International Journal of

INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

An Embedded VGG 22 Model for Gender Classification in Crowd Videos

Priyanka Singh *[1], Dr. Rajeev Vishwakarma [2]

Submitted: 16/11/2023 **Revised**: 28/12/2023 **Accepted**: 08/01/2024

Abstract: This study presents the development and optimization of an embedded VGG model for the purpose of gender classification in crowd videos. Traditional VGG models offer robust feature extraction for image classification but are computationally intensive, rendering them less practical for real-time analysis in embedded systems with limited resources. Addressing this, the research explores the implementation of VGG11 through a VGG22 architecture, analyzing their performance in terms of accuracy, precision, recall, and F1-score. The findings indicate a trend of increasing performance with deeper architectures, with the VGG22 model achieving the highest scores across all metrics. The research methodology involved adapting the VGG architecture to the constraints of embedded systems through model compression techniques such as pruning and quantization, alongside optimization strategies like knowledge distillation. The models were evaluated using standard gender classification datasets, with a particular focus on the challenging conditions of crowd video data. The results confirmed that with careful optimization, it is possible to maintain high accuracy in gender classification while significantly reducing the computational demands of the model.

Keywords: *VGG11, VGG13, VGG16, VGG19, VGG22, Gender , Deep Convolutional Neural Network.*

1. Introduction

In the realm of computer vision, gender classification stands as a cornerstone task with substantial applications spanning from targeted advertising to enhanced user interfaces and security systems. The proliferation of crowd-sourced video data, captured by the omnipresent surveillance cameras, has necessitated the development of automated and reliable gender classification systems [1]. These systems are expected to function accurately despite the challenges posed by uncontrolled environments, such as variable lighting, diverse camera angles, and the presence of occlusions and motion blur. The introduction of deep learning, particularly Convolutional Neural Networks (CNNs), has significantly advanced the capabilities of such classification systems. Among the various CNN architectures, the Visual Geometry Group's VGG models have garnered attention due to their depth and robust feature extraction capabilities, leading to impressive performance in image classification tasks.

However, the deployment of these sophisticated models in real-world scenarios is not without its challenges. The computational intensity required to run deep CNNs like VGG often exceeds the processing capabilities of the embedded systems that operate at the edge of the network, where real-time processing is crucial. To address this, there is a growing interest in developing an embedded VGG model that maintains the accuracy of its more computationally demanding counterparts while adhering to the constraints of embedded systems [2]. This involves optimizing the network architecture and leveraging model compression techniques to reduce the size and computational load without significantly impacting performance.

This work focuses on adapting the VGG model for gender classification in crowd videos, specifically tailored for embedded systems. The goal is to achieve a balance between model complexity and computational efficiency, enabling accurate and real-time gender classification suitable for deployment in embedded devices. By compressing and optimizing the VGG architecture, the model can be made suitable for on-device processing, thus reducing the latency and bandwidth requirements associated with cloud processing.

Moreover, this research delves into the practical considerations of deploying such models in real-world scenarios, including the ethical implications and privacy concerns of automated gender classification [3]. It aims to provide a solution that is not only technically feasible but also socially responsible. The expected outcome is a streamlined VGG-based model that sets a new standard for gender classification in crowd videos, paving the way for more advanced and accessible computer vision applications in everyday technology.

[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

2. Background Study

2.1 Overview of Gender Classification Challenges in Crowd Videos

Gender classification within crowd videos encapsulates a multitude of challenges, forming a complex problem space for machine learning and computer vision. This complexity arises primarily due to the unconstrained nature of crowd environments where factors such as diverse lighting conditions, varying angles and distances of faces from the camera, motion blur due to movement, occlusions caused by objects or other individuals, and the sheer diversity in clothing and physical appearances add layers of difficulty. Each of these factors can significantly degrade the quality of the input data for classification models. Furthermore, the scalability of gender classification algorithms to process data from multiple video streams simultaneously while maintaining high accuracy and low latency presents a technical hurdle.

These algorithms must be robust enough to handle the high variability and yet remain efficient enough to provide real-time analytics, a critical requirement for applications such as surveillance, behavioral analysis, and crowd management. In crowd videos, faces are rarely perfectly posed; they appear at different orientations and are often partially visible. This results in a limited availability of clear, frontal facial features, which are crucial for accurate gender classification. Additionally, the variation in facial expressions and the transient presence of individuals in video frames complicate the extraction of reliable gender-specific features [4]. The dynamic nature of crowds means that the algorithms must be adaptive, learning from new data and potentially retraining or updating models in the field to maintain accuracy over time. This is coupled with the need for privacy preservation and ethical considerations in deploying gender classification systems in public spaces.

2.2 Evolution of Gender Classification Techniques

The evolution of gender classification techniques is a tale of constant innovation spurred by technological advancements and the growing complexity of application demands. In the initial stages, gender classification relied heavily on geometric feature-based approaches, where algorithms focused on measuring distances and angles between key facial landmarks. These techniques, while foundational, were limited by their reliance on high-quality, frontal images and suffered significantly in accuracy when faced with real-world conditions. As machine learning became more sophisticated, statistical models like Support Vector Machines (SVM), Linear Discriminant Analysis (LDA), and simple neural networks came into play. These models could handle more variability in facial features and were better at managing differences in poses and expressions. They were, however, still limited in their ability to deal with low-resolution images and partial occlusions [5].

The breakthrough came with the advent of deep learning, particularly the development of Convolutional Neural Networks (CNNs). These models could learn hierarchical representations of face images, making them adept at recognizing and classifying gender even from low-quality images and under varied lighting conditions. Techniques such as transfer learning, where a model trained on one task is adapted for another, further enhanced the ability of CNNs to generalize from one dataset to another. In recent times, the focus has shifted towards creating models that are not only accurate but also efficient enough to run in real-time on embedded systems with limited computational resources. This has led to the development of lightweight deep learning models and the use of edge computing to bring processing closer to where data is captured. Simultaneously, there has been a growing emphasis on addressing bias and ensuring fairness in gender classification algorithms [6]. This involves training models on diverse datasets that represent different ethnicities, ages, and other demographic factors to avoid perpetuating stereotypes or inequality. The trajectory of gender classification techniques reflects a journey from rule-based algorithms to data-driven models, paralleling the broader trends in artificial intelligence and computational technology. The field continues to evolve rapidly, with ongoing research focusing on improving the robustness, efficiency, and ethical aspects of gender classification.

2.3 Deep Learning Models for Image and Video Analysis

Deep learning models have revolutionized image and video analysis, providing powerful tools for a myriad of applications ranging from facial recognition to autonomous driving. At the heart of this revolution is the Convolutional Neural Network (CNN), which has become synonymous with deep learning in visual domains. CNNs are adept at automatically and adaptively learning spatial hierarchies of features from image data. They are composed of multiple layers of convolutional filters that apply various transformations to the input, capturing features like edges, textures, and patterns that are essential for image classification tasks.

Another critical model in deep learning for image and video analysis is the Recurrent Neural Network (RNN), especially its more advanced variants like Long Short-Term Memory (LSTM) networks. RNNs are designed to handle sequential data and are thus well-suited for video analysis where temporal dependencies between frames are crucial. Transfer learning, where a model developed for one task is reused as the starting point for a model on a second task, has been a key factor in the success of deep learning for image and video analysis [7]. Models pre-trained on large datasets, such as ImageNet, can be fine-tuned with a smaller amount of data for tasks like gender classification in crowd videos. Generative Adversarial Networks (GANs) have also made significant contributions, particularly in image generation and enhancement. They can create high-resolution images from low-resolution inputs, which is beneficial for improving the quality of video frames before analysis.

Autoencoders, particularly variational autoencoders, are used for unsupervised learning tasks such as feature extraction and dimensionality reduction in image data, which are essential for clustering and anomaly detection in video streams. Attention mechanisms, which allow models to focus on specific parts of an image, have improved the performance of deep learning models in tasks where the context and location of objects in the image are important.

Deep learning in image and video analysis is not without its challenges [8]. Models require large amounts of data and computational power to train, and there are ongoing concerns regarding the interpretability of the models and the biases that can be encoded within them. Despite these challenges, deep learning models continue to push the boundaries of what's possible in image and video analysis, providing more accuracy and depth to the interpretation of visual data. As computational power increases and more sophisticated models are developed, their impact on image and video analysis will only grow stronger.

2.4 Embedded Systems in Image Processing

Embedded systems have become a cornerstone for real-time image processing due to their ability to integrate software and hardware to perform dedicated tasks efficiently. These systems are characterized by their resource-constrained nature, requiring optimized algorithms that can run with limited memory and computational power while still delivering fast and reliable results. In the context of image processing, embedded systems are typically designed to handle specific tasks such as facial recognition, object tracking, or gender classification in video streams. These tasks demand not only accuracy but also the ability to process data in real time, which can be challenging given the computational intensity of most image processing algorithms, especially those involving deep learning models. To address this, there has been a surge in the development of compact and efficient deep learning architectures that can operate within the constraints of embedded systems [9]. Techniques such as network pruning, quantization, and knowledge distillation are used to reduce the size of the models and the complexity of computations. This involves trimming unnecessary network weights, reducing the precision of the numerical representations, and transferring knowledge from a large model to a smaller one, respectively. Another critical aspect of embedded systems in image processing is the use of specialized hardware accelerators such as Field-Programmable Gate Arrays (FPGAs) and Graphics Processing Units (GPUs).

These accelerators can perform parallel processing, which is highly beneficial for the matrix and vector operations that are fundamental in image processing tasks. Embedded systems often employ edge computing, where data processing occurs close to the data source, minimizing latency and reducing the need for constant connectivity to central servers. This approach is particularly useful in scenarios where quick decision-making is crucial, such as autonomous vehicles or security surveillance systems [10]. The integration of artificial intelligence (AI) with embedded systems has also been a significant advancement, allowing for more intelligent and adaptive image processing. AI-enabled embedded systems can learn from new data, improve their performance over time, and make decisions autonomously. Despite their potential, embedded systems in image processing must navigate challenges such as power consumption, heat dissipation, and the need for robustness against diverse operational environments. As technology continues to advance, the capabilities of these systems are expanding, enabling more complex and sophisticated image processing applications to be deployed in real-world scenarios [11].

3. Literature Review

3.1 Gender classification in computer vision

Gender classification in computer vision is a task that involves the identification and categorization of individuals into gender categories based on their facial features or other form factors. It is a subset of biometric classification that leverages the capabilities of artificial intelligence (AI) and, more specifically, machine learning algorithms to discern gender from images or video feeds. The process generally involves several steps, starting with face detection, where the system locates the face within an image. Once a face is detected, feature extraction occurs, which isolates and identifies the unique attributes of the face that are relevant to distinguishing gender. In the early days, these features were often handcrafted based on domain knowledge, such as the distance between facial landmarks. With the advent of deep learning, particularly Convolutional Neural Networks (CNNs), gender classification systems have made significant strides. These networks learn to identify gender-distinguishing features directly from the data, often resulting in higher accuracy levels than methods relying on hand-engineered features [12].

CNNs can handle a wide range of variances in facial images, including different expressions, poses, and lighting conditions, which are common in real-world scenarios.

However, gender classification systems face challenges related to the diversity and bias of training datasets. Systems trained on non-representative datasets may not perform equitably across all demographics. Hence, there's an ongoing effort in the community to create more balanced datasets and design algorithms that are fair and unbiased. Privacy and ethical considerations are also paramount, as gender classification can be sensitive. Ensuring that these systems are used responsibly and with consent is a critical concern for developers and end-users alike [13].

3.2 Deep learning for image and video analysis

Deep learning has fundamentally transformed the field of image and video analysis. The core of these advancements is the Convolutional Neural Network (CNN), a deep learning architecture specifically designed to handle pixel data. CNNs and their variants have become the go-to methods for tasks such as object detection, image segmentation, facial recognition, and video analysis due to their ability to automatically extract and learn feature hierarchies from raw data. In image analysis, CNNs can identify patterns and objects in images with a high degree of accuracy, even under variable conditions such as different lighting or angles. For video analysis, CNNs are often combined with Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks, to capture temporal dependencies and analyze sequences of frames. Transfer learning, where a model pre-trained on a large dataset is fine-tuned for a specific task, has been particularly effective for image and video analysis, allowing for high performance even with smaller datasets [14].

This approach is widely used in gender classification, where pre-trained models are adapted to focus on the features relevant to distinguishing between genders. Generative Adversarial Networks (GaN) have also made significant strides, particularly in video synthesis and in generating high-resolution images from low-resolution inputs, which is beneficial for video enhancement and restoration. Moreover, the use of attention mechanisms in deep learning models allows the network to focus on the most informative parts of an image, enhancing performance on tasks such as scene understanding and anomaly detection in videos. However, deep learning models, particularly those used for video analysis, require substantial computational resources. This necessitates the use of powerful GPUs and has led to the development of model compression techniques to allow deployment on less powerful devices, such as mobile phones and embedded systems.

3.3 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a class of deep neural networks that have become the backbone of many computer vision systems. CNNs are particularly well-suited for analyzing visual imagery and are used extensively in image and video recognition, recommender systems, image classification, medical image analysis, natural language processing, and other machine learning tasks [15]. The architecture of a CNN is designed to mimic the human visual system and is composed of one or more convolutional layers often followed by pooling layers, fully connected layers (dense layers), and normalization layers. Here’s a breakdown of the key components in a CNN:

1. **Convolutional Layer**: This is the core building block of a CNN that does most of the computational heavy lifting. It applies a convolution operation to the input, passing the result to the next layer. This layer's parameters consist of a set of learnable filters or kernels, which have a small receptive field but extend through the full depth of the input volume. As the filter slides (or convolves) around the input image, it learns features from the image.
2. **Activation Function**: After each convolution operation, an activation function like ReLU (Rectified Linear Unit) is used to introduce non-linear properties into the network. This helps the network to learn more complex patterns in the data.
3. **Pooling (Subsampling or Downsampling)**: This layer reduces the spatial size of the representation to reduce the amount of parameters and computation in the network. Pooling layers partition the input image into a set of non-overlapping rectangles and, for each such sub-region, outputs the maximum (Max Pooling) or average (Average Pooling).
4. **Fully Connected Layer**: Neurons in a fully connected layer have full connections to all activations in the previous layer. Their activations can hence be computed with a matrix multiplication followed by a bias offset.
5. **Normalization Layer**: Layers like Batch Normalization are used to stabilize learning by normalizing the input to each unit to have zero mean and unit variance. This helps in speeding up the training of the network and reduces the sensitivity to network initialization.
6. **Dropout**: This is a regularization technique where, during training, some layer outputs are randomly ignored or "dropped out" to prevent overfitting.

CNNs are effective for the following reasons:

- **Parameter Sharing**: A feature detector (such as a filter) that is useful in one part of the image is probably useful across the entire image.

- **Local Connectivity**: Focusing on local connectivity allows the network to concentrate on low-level features in the early layers, and then assemble them into higher-level features in later layers.

3.4 The VGG architecture

The VGG architecture is a defining model in the world of deep learning for computer vision, introduced by the Visual Geometry Group from Oxford University, which is where it gets its name. It was a breakthrough for its simplicity and depth at the time it was introduced [16].

Here are the main characteristics of the VGG architecture:

1. **Simplicity**: Unlike many preceding models, VGG's architecture is uniform. It uses a series of convolutional layers with small receptive fields followed by max-pooling layers, then concludes with a stack of fully connected layers.
2. **Small Receptive Fields**: The convolutional layers use small 3×33×3 filters, which is the smallest size to capture the notion of left/right, up/down, center. The small filter size allows for deeper architectures while keeping the number of parameters down.
3. **Depth**: VGG models are deep, with configurations going up to 19 layers. This depth is achieved by stacking convolutional layers on top of each other before applying a max-pooling layer. The depth of the network allows it to learn a hierarchy of features at various levels of abstraction.
4. **Channels**: The number of filters in the convolutional layers starts at 64 and is doubled after each max-pooling layer, until it reaches 512.
5. **Fully Connected Layers**: After a series of convolutional and max-pooling layers, the architecture is concluded with three fully connected layers where the final layer is a softmax classification layer.
6. **ReLU Activation**: The VGG architecture uses ReLU (Rectified Linear Unit) activation function throughout the network for introducing non-linearity.
7. **Fixed Input Size**: The architecture is designed to work on a fixed input image size of 224×224.
8. **Use of Max-Pooling**: Between the convolutional layers, max-pooling is used for spatial downsampling.

There are several variants of the VGG architecture, commonly referred to as VGG11, VGG16, and VGG19, where the numbers denote the number of layers that have weights. The most popular variants are VGG16 and VGG19, which include 16 and 19 layers respectively.

1. **VGG11**: This version has 11 layers with weights, consisting of 8 convolutional layers and 3 fully connected layers. The convolutional layers use filters with a very small receptive field of 3×33×3 (which is the smallest size to capture the notion of left/right, up/down, center), and the network uses max pooling layers to reduce volume size.
2. **VGG13**: The VGG13 model is similar to VGG11 but has 13 layers with weights; it includes 10 convolutional layers with more filters in the middle layers compared to VGG11.
3. **VGG16**: One of the most popular variants, VGG16 consists of 16 layers with weights. There are 13 convolutional layers and 3 fully connected layers. The convolutional layers are organized in blocks, with each block followed by a max-pooling layer for spatial down-sampling.
4. **VGG19**: The deepest standard VGG model, VGG19, has 19 layers with weights. It includes 16 convolutional layers arranged in a similar block structure as VGG16 but with more convolutional layers in the later blocks, and 3 fully connected layers.

3.5 Related work on gender classification in videos

This research delves into enhancing facial recognition through the nuanced estimation of age and gender, even under less controlled, real-world conditions. By leveraging the strengths of Deep Convolutional Neural Networks (DCNNs), the study has shown that pre-trained CNN models adapted for this purpose outshine the purpose-built GilNet model. Particularly, models adapted from domains closely related to age and gender tasks, like the VGG-Face CNN, have shown superior performance, emphasizing the benefits of domain-specific transfer learning [1]. In another exploration, gender determination from fingerprint images via deep learning techniques is presented. Without human-led image preprocessing, models such as VGG-19, ResNet-50, and EfficientNet-B3 have been trained from the ground up, with EfficientNet-B3 showing remarkable accuracy. This advancement underscores the potential of end-to-end deep learning for biometric identification [2]. The paper also critiques traditional methods for organizing real-world facial photos by age and gender, noting their limitations in handling variance. It proposes a novel CNN framework that refines the feature extraction and classification processes, achieving state-of-the-art performance on benchmark datasets after pretraining and fine-tuning on relevant datasets [3].

Additionally, the text reviews studies examining the influence of image preprocessing, model initialization, and architecture on the recognition of age and gender in facial images. It highlights the role of proper initialization and preprocessing in achieving top-notch gender recognition performance using neural networks and the Layer-wise Relevance Propagation (LRP) algorithm for insightful model visualization [4]. Lastly, the text addresses the ethical aspect of automated gender classification, acknowledging the technology's susceptibility to biases across different gender and racial groups. It reports on studies showing disproportionate error rates, especially for darker-skinned individuals and women, and suggests that architectural differences in algorithms and training data imbalances contribute to this bias.

It proposes that facial morphological differences influenced by genetic and environmental factors could explain the lower performance among certain demographics, such as Black females [5]. Gender classification has become a dynamic field in research, with significant efforts dedicated to enhancing its application in areas like monitoring, surveillance, and human-computer interaction. Despite advancements, there's a noted gap in the performance of current methods on live images. The emergence of deep learning has, however, marked a notable improvement in complex tasks across different domains. This study leverages the TensorFlow framework to explore the efficacy of deep learning in gender classification, utilizing Keras models with pre-trained ImageNet weights. A comparison of models including VGG16, ResNet-50, and MobileNet is made, with a database comprising primarily Asian faces, revealing VGG-16 as the most accurate [6].

The paper also ventures into novel territory with an analysis of age and gender extraction from ear images, an area less explored in biometrics. Employing both geometric and appearance-based features, and using deep learning models fine-tuned on an extensive ear dataset, the study finds appearance-based methods superior. However, it suggests that age classification from ear images still requires further research [7]. For real-time applications requiring rapid gender classification, the paper presents a streamlined 16-layer network derived from VGG-16, optimized for performance without compromising accuracy, even on less powerful devices. The network utilizes Fisher's Linear Discriminant Analysis to prune less relevant neurons, achieving impressive reductions in size and computing time while maintaining high accuracy on standard datasets [8]. The significance of age and gender classification has grown across various sectors, from social media analytics to business demographics. The paper discusses the application of transfer learning to enhance classification performance on facial images, using an ensemble of CNN models for age estimation.

The results showcase high accuracy, particularly with the VGG14 model, emphasizing the potential of combining multiple deep learning approaches [9]. In the context of social media, age and gender classification's relevance has only increased with the advent of AI, which has boosted performance in visual recognition. However, the paper notes that the precision of mobile-friendly networks does not always match their larger counterparts, highlighting a trade-off between accuracy and accessibility [10]. The paper also touches on the broader field of pattern recognition and its critical role in developing intelligent systems, including biometrics for security. It presents a gender prediction and age estimation system based on CNNs, tested on widely recognized datasets, demonstrating substantial improvements in system performance and accuracy [11].

Lastly, the research delves into the broader area of soft biometrics, which encompasses traits like age and gender, crucial for enhancing communication between humans and machines. This comprehensive review covers contributions in gender classification and age estimation using neural networks, discussing datasets, findings, and metrics for a clear understanding of the research landscape and outlining potential future research directions [12]. Automatic gender recognition has become increasingly relevant, driven by the proliferation of social media and online networking platforms.

Despite this, current systems struggle to match the performance of related face recognition tasks, especially when dealing with images from the physical world. In this paper, we demonstrate how applying deep learning, specifically Deep Convolutional Neural Networks (D-CNN), enhances gender classification performance. We introduce an efficient VGGNet-based convolutional network architecture designed to perform well even when the training data is scarce. Our experiments show that this approach significantly outperforms existing methods in real-world gender recognition tasks [13]. Furthermore, Bone Age Assessment (BAA) is crucial in medical and forensic fields, and gender classification plays a vital role in it.

We present a novel framework that uses deep learning to classify gender and predict age from a single hand radiograph. Employing the VGG-16 model and transfer learning techniques, we achieve 79.6% accuracy in gender classification and highly precise age predictions [14]. Age and gender classification of faces is a key area of research with many practical applications. Convolutional Neural Networks (CNNs) have emerged as a particularly effective tool for this, thanks to their feature extraction and classification capabilities. In our study, we propose a CNN-based model that handles the variability of real-world faces through robust preprocessing and pretraining on a large dataset with unfiltered labels. By incorporating dropout and data augmentation, we mitigate overfitting, allowing the model to generalize well. Our model demonstrates superior performance on the OIU-Adience dataset, achieving 84.8% and 89.7% accuracy in age and gender classification, respectively [15].

Lastly, we explore the use of competition-winning deep neural networks with pretrained weights for gender recognition and age estimation. Transfer learning is applied to VGG19 and VGGFace models, and we assess various training techniques and parameters to enhance prediction accuracy. A hierarchical CNN system is evaluated that classifies by gender before predicting age with gender-specific models. The results are impressive, with a gender recognition accuracy of 98.7% and a mean age estimation error of 4.1 years, showcasing the efficacy of repurposing existing convolutional filters for new classification tasks [16]. This paper introduces multimodal deep neural network frameworks that use both profile face and ear images for age and gender classification. By incorporating ear appearance, a less common biometric modality, we aim to improve the accuracy of extracting these soft biometric traits. Our end-to-end deep learning frameworks employ various fusion strategies at data, feature, and score levels, enhanced by domain adaptation and a combination of center and softmax loss. Extensive testing on UND-F, UND-J2, and FERET datasets shows our multimodal system's high accuracy, surpassing current methods that rely solely on profile faces or ear images [17].

In live video, facial gender classification faces challenges like motion blur and variable lighting conditions. To address these, we propose a Multi-Branch Voting CNN (MBV-CNN) that detects faces, enhances their brightness adaptively, and employs three CNN branches, culminating in a majority voting mechanism. This significantly increases accuracy on both the LFW dataset and our Gender Classification for Live Videos (GCLV) dataset [18]. Gender classification from facial images is further explored using a transfer learning approach with the VGG16 CNN model, pre-trained on a large dataset of natural images. Fine-tuning this model on a well-balanced dataset yields better results than previous state-of-the-art methods on the LFW-Gender dataset [19].

The paper also examines the use of various CNN architectures, like VGG16, Inception V3, and ResNet50, to address the challenges of lighting, pose, and expression in gender classification. Through comprehensive dataset processing and optimization, the highest accuracy achieved is 95.10% [20]. In addition, we investigate gender classification using Near-Infrared Periocular iris images, applying deep learning to identify relevant features from a small dataset. A CNN trained from scratch using data augmentation outperforms pre-trained models like VGG and Resnet, achieving an accuracy of 85.48% [21]. A prototype is proposed for real-time gender and emotion classification using a robust convolutional neural network and depth-wise separable convolution for emotion recognition.

This method provides high accuracy for gender classification (95%) and facial emotion recognition (67%), with a significant reduction in model size due to fewer hyperparameters [22]. Lastly, the paper evaluates the performance of deep learning models for binary gender classification from fingerprints, highlighting the challenge of limited data and the need for time-efficient models. By applying data augmentation and transfer learning, the VGG-19 model shows the best performance, with a testing accuracy of 71.9% [23]. This paper delves into gender recognition as a component of human activity analysis, which is crucial for data mining and machine learning applications. To improve accuracy and model generalization, we gathered a dataset of five million weakly labeled facial images and conducted three experiments.

These experiments compare the performance of convolutional neural networks (CNNs) of various depths and a support vector machine using local binary patterns, assess the impact of contextual data on accuracy, and evaluate CNNs' cross-database generalization abilities. Our findings set new benchmarks, with a 98.90% accuracy on the Labeled Faces in the Wild (LFW) dataset and a 91.34% accuracy on the Images of Groups (GROUPS) dataset for cross-database gender classification [24]. Addressing the automatic gender classification challenge, this analysis paper emphasizes the difficulties presented by low-resolution images and occlusion in datasets. We explore the effectiveness of deeper CNNs trained on separate facial components and compare the results with leading gender classification methods. The study also considers how network configurations and parameters impact classification, offering insights into age-related gender distinctions. The technique proposed shows promise, especially with larger crop sizes, and can accurately classify gender using images of the mouth, nose, and face (excluding eyes) [25].

Lastly, we introduce a multimodal, multitask deep CNN framework for age and gender classification, incorporating ear and profile face biometrics. We experiment with data, feature, and score fusion methods to merge information from these biometrics, using VGG-16 and ResNet-50 models enhanced with center loss for sharper feature discrimination. A two-stage fine-tuning process is implemented to further refine the models' representational abilities. Testing on the FERET, UND-F, and UND-J2 datasets demonstrates the utility of ear and profile face images in extracting soft biometric traits.

The study reveals that these images are viable alternatives to frontal face views for biometric recognition systems, with the multimodal system achieving superior accuracy in age and gender classification, surpassing unimodal methods and previous state-of-the-art techniques [26]. The advent of deep learning has redefined the processing and analysis of large data sets, especially after Geoffrey Hinton's 2006 model that utilized multiple hidden layers in neural networks. This was initially challenging due to computational complexities, but the emergence of GPU technology has greatly enhanced the efficiency of such computations, leading to a resurgence in deep learning's popularity. Convolutional Neural Networks (CNNs), a subset of deep learning models, have become particularly prominent in image classification tasks. This study introduces a new CNN model specifically designed for gender classification, trained and tested using the Adience dataset, achieving an 88.5% accuracy rate. This model outperforms traditional machine learning and other CNN models [27].

Age Group Classification (AGC) remains a complex task due to the variability in human features. This study presents an AGC system that uses a two-stage process: preprocessing (including face detection, gamma correction, and normalization) and classification via the VGG16 architecture, which incorporates convolution, max-pooling, and activation functions. Tested on the MORPH database, the system attains over 90% accuracy across various age groups with the VGG16 architecture [28].

The paper also investigates gender classification using near-infrared images of the periocular region, a valuable trait for forensic applications. Employing two CNN-based approaches, one featuring a pre-trained CNN for feature extraction and an SVM for classification, and the other using an end-to-end classifier through fine-tuning a pre-trained CNN, the study achieves high accuracy compared to baseline methods, especially on non-ideal image datasets [29]. Furthermore, gender classification from face images, despite challenges like complex backgrounds and varying lighting conditions, is explored using CNN and Alex Net models. Both models prove effective, with comparative analyses demonstrating their capability in gender classification from facial images [30]. Lastly, face gender recognition is vital for enhancing human-robot interactions.

This work compares the performance of SVM and CNN models when paired with hand-crafted, deep-learned, and fused features. Conducting tests on the Adience and LFW datasets and using statistical methods to validate the results, it is found that SVMs perform best with fused features, while CNNs excel with deep-learned features, with CNN significantly outperforming SVM in terms of accuracy and other metrics [31]. Classifying facial gender and detecting smiles in uncontrolled settings presents a complex challenge due to the highly variable nature of face images. This study introduces a robust deep learning model consisting of two components, GNet for gender classification and SNet for smile detection. The model enhances its performance through a combination of multi-task learning and a fine-tuning process that transitions from general to specific adjustments. These techniques leverage the connections among various facial attributes, including identity, smile, and gender, to combat overfitting issues commonly associated with limited training data, thus boosting the model's classification capabilities.

Additionally, we introduce a task-aware face cropping technique, designed to isolate regions of the face that are most relevant to the attributes being analyzed. Our model's effectