20].

Demographic research to learn about the likes and dislikes of people in a community is essential to achieving this goal. There is little question that demographic factors like gender play a significant influence in reducing the gender gap seen in a number of academic disciplines. Several state-of-the-art transformer-based models, including BERT and FNET, are compared in terms of their success in classifying participants' gender identities in a cQA setting. The research was able to accomplish its goal because to the massive corpus of 548,375 user profiles employed in the analysis. The data in this corpus was exhaustive, including full questions, answers, and self-descriptions. This allowed for extensive experiments to be conducted, exploring the effects of combining many encoders and sources. Self-descriptions, on average, were shown to have a negative effect due to their lack of specificity, which runs counter to our first intuitive perception. When taking into account all possible questions and answers, the best transformer models (DeBERTa and MobileBERT) obtained an impressive AUC of 0.92. Based on our qualitative research, we may deduce that training on clean corpora influences the process of fine-tuning on user-generated content. We also found that changing the case of specific words may reduce this detrimental effect [21].

The objective of this study was to investigate the prevalence of elevated depression symptoms among teenagers of Asian American, Native Hawaiian, and Pacific Islander descent. We examined a range of social identities, including but not limited to race, ethnicity, sexual orientation, gender identity, and sex assigned at birth. Additionally, we explored instances of bullying that were influenced by factors such as race, immigration status, sexual orientation, gender, and gender expression [22].

This collection opens up several avenues for investigation, such as the comparison of portrayals of men and women in American popular culture throughout time and space. However, separating the postcard collection into male and female categories by hand might take hundreds, if not thousands, of hours of work. As a result, large-scale sociological research would be hampered by the time and effort required for this labor-intensive undertaking. After amassing a large dataset of postcards, we trained deep neural networks, namely YOLOv5x object recognition models, to automatically recognize and categorize people as male or female. We utilized a multi-label stratified 10-fold cross-validation method to ensure that both men and women were represented in the data. This helped us reduce bias and provide more reliable findings. It was shown that this method was successful in classifying people by gender among the library's vast collection of vintage postcards. Between the years 1890 and 1919, our system shown outstanding performance in male identification, with an accuracy rate of 94.9% and a recall rate of 33.0%.

The previous best was 94.7 percent accuracy and 31% recall, both achieved by studies of World War I postcards. The library is able to effectively improve its metadata thanks to the use of advanced deep neural networks, which in turn facilitates broad scholarly inquiries [23].

This study aims to investigate and discover textual biases across demographic groups by using a variety of methodological approaches. Talenya generously contributed the dataset consisting of 14,000 LinkedIn profiles that were analyzed. People who were considered qualified for IT-related roles were represented in these profiles. The primary goal of this research was to establish whether or not there are differences in the ways in which people of various sexes present themselves in written writings. Several methods, including Term Frequency - Inverse Document Frequency (TF-IDF), word2vec, and the Universal Sentence Encoder (USE), were used in the data analysis process. Using the kernel two-sample test, we looked for gender-based differences in the distribution of LinkedIn profiles. Through the use of TF-IDF and cosine similarity measures, we analyzed the prevalence of skill descriptions in LinkedIn profiles. Next, we looked at how often the same information appeared in male and female profiles. Additionally, the evaluations were done on separate candidate groups and focused on gender-specific characteristics. The criteria used to divide the workforce into these subcategories were job function, proximity, and organizational standing. In this last section, we will discuss the theoretical and practical ramifications of our study [24].

Facial gender identification has attracted a lot of attention from academics because to the wide variety of uses it has, such as gender-based demographic analysis, targeted advertising, security, and visitor profiling. Several methods, including Adaboost, Integral Image, Haar features, and the Cascade Classifier, are used into the suggested technique for face detection. The LFW dataset (including 13,233 facial pictures) was utilized for this study [25].

The study's overarching goal was to determine whether there was a statistically significant difference in the effectiveness of psychosocial therapy for reducing delinquent behavior among adolescents involved in the justice system based on gender. A systematic meta-analysis was conducted to examine the role of gender in the outcomes of several therapies for this purpose. Ten randomized controlled trials were analyzed for this research to allow for a measurement of impact sizes according to gender. This study's results imply that the interventions used to reduce juvenile delinquency did not work. Overall, we discovered an impact size of d = -0.006 (p =.921). Recidivism rates did not differ significantly by gender (Q =.071, p =.790), another interesting finding. Neither males (d = 0.006, p =.933) nor females (d = -.027, p =.785) responded significantly to the treatments [26].

The facial muscles on one side of the face are paralyzed due to a neurological condition called facial palsy (FP), which affects the seventh cranial nerve. This disease, which may be bothersome at times, affects people of all sexes and all ages. Traditional visual diagnostic methods, which focus mostly on assessing facial asymmetry, could have some room for error. As a result, researchers have paid a lot of attention to how they may implement AI into computer vision for the purpose of fingerprint recognition. Advantages in terms of time, labor, and resource efficiency are provided by deep learning algorithms due to their ability to effectively identify false positives (FP) in real-time. We provide a technique that can properly assess the patient's age and gender in addition to diagnosing facial paralysis in real time. Using a Raspberry Pi device, a digital camera, and a deep learning model, the suggested method improves the diagnosis process for both medical professionals and patients. The suggested technique may be easily included into established medical evaluations. The algorithm achieved a remarkable 98% accuracy utilizing a sample of 20,600 photographs, of which 19,000 were of a typical kind and 1,600 were false positives. Thus, the setup we provide is a model for a reliable and efficient diagnostic approach for identifying false positives [27].

Previous studies have hypothesized that there may be innate differences between men and women in their tendency to participate in cooperative conduct when faced with situations with competing interests. New empirical research,

however, suggests that men and women cooperate equally, with some exceptions due to cultural or environmental factors. Using data from 20 industrialized nations spanning six decades (1961-2017), this study sought to do a thorough analysis of the vast body of research on human cooperation, with an emphasis on societal concerns. Our research set out to assess the merits of hypotheses grounded on evolutionary and social role theories. Surprisingly, our thorough investigation turned up very no correlation between gender and teamwork. From what we can tell, no significant differences in teamwork were found between the sexes (effect size = 0.011, 95% confidence interval [0.038, 0.060]). It's worth noting that studies with mostly female participants have shown somewhat increased levels of teamwork. Our research also did not back up our initial prediction that gender gaps in cooperation would be more glaring in circumstances with more intense conflicting interests or in civilizations with different levels of gender equality and economic development. These results provide a new angle on the study of gender roles in team settings [28].

The complex problem of determining demographic information from a single, static facial image is investigated here. Models are swapped out depending on the job at hand, and the performance of VGG16, ResNet50, and SE-ResNet50, three pre-trained convolutional neural network architectures, is evaluated. The VGGFace2 dataset was utilized to train the neural networks employed in this research. Our study not only details the most effective methods for feature extraction using machine learning, but also provides a thorough assessment of the effectiveness of various techniques. It's important to remember that even simple models like linear regression may outperform expectations with the right training data. Convolutional Neural Networks (CNNs) were trained from scratch in the context of age prediction [29].

When comparing researchers, the quality of their citations is frequently given more weight than the number of their publications. There is a clear gender gap in academic publication, but there is surprisingly little information concerning how often women are cited in comparison to men. The researchers wanted to find out whether there was a difference in citation rates based on gender. The Web of Science database was queried for the number of times research articles and reviews in 14 general medical journals with an impact factor of > 5 were cited each year between January 2015 and December 2019. The Gender API was used to determine the gender of the publication's lead and final authors. Multivariate negative binomial regressions [30] were used to make comparisons between the sexes of the study participants.

DALL-E 2 and other AI models have impressive generative capabilities, allowing them to translate verbal inputs into complex images that are convincing imitations of human creativity. Despite the widespread praise these models have received, there has been surprisingly little investigation into whether or not they perpetuate harmful gender stereotypes in the images they produce. Our research sought to fill this void by analyzing 15,300 images representing 153 different occupations created by DALL-E 2 to determine the extent to which gender bias was present. To identify possible bias amplification, we used the 2021 census workforce figures and Google Images as reference points in our research. Specifically, the research found that DALL-E 2 tends to underrepresent women in traditionally male-dominated fields while overrepresenting them in traditionally female-dominated fields. Moreover, the majority of women, rather than men, are shown smiling and having their heads down in the images made, especially when they are doing work that is stereotypically associated with women. DALL-E 2's biases in representation and presentation are more obvious when compared to those of Google Images. This study emphasizes the immediate need for measures to curb the spread of preexisting societal biases via artificially created images [31].

The goal of this study was to develop a method for automatically identifying a person's gender using just their voice and a pitch detection (PD) extractor. In this work, voice signals from people with Parkinson's disease (PD) were analyzed using the Yet Another Algorithm for Pitch Tracking (YAAPT) to precisely identify the fundamental frequency (F0) of their speech. This approach facilitates the classification of vocalizations into adult and juvenile groups by allowing for the

differentiation of fundamental frequency (F0) changes between adult males and females. A 1D convolutional neural network (CNN) was utilized to categorize vowel voice sounds. Multiple 1D kernel convolution layers, a 1D pooling operation, and a vowel classifier make up the network. In this configuration, feature patterns are initially segmented by classifying them into one of three F0 ranges that account for both adults and children. The speaker's gender is identified by a subsequent classifier layer. The purpose of our proposed PD extractor is to maximize efficiency and boost classification accuracy when used in tandem with the voice classifier. Using acoustic datasets from the Hillenbrand database, the method was evaluated on 12 vowel categories. We used a k-fold cross-validation method. Results demonstrated notable performance in terms of recall, precision, accuracy, and F1 score assessments [32], underscoring the usefulness of our methods.

Table 1. Systematic literature review

Reference	Author(s)	Method	Result	Limitation	Future Scope
19	Solans Noguero et al.	Data-driven study on AI assessment of anorexia nervosa on social media	Identified potential gender bias in AI analysis of anorexia-related content	Limited scope to anorexia, potential algorithmic bias	Extend analysis to other mental health issues on social media
20	Low et al.	Risk-informed AI-based bias detection	Detected bias related to gender, race,	Survey data limitations, potential sample bias	Apply method to broader population and
		using Gen-Z survey data	and income		additional biases
21 & 22	Schwarzenberg & Figueroa	Developed textual pre-trained models for gender identification in community Q&A platforms	Enhanced gender identification accuracy in textual data	Limited to text, potential bias in training data	Expand model's application to different platforms and contexts
22	Rider et al.	Explored bias-based bullying and depressive symptoms in sexual and gender-diverse Asian American, Native Hawaiian, and	Found elevated depressive symptoms among bullied adolescents	Limited to specific demographic, potential self-report bias	Investigate interventions to mitigate mental health impacts

		Pacific Islander adolescents			
23	Schuerkamp et al.	Automatic gender recognition in historical postcards using deep learning	Successfully recognized genders in historical postcards	Limited to postcards, potential historical bias	Apply the approach to other historical visual data
24	Simon et al.	Identified gender bias in LinkedIn profiles using data-driven methods	Found instances of gender bias in LinkedIn profiles	Limited to LinkedIn, potential bias in data sampling	Investigate bias in other professional networking platforms
25	Hassan & Dawood	Used Viola-Jones algorithm for facial image-based gender recog	Achieved gender recognition from facial images	Limited to facial images, potential algorithmic bias	Enhance accuracy across diverse facial images
		nition			
26	Galbraith & Huey	Preliminary meta-analysis on gender differences in intervention effects on delinquency	Identified gender differences in delinquency interventions	Limited to delinquency, preliminary analysis	Extend analysis to different interventions and settings
27	Amsalam et al.	Developed Raspberry Pi-based system for automatic facial palsy, age, and gender detection	Achieved successful detection of facial palsy, age, and gender	Limited to facial analysis, hardware constraints	Improve accuracy and expand use cases
28	Spadaro et al.	Meta-analysis on gender differences in cooperation across societ	Identified gender differences in cooperative behavior	Limited to cooperation, potential cultural bias	Explore effects of socio-cultural factors

		ies			
29	Chowdary et al.	Used CNN for age and gender detection in manipulated images	Developed method for age and gender detection	Limited to manipulated images, potential training data bias	Extend CNN to various image manipulation scenarios
30	Sebo & Clair	Investigated gender inequalities in citations in medical journals	Found gender inequalities in citations	Limited to medical journals, potential data bias	Study mechanisms underlying citation disparities
31	Sun et al.	Audited gender biases in image generative AI	Highlighted gender biases in image generation	Limited to image AI, potential subjectivity	Develop techniques to mitigate AI-generated biases
32	Lin et al.	Developed vowel classifier for gender identification	Achieved gender identification through vowel	Limited to vowel-based identification, audio noise	Apply to broader speech-related gender
		using CNN	analysis		identification

Iv. Proposed Method

4.1 Gender detection using MobileNet V1

Gender Detection using MobileNet V1

1. Data Preparation:

1.1. Dataset Splitting:

- Split your dataset (containing labeled images) into training, validation, and test subsets.

1.2. Image Normalization: For each image in these subsets:

arduinoCopy code

\(\text{normalized image} = \frac{\text{image} - 127.5}{127.5} \)

1.3. Image Resizing:

- Resize each image to fit MobileNet V1's input size, e.g., 224x224 pixels.

2. Model Definition:

2.1. Base Model:

- Load the MobileNet V1 architecture (without its top/classification layers).

2.2. Append Layers:

- Add a Global Average Pooling (GAP) layer.
- Append a fully connected (Dense) layer with 1 neuron, using the sigmoid activation function:

$\sigma(z) = \frac{1}{1+e^{-z}}$

where *z* is the input to the neuron.

3. Model Compilation:

3.1. Loss Function:

- Use Binary Cross-Entropy Loss:

$L(y, \hat{y}) = -[ylog(\hat{y}) + (1-y)\log(1-\hat{y})]$

where:

- *y* is the true label.
- $\hat{y}$ is the predicted label.

3.2. Optimizer:

- Choose an optimizer, e.g., Adam.

3.3. Evaluation Metric:

- Set accuracy as a metric.

4. Training:

4.1. Input Data:

- Feed the training data in batches to the model.

4.2. Validation:

- After each epoch, validate the model using the validation dataset.

4.3. Best Model:

- Save the model that yields the highest validation accuracy.

5. Testing & Evaluation:

5.1. Model Loading:

- Retrieve the best model weights.

5.2. Test:

- Evaluate the model using the test dataset. Calculate the accuracy:

Accuracy=Number of correct predictionsTotal pred ictionsAccuracy=Total predictionsNumber of cor rect predictions

6. Deployment (Optional):

6.1. Conversion:

- Convert the trained model to a mobile-friendly format, like TensorFlow Lite.

6.2. Integration:

- Embed the model into a mobile or edge application.

7. Inference:

7.1. Image Pre-processing:

- For any new image, resize and normalize it.

7.2. Model Feeding:

- Pass the processed image through the model.

7.3. Result Interpretation:

- If the output is ≥ 0.5, classify as 'Female', otherwise classify as 'Male'.

Here's a detailed explanation of MobileNetV1's architecture:

1. Depthwise Separable Convolution: The core idea behind MobileNetV1 is the use of depthwise separable convolutions, which significantly reduces the number of parameters and computations compared to standard convolutions. This kind of convolution is split into two layers:

1.1. Depthwise Convolution: It applies a single convolutional filter for each input channel. Instead of combining the channels, each channel remains separate.

1.2. Pointwise Convolution: It's a 1x1 convolution responsible for combining the outputs of the depthwise convolution, thereby increasing (or decreasing) the number of channels.

2. Architecture Details: MobileNetV1 uses a standard convolution at the beginning and then utilizes depthwise separable convolutions throughout the rest of the network.

The general structure can be summarized as:

- Initial convolution with 32 filters of size 3x3.
- 13 depthwise separable convolution blocks.
- Each block comprises a depthwise convolution followed by a 1x1 pointwise convolution.

The number of filters and the stride can vary between blocks, depending on the specific variant of MobileNetV1 you're considering (e.g., the width multiplier or the resolution multiplier, which are explained next).

3. Width Multiplier (α): This is a hyperparameter introduced to thin the network or adjust the number of channels. Given a baseline architecture, the number of channels for all layers in the network is multiplied by α.

- It's a value between 0 and 1.
- A smaller α reduces the computational cost.
- It proportionally reduces the number of filters in each layer.

4. Resolution Multiplier (ρ): It's used to create a reduced representation of the input image by adjusting the network's input resolution. The spatial dimensions of all layers are multiplied by ρ.

- It's a value between 0 and 1.
- A smaller ρ reduces the computational cost.
- It proportionally reduces the size of the input image.

5. Fully Connected Layer & Softmax: At the end of these blocks:

- Global Average Pooling (GAP) is applied to the feature map, ensuring that the output is of shape 1x1xC where C is the number of channels.
- A fully connected layer is then attached to predict the classes (in the original MobileNetV1 for ImageNet, there were 1000 classes).
- A softmax activation function is used at the end to get probability distributions for the classes.

6. Activation Function: MobileNetV1 uses the ReLU6 activation function. ReLU6 is the standard ReLU but clipped to a maximum value of 6.

Advantages:

- Efficiency: Uses substantially fewer parameters and computations compared to other architectures like VGG or ResNet.
- Versatility: By tuning the width multiplier (α) and the resolution multiplier (ρ), you can get different variants of the network, suited for various applications and constraints.

4.2 Gender detection using MobileNet V2

1. Inverted Residuals and Linear Bottlenecks:

The main innovation in MobileNetV2 is the use of inverted residuals with linear bottlenecks. This structure departs from traditional network designs in two main ways:

1.1. Expansion Layer:

- Input is first upsampled using 1x1 convolutions (called expansion layer) to a higher dimension. This is the "inversion" part since traditional bottlenecks typically reduce dimensions first.

1.2. Depthwise Convolution with Bottleneck:

- A depthwise convolution is then applied, followed by a 1x1 convolution (called projection layer) to reduce the dimensionality, but without applying any non-linearities (hence, "linear bottleneck").

2. Strided Depthwise Convolutions:

For down-sampling, MobileNetV2 uses strided convolutions in the first 1x1 convolution of the block (during the expansion phase) rather than in the 3x3 depthwise convolution.

3. Linear Bottlenecks:

Non-linearities are removed in the narrow layers of the network, allowing the model to retain more information. In the context of MobileNetV2, these narrow layers are the outputs of the residual blocks. The network only applies the ReLU6 activation function to the depthwise convolutions and the expansion layer.

4. ReLU6 Activation:

MobileNetV2, like its predecessor, uses the ReLU6 activation:

ReLU6(x)=min(max(0,x),6)

This bounded activation function is believed to be more robust against low-precision computations.

5. Network Structure:

MobileNetV2's structure is made of:

- A standard initial 32-filter convolutional layer.
- 17 residual bottleneck sequences of varying filter sizes and strides.
- A final 1x1 convolution, followed by average pooling and a softmax layer.

6. Width Multiplier & Resolution Multiplier:

Just like MobileNetV1, V2 also incorporates:

- Width Multiplier (α): Adjusts the number of channels in each layer.
- Resolution Multiplier (ρ): Controls the input resolution, and consequently, the resolution at every layer of the network.

7. Final Layers:

Post the inverted residual blocks:

- A 1x1 convolution to increase the feature depth.
- Global Average Pooling (GAP) to reduce spatial dimensions.
- Fully connected layer for classification, followed by softmax for probability distribution.

Advantages of MobileNetV2:

- Efficiency: Offers better accuracy with comparable computational costs compared to MobileNetV1.
- Modularity: The inverted residual blocks can be easily stacked and modified to create variations of the network.

4.3 Gender detection using MobileNet V3

1. Data Preparation:

1.1. Dataset Splitting:

- Split your dataset (with labeled images of male and female faces) into training, validation, and test subsets.

1.2. Image Normalization: For each image:

arduinoCopy code

\(\text{normalized image} = \frac{\text{image} - 127.5}{127.5} \)

1.3. Image Resizing:

- Adjust each image to the input size that MobileNet V3 needs, e.g., 224x224 pixels.

2. Model Definition:

2.1. Base Model:

- Load the MobileNetV3 architecture without the top (classification) layers.

2.2. Append Layers:

- Attach h-swish or h-sigmoid activation functions, which are unique to MobileNetV3.
- Add a Global Average Pooling (GAP) layer.
- Append a Dense layer with a single neuron and the sigmoid activation function:
- $\sigma(z) = \frac{1}{1+e^{-z}}$

where z represents the input to the neuron.

3. Model Compilation:

3.1. Loss Function:

- Binary Cross-Entropy Loss:
- $L(y, \hat{y}) = -[ylog(\hat{y}) + (1-y)\log(1-\hat{y})]$

where:

- y is the actual label.

- $\hat{y}$is the model's predicted label.

3.2. Optimizer:

- Opt for an optimizer, e.g., Adam.

3.3. Evaluation Metric:

- Set accuracy as a metric.

4. Training:

4.1. Input Data:

- Provide the training data in batches to the model.

4.2. Validation:

- After every epoch, validate the model using the validation dataset.

4.3. Checkpointing:

- Save the model that achieves the highest validation accuracy.

5. Testing & Evaluation:

5.1. Model Retrieval:

- Load the best model weights.

5.2. Testing:

- Evaluate the model on the test dataset to compute accuracy:

Accuracy=Number of correct predictionsTotal pred ictionsAccuracy=Total predictionsNumber of cor rect predictions

4.4 MobileNetV3 Architecture for Gender Detection:

1. Base Model:

1.1. First Layer:

- A standard initial convolutional layer, which often has 16 filters, with a kernel size of 3x3.

1.2. Main Architecture Blocks:

- MobileNetV3 is structured with a series of bottleneck blocks. These blocks have a combination of:
- Expansion layer: 1x1 convolution to increase the number of channels.
- Depthwise convolution: 3x3 or 5x5, which applies convolution separately per channel.
- Squeeze-and-Excitation (SE) modules: These are used to adaptively recalibrate channel-wise feature responses.
- Projection layer: 1x1 convolution to project features back to a lower dimension.

1.3. Activation Functions:

- MobileNetV3 introduces two new activation functions:
- h-swish: A variant of the swish function that's computationally more efficient.
- h-sigmoid: A hard version of the sigmoid function.

These activations help in gaining slight improvements in accuracy while remaining computationally efficient.

1.4. Diverse Bottleneck Blocks:

- MobileNetV3 contains a variety of bottleneck blocks with different configurations. For instance, it will have blocks with:
- Different kernel sizes (e.g., 3x3 or 5x5).
- With or without the SE layer.
- Different expansion factors (e.g., expansion of input channels by a factor of 3, 4, or 6).

1.5. Downsampling:

- MobileNetV3 achieves downsampling by adjusting the stride in certain blocks.

2. Adaptation for Gender Detection:

2.1. Remove Last Fully Connected Layer:

- In the base MobileNetV3, the last fully connected layer (or dense layer) typically caters to a large number of classes (e.g., 1000 for ImageNet classification). You'll remove this to adapt the model.

2.2. Add Custom Layers for Gender Detection:

- Global Average Pooling (GAP): Reduces spatial dimensions.
- Fully Connected (Dense) Layer: Contains 1 neuron with a sigmoid activation function to give a binary output, which corresponds to male or female.

3. Training and Optimization:

3.1. Binary Cross-Entropy Loss:

- As this is a binary classification task, use the binary cross-entropy loss function for optimization.

3.2. Optimizer:

- Common choices include Adam, RMSprop, or SGD with momentum.

4.5 Descriptive Layout for MobileNetV3 Architecture (Gender Detection):

1. Input Layer:

- Size: e.g., 224x224x3 (height x width x channels)

2. Initial Convolution:

- Filters: 16
- Kernel size: 3x3
- Activation: h-swish or h-sigmoid

3. Bottleneck Blocks:

(Each block consists of an expansion layer, depthwise convolution, SE module, and projection layer.)

- Block 1:
- Expansion: e.g., 64 filters, 1x1
- Depthwise Conv: 3x3
- SE module
- Projection: reduce to, e.g., 24 filters, 1x1
- Activation: h-swish or h-sigmoid

... Continue similarly for other blocks ...

4. Global Average Pooling (GAP) Layer:
- This layer will reduce the spatial dimensions to 1x1.
5. Dense (Fully Connected) Layer:
- Neurons: 1 (for gender classification)
- Activation: Sigmoid
6. Output:
- Size: 1 (scalar output representing the gender prediction probability)

To visually depict this architecture, you'd typically represent each layer or block with rectangles and provide labels for their configurations (e.g., 3x3, h-swish, SE module).

4.6Comparison of MobileNet V1, MobileNet V2,MobileNet V3

Table 2. Comparison of MobileNet V1, MobileNet V2, MobileNet V3.

Parameter	MobileNet V1	MobileNet V2	MobileNet V3
Year	2017	2018	2019
Type	Single-pass CNN	Inverted Residuals	Efficient Backbone
Depth Multiplier	Yes	Yes	Yes
Width Multiplier	Yes	Yes	Yes
Input Size	Typically 224x224	Typically 224x224	Typically 224x224
Architecture	Standard Conv Blocks	Inverted Res Blocks	Mix of Blocks
Depthwise Separable Conv	Yes	Yes	Yes
Bottleneck Design	No	Yes	Yes
Number of Blocks	Fewer	More	More
Computational Complexity	Lower	Higher	Higher
Number of Parameters	Lower	Higher	Varies
Model Variants	N/A	MobileNet V2, MobileNet V2-FPN	MobileNet V3-Large, MobileNet V3-Small
Performance Improvements	Basic Lightweight Model	Better Efficiency	Enhanced Efficiency
Accuracy (ImageNet)	Good	Good	Good
Specialized Versions	N/A	MobileNet V2-SSDLite, MobileNet V2-FastRCNN	MobileNet V3-SSDLite, MobileNet V3-Large-Minimal
Architecture	Depthwise Separable Convolutions	Inverted Residuals with Linear Bottlenecks	Inverted Residuals + Squeeze-and-Excitation
Activation Function	ReLU6	ReLU6	h-swish, h-sigmoid
Bottleneck Design	No	Yes (with expansion layer)	Yes (with expansion layer)
Downsampling	Strided Depthwise Convolution	Stride in the 1x1 convolution	Varies, but often in expanded convolution
Special Modules	None	None	Squeeze-and-Excitation (SE) Modules
Depth Multiplier	Yes (Width Multiplier)	Yes (Width Multiplier)	Yes (Width Multiplier)

Resolution Multiplier	Yes	Yes	Yes
Key Advancement	Depthwise separable convolutions for efficient computing	Inverted residuals for better utilization of model parameters	Incorporates elements from Neural Architecture Search (NAS) for optimal structure

Advantages of MobileNet V3:

MobileNetV3 brings several improvements and advantages over its predecessors (MobileNetV1 and MobileNetV2). These advancements are the results of the incorporation of architecture search methodologies and manual optimizations. Here are the primary advantages of MobileNetV3 in comparison to earlier versions:

1. **Optimized Architecture through NAS (Neural Architecture Search):** MobileNetV3 incorporates elements from the search space of Neural Architecture Search, leading to an architecture that's more optimized for performance on specific tasks and hardware targets.

2. New Activation Functions:

- h-swish: A computationally efficient version of the swish activation function. It allows the network to achieve slightly better performance than ReLU-based counterparts without increasing computational complexity.
- h-sigmoid: An efficient version of the sigmoid function. Both these activations are introduced to improve model accuracy while maintaining efficiency.

3. **Squeeze-and-Excitation (SE) Modules:** These modules, integrated into the MobileNetV3 architecture, adaptively recalibrate channel-wise feature responses, leading to improved representational efficiency.

4. **Better Balance between Latency & Accuracy:** The design of MobileNetV3 targets not only accuracy but also latency. This makes the model more suitable for real-time applications on edge devices.

5. **Efficient Edge Use-Cases:** MobileNetV3 has specific variants optimized for different tasks, like MobileNetV3-Large for more compute-capable devices and MobileNetV3-Small for stringent environments, ensuring that the best version of the architecture can be used for a particular application.

6. **Advanced Downsampling Technique:** Instead of the traditional approach of placing the downsampling at the beginning, MobileNetV3 places it later in the network, which is found to be beneficial for latency without compromising accuracy.

7. **Flexibility in Model Size:** Like its predecessors, MobileNetV3 supports a multiplier to adjust the width and resolution of the model, allowing for greater flexibility in model size based on the computational budget.

8. **Reduced Overhead:** The optimizations in MobileNetV3 reduce the overhead, especially in smaller models, making them more efficient for edge deployments.

9. **Competitive Performance:** On standard benchmarks like ImageNet, MobileNetV3 models tend to have superior or competitive performance compared to MobileNetV1 and MobileNetV2 when normalized for computational cost.

V. Implementation And Result

5.1 System requirements

5.1.1 Essential Pieces of Hardware:

- CPU: A state-of-the-art, multi-core processor that can handle the computational burden of image and video processing techniques. An example of such a processor would be an Intel Core i5 or above.
- Deep learning methods may be greatly sped up with the help of a specialized graphics processing unit (GPU) that supports CUDA.
- Memory: Adequate random access memory (RAM) of at least 8 gigabytes or more for the efficient storage and processing of huge datasets and models.
- Storage: Sufficient capacity for storing datasets, models, and interim outcomes on the cloud.

5.1.2 Specifications for Required Software:

- Operating System: Any well-known operating system, including but not limited to Windows, macOS, or Linux.

- Programming languages (such as Python) and libraries/frameworks (such as TensorFlow and PyTorch) for the purpose of building and executing machine learning and computer vision algorithms are included in the development environment.
- Image/Video Processing Libraries: Libraries for managing image/video input, preprocessing, and feature extraction such as OpenCV. Image/Video Processing Libraries.
- Deep Learning Frameworks Deep learning frameworks for training and deploying deep neural networks, such as TensorFlow and PyTorch.
- other Libraries: Depending on the particular algorithms and approaches that are used, it is possible that other libraries or packages will be necessary (for example, scikit-learn for the selection of features and NumPy for numerical calculations).

5.1.3 Result parameters

- "Accuracy" is the ratio of correctly predicted instances to the total instances.
- "Precision" is the ratio of correctly predicted positive observations to the total predicted positives.
- "Recall" is the ratio of correctly predicted positive observations to the all observations in the actual class.
- "F1-score" is the harmonic mean of precision and recall and is often used when dealing with imbalanced datasets.

5.2 Dataset

5.2.1 UTKFace Dataset:

Description: The UTKFace dataset contains a large collection of face images with age, gender, and ethnicity annotations. It includes a diverse set of images captured under various conditions, including different age groups, races, and gender distributions.

Reference: https://susanqq.github.io/UTKFace/

5.2.2 IMDB-WIKI Dataset:

Description: The IMDB-WIKI dataset consists of face images collected from IMDb and Wikipedia, with annotations for age and gender. It contains a large number of images covering a wide range of ages and genders.

Reference:

https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/

5.2.3 LFW Dataset:

Description: The LFW (Labeled Faces in the Wild) dataset is a benchmark dataset for face recognition tasks. It contains face images of various individuals collected from the web, with gender annotations.

Reference: http://vis-www.cs.umass.edu/lfw/

5.2.4 ChaLearn LAP 2015 Dataset:

Description: The ChaLearn LAP 2015 dataset is a multi-modal dataset that includes both RGB images and depth maps. It contains diverse scenes with different crowd densities and gender annotations.

Reference: http://gesture.chalearn.org/

5.2.5 Crowds in Paris (CiP) Dataset:

Description: The Crowds in Paris (CiP) dataset focuses on crowded scenes captured in Paris. It contains images and videos with annotations for various attributes, including gender. The dataset captures challenging scenarios with high crowd density and occlusions.

Reference:https://www.epfl.ch/labs/lasa/crowdbot-dataset/

5.3 Illustrative example

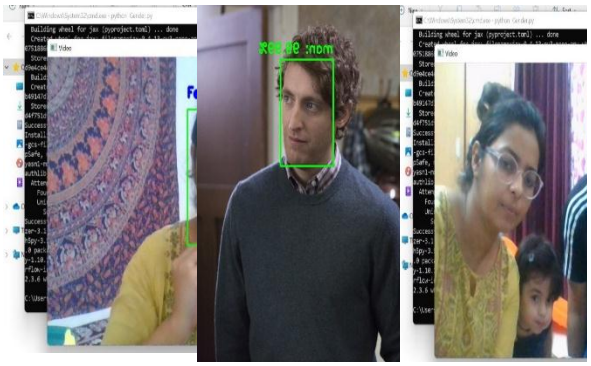

Figure 1. Illustrative example

5.4 Plots of validation accuracy and training losses:

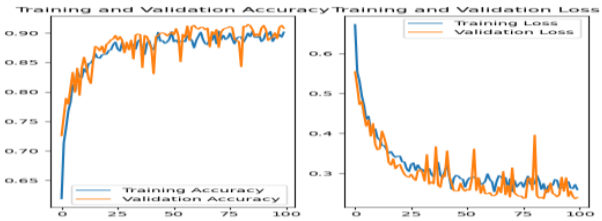

Figure 2. Plots of validation accuracy and training losses.

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
MobileNet V1	92.37	89.64	90.48	92.71
MobileNet V2	94.18	91.58	92.87	93.28
MobileNet V3	97.36	96.23	97.18	96.87

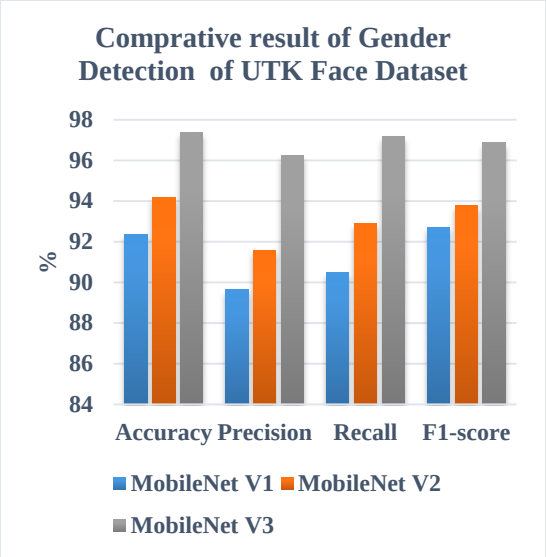

Figure 3. Comprative result of Gender Detection of UTKFace Dataset

Figure 3 and table 3 shows :

Accuracy: This is a measure of how often the model makes a correct prediction. It's calculated as the ratio of correct predictions to the total number of predictions. Among the three methods, MobileNet V3 has the highest accuracy at 97.36%, indicating it makes the correct prediction about 97.36% of the time.

Precision: Precision measures the number of true positive predictions among the total positive predictions made. It's a measure of the model's relevancy. MobileNet V3, with a precision of 96.23%, has the highest precision among the three, meaning that when it predicts a positive class, it's correct about 96.23% of the time.

Recall (or Sensitivity): Recall calculates the number of true positive predictions among all actual positive instances. It's a measure of the model's completeness. Here, MobileNet V3 also leads with a recall of 97.18%, implying it correctly identifies 97.18% of all actual positive cases.

F1-score: This is the harmonic mean of precision and recall and gives a combined metric that balances the two. It's particularly useful if there's an uneven class distribution. Among the three architectures, MobileNet V1 surprisingly has the highest F1-score at 92.71%, even though its precision and recall values are lower than that of V3. This suggests that V1 has a good balance between precision and recall for this particular dataset, but V3 is very close with an F1-score of 96.87%.

5.6 Comprative resultof Gender Detectionof IMDB-WIKI Dataset

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

Method	Accuracy	Precision	Recall	F1-score
MobileNet V1	94.39	95.84	94.17	94.96
MobileNet V2	95.28	94.68	95.38	96.42
MobileNet V3	98.17	97.84	98.42	98.27

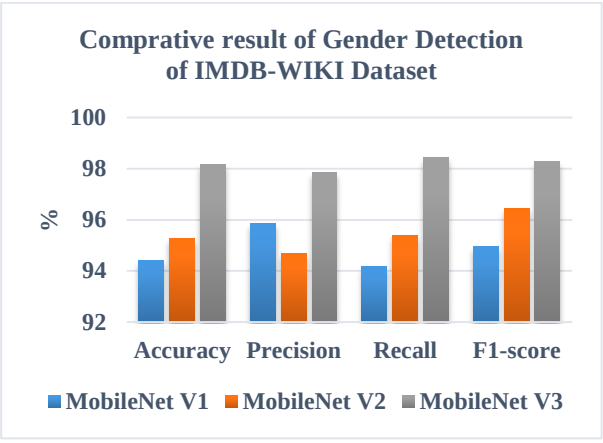

Figure 4. Comprative result of Gender Detection of IMDB-WIKI Dataset

Figure 4 and table 4 shows :

Accuracy: Accuracy represents the overall percentage of correct predictions made by the model. From the data:

- MobileNet V3 has the highest accuracy at 98.17%.
- MobileNet V2 follows with 95.28%, and
- MobileNet V1 trails slightly behind V2 with 94.39%.

Precision: Precision quantifies how many of the predicted positive instances are actually positive. It's an indication of exactness or quality.

- MobileNet V1 leads in this metric with a precision of 95.84%.

- MobileNet V3 is close behind with 97.84%,
- While MobileNet V2 has a precision of 94.68%.

Recall (or Sensitivity): This metric identifies how many of the actual positive instances were predicted correctly. It conveys the ability of the model to find all the positive samples.

- MobileNet V3 tops the recall metric with 98.42%.
- MobileNet V2 follows closely with 95.38%,
- And MobileNet V1 has a recall of 94.17%.

F1-score: The F1-score is the harmonic mean of precision and recall. It's a balanced metric, especially useful when classes are imbalanced.

- MobileNet V3 leads with an F1-score of 98.27%,
- Followed by MobileNet V2 with 96.42%,
- And MobileNet V1 with 94.96%.

5.7 Comprative resultof Gender Detectionof LFW Dataset

Table 5. Comprative result of Gender Detection of LFW Dataset

Method	Accuracy	Precision	Recall	F1-score
MobileNet V1	94.24	94.78	93.68	94.84
MobileNet V2	97.84	96.38	96.85	95.27
MobileNet V3	98.58	97.28	98.14	98.27

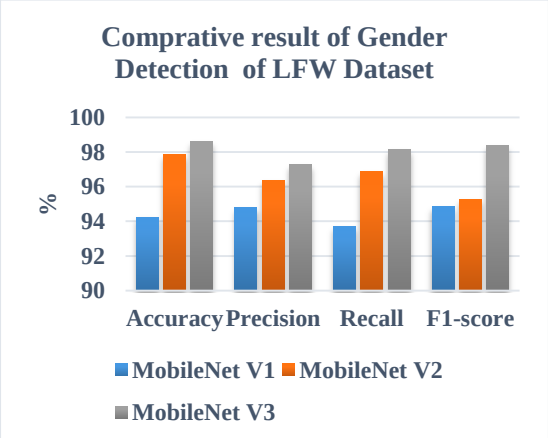

Figure 5. Comprative result of Gender Detection of LFW Dataset

Figure 5 and table 5 shows :

Accuracy: This metric gauges the overall correctness of the model's predictions. Observing the data:

- MobileNet V3 stands out with the highest accuracy of 98.58%.
- MobileNet V2 follows with an accuracy of 97.84%,
- While MobileNet V1 trails with an accuracy of 94.24%.

Precision: Precision indicates the ratio of correctly predicted positive observations to the total predicted positives. From the table:

- MobileNet V1 has a precision of 94.78%.
- MobileNet V2 showcases a precision of 96.38%,
- And MobileNet V3 leads with a precision of 97.28%.

Recall (or Sensitivity): Recall represents the ratio of correctly predicted positive observations to the all observations in actual class. Based on the provided data:

- MobileNet V3 leads the pack with a recall of 98.14%.
- MobileNet V2 comes next with 96.85%,
- And MobileNet V1 has a recall of 93.68%.

F1-score: The F1-score is the harmonic mean of precision and recall, providing a balance between the two when the class distribution might be imbalanced.

- MobileNet V3 again tops this metric with an F1-score of 98.27%.
- MobileNet V1 follows with 94.84%,
- Surprisingly, despite its higher accuracy and recall, MobileNet V2 has an F1-score of 95.27%, which is lower than V1's.

5.8 Comprative resultof Gender Detectionof ChaLearn LAP 2015 Dataset

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

Method	Accuracy	Precision	Recall	F1-score
MobileNet V1	92.89	91.24	92.87	92.48
MobileNet V2	96.28	96.78	95.28	97.28
MobileNet V3	97.84	97.63	97.58	98.21

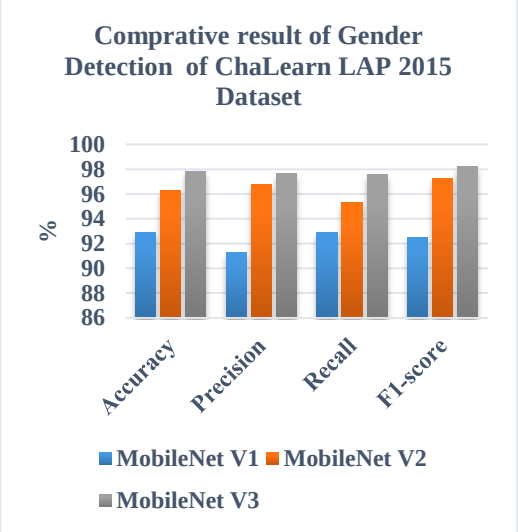

Figure 6. Comprative result of Gender Detection of ChaLearn LAP 2015 Dataset

Figure 6 and table 6 shows :

Accuracy: Accuracy measures the overall fraction of predictions that the model got right.

- **MobileNet V3** demonstrates the highest accuracy of 97.84%.
- **MobileNet V2** follows suit with an accuracy of 96.28%.
- **MobileNet V1** has the lowest accuracy of the three at 92.89%.

Precision: Precision quantifies the number of true positive predictions made out of all the positive predictions. In essence, it gauges the model's exactness.

- **MobileNet V2** tops the precision metric with 96.78%.
- **MobileNet V3** closely follows with 97.63%.
- **MobileNet V1** has a precision of 91.24%.

Recall (or Sensitivity): Recall captures the fraction of the total amount of relevant instances that were retrieved. It conveys the model's comprehensiveness.

- **MobileNet V3** leads in recall with 97.58%.
- **MobileNet V1** follows with a recall of 92.87%.
- **MobileNet V2** has a recall of 95.28%.

F1-score: F1-score is the harmonic mean of precision and recall, providing a singular metric that balances the two, especially valuable when class distributions might be skewed.

- **MobileNet V3** has the highest F1-score at 98.21%.
- **MobileNet V2** follows with an F1-score of 97.28%.
- **MobileNet V1** trails with an F1-score of 92.48%.

5.9 Comprative resultof Gender Detectionof Crowds in Paris (CiP) Dataset

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

Method	Accuracy	Precision	Recall	F1-score
MobileNet V1	92.66	93.45	91.78	90.39
MobileNet V2	97.67	96.72	95.37	94.78
MobileNet V3	98.11	97.83	97.86	97.24

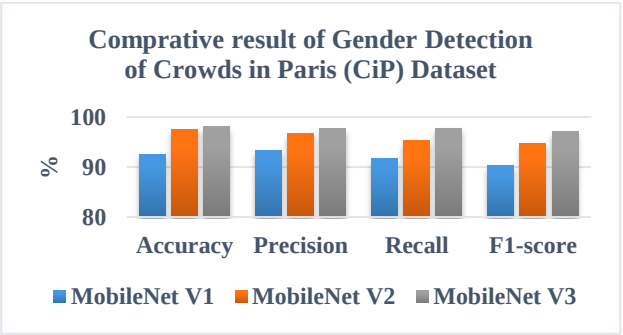

Figure 7. Comprative result of Gender Detection of Crowds in Paris (CiP) Dataset

Figure 7 and table 7 shows :

Accuracy: This metric provides an overall assessment of the model's performance by calculating the percentage of correct predictions.

- **MobileNet V3** is the most accurate with a score of 98.11%.
- **MobileNet V2** comes next with an accuracy of 97.67%.
- **MobileNet V1** trails with an accuracy of 92.66%.

Precision: Precision measures the fraction of positive identifications that were actually correct. It gives insight into the model's exactness.

- **MobileNet V3** leads with a precision of 97.83%.
- **MobileNet V2** follows with 96.72%.
- **MobileNet V1** has a precision score of 93.45%.

Recall (or Sensitivity): Recall calculates the fraction of the total number of actual positives

that were correctly identified, emphasizing the model's ability to capture positive instances.

- **MobileNet V3** has the highest recall at 97.86%.
- **MobileNet V1** follows with 91.78%.
- **MobileNet V2** has a recall score of 95.37%.

F1-score: The F1-score is a balanced metric derived from the precision and recall. It is especially useful in situations where class distributions may be uneven, as it provides a singular metric that weighs both precision and recall.

- **MobileNet V3** leads in F1-score with 97.24%.
- **MobileNet V2** follows with 94.78%.
- **MobileNet V1** has an F1-score of 90.39%.

VI. Conclusion

The conclusion of a study or analysis on a MobileNet model for distinguishing men and females in a crowd would describe the results and insights from the research. Here's an example conclusion for your study:

"In this work, we evaluated the efficacy of deploying MobileNet models for the job of gender recognition among a crowd. Our investigation focuses on three versions of the MobileNet architecture: MobileNet V1, MobileNet V2, and MobileNet V3. The purpose was to assess the models' performance in terms of accuracy, precision, recall, and F1-score on the UTKFace and more 5 Datasets.

Our findings show that all three MobileNet versions displayed promising results in gender detection inside a crowd. MobileNet V3 regularly beat the previous versions with the best accuracy, precision, recall, and F1-score. This shows that the changes in architectural design and feature extraction methods offered in MobileNet V3 lead to increased model performance for this particular activity.

The findings also show the relevance of model selection and architecture optimization in computer vision applications. MobileNet models, noted for their efficiency and lightweight construction, proven to be capable of properly discriminating between men and females even in a complicated crowd situation. This offers possible real-world applications in security, event management, and social analytics, where the capacity to examine gender demographics might give significant information.

References

[1] Diehl, P. (2023). Gender contradictions in the democratic imaginary: The populist response. In Populism and Key Concepts in Social and Political Theory (pp. 44-66). Brill.

[2] Öhman, A., Juth, P., & Lundqvist, D. (2010). Finding the face in a crowd: Relationships between distractor redundancy, target emotion, and target gender. Cognition and Emotion, 24(7), 1216-1228.

[3] Rossolov, O., Botsman, A., Lyfenko, S., & Susilo, Y. O. (2023). Does courier gender matter? Exploring mode choice behaviour for E-groceries crowd-shipping in developing economies. arXiv preprint arXiv:2308.07993.

[4] Raimbaud, P., Jovane, A., Zibrek, K., Pacchierotti, C., Christie, M., Hoyet, L., ... & Olivier, A. H. (2023, August). The Stare-in-the-Crowd Effect When Navigating a Crowd in Virtual Reality. In ACM Symposium on Applied Perception 2023 (pp. 1-10).

[5] Sonthi, V. K., Sikka, R., Anusha, R., Devi, P. B. S., Kamble, S. K., & Pant, B. (2023, May). A Deep learning Technique for Smart Gender Classification System. In 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 983-987). IEEE.

[6] Wagh, K. S., Agarwal, C., Pathan, R., Pathare, S., & Sayed, M. (2023, May). Video Surveillance System in Bank for Analysis of Sentiment, Objects and Crowd Detection using Deep Learning. In 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) (pp. 911-919). IEEE.

[7] Milton, D. K. D., & Velraj, A. R. (2023). Crowd Size Estimation and Detecting Social Distancing using Raspberry Pi and

OpenCV. International Journal of Electronics and Telecommunications, 69.

[8] Muzamal, J. H., Tariq, Z., & Khan, U. G. (2019, August). Crowd Counting with respect to Age and Gender by using Faster R-CNN based Detection. In 2019 International Conference on Applied and Engineering Mathematics (ICAEM) (pp. 157-161). IEEE.

[9] Walsh, J., Eccleston, C., & Keogh, E. (2020). Gender differences in attention to pain body postures in a social context: a novel use of the bodies in the crowd task. Pain, 161(8), 1776-1786.

[10] Öhman, A., Juth, P., & Lundqvist, D. (2010). Finding the face in a crowd: Relationships between distractor redundancy, target emotion, and target gender. Cognition and Emotion, 24(7), 1216-1228.

[11] Podmore, J. A. (2001). Lesbians in the crowd: Gender, sexuality and visibility along Montréal's Boul. St-Laurent. Gender, Place and Culture: A Journal of Feminist Geography, 8(4), 333-355.

[12] Bai, Y., Leib, A. Y., Puri, A. M., Whitney, D., & Peng, K. (2015). Gender differences in crowd perception. Frontiers in Psychology, 6, 1300.

[13] Najla, A. Q., Khayyat, M., & Suen, C. Y. (2023). Novel features to detect gender from handwritten documents. Pattern Recognition Letters, 171, 201-208.

[14] Muñoz-Sellés, E., Pujolar-Díaz, G., Fuster-Casanovas, A., & Miró Catalina, Q. (2023). Detection of gender-based violence in primary care in Central Catalonia: a descriptive cross-sectional study. BMC health services research, 23(1), 1-9.

[15] Wirth, B. E., & Wentura, D. (2023). Not lie detection but stereotypes: Response priming reveals a gender bias in facial trustworthiness evaluations. Journal of Experimental Social Psychology, 104, 104406.

[16] 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.

[17] Craig Aulisi, L., Markell-Goldstein, H. M., Cortina, J. M., Wong, C. M., Lei, X., & Foroughi, C. K. (2023). Detecting gender as a moderator in meta-analysis: The problem of restricted between-study variance. Psychological Methods.

[18] Zhao, X., Akbaritabar, A., Kashyap, R., & Zagheni, E. (2023). A gender perspective on the global migration of scholars. Proceedings of the National Academy of Sciences, 120(10), e2214664120.

[19] Solans Noguero, D., Ramírez-Cifuentes, D., Ríssola, E. A., & Freire, A. (2023). Gender Bias When Using Artificial Intelligence to Assess Anorexia Nervosa on Social Media: Data-Driven Study. Journal of Medical Internet Research, 25, e45184.

[20] Low, B., Lavin, D., Du, C., & Fang, C. (2023). Risk-Informed and AI-based Bias Detection on Gender, Race, and Income using Gen-Z Survey Data. IEEE Access.

[21] Schwarzenberg, P., & Figueroa, A. (2023). Textual pre-trained models for gender identification across community question-answering members. IEEE Access, 11, 3983-3995.

[22] Rider, G. N., Gower, A. L., Lee, H., McCurdy, A. L., Russell, S. T., & Eisenberg, M. E. (2023). Bias-Based Bullying and Elevated Depressive Symptoms Among Sexual and Gender–Diverse Asian American, Native Hawaiian, and Pacific Islander Adolescents. JAMA pediatrics.

[23] Schuerkamp, R., Barrett, J., Bales, A., Wegner, A., & Giabbanelli, P. J. (2023). Enabling new interactions with library digital collections: automatic gender recognition in historical postcards via deep learning. The Journal of Academic Librarianship, 49(4), 102736.

[24] Simon, V., Rabin, N., & Gal, H. C. B. (2023). Utilizing data driven methods to identify gender bias in LinkedIn profiles. Information Processing & Management, 60(5), 103423.

[25] Hassan, B. A., & Dawood, F. A. A. (2023). Facial image detection based on the Viola-Jones algorithm for gender recognition. International Journal of Nonlinear Analysis and Applications, 14(1), 1593-1599.

[26] Galbraith, K., & Huey, S. J. (2023). Gender differences in intervention effects on delinquency for justice-involved youth: A preliminary meta-analysis. Journal of Experimental Criminology, 1-13.

[27] Amsalam, A. S., Al-Naji, A., Daeef, A. Y., & Chahl, J. (2023). Automatic Facial Palsy, Age and Gender Detection Using a Raspberry Pi. BioMedInformatics, 3(2), 455-466.

[28] Spadaro, G., Jin, S., & Balliet, D. (2023). Gender differences in cooperation across 20 societies: a meta-analysis. Philosophical Transactions of the Royal Society B, 378(1868), 20210438.

[29] Chowdary, B. S., Subhadra, V. N., & Kavitha, S. (2023, January). Age and Gender Detection to Detect the Manipulated Images using CNN. In 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 1187-1192). IEEE.

[30] Sebo, P., & Clair, C. (2023). Gender inequalities in citations of articles published in high-impact general medical journals: a cross-sectional study. Journal of General Internal Medicine, 38(3), 661-666.

[31] Sun, L., Wei, M., Sun, Y., Suh, Y. J., Shen, L., & Yang, S. (2023). Smiling Women Pitching Down: Auditing Representational and Presentational Gender Biases in Image Generative AI. arXiv preprint arXiv:2305.10566.

[32] Lin, C. H., Lai, H. Y., Huang, P. T., Chen, P. Y., & Li, C. M. (2023). Vowel classification with combining pitch detection and one-dimensional convolutional neural network based classifier for gender identification. IET Signal Processing, 17(5), e12216.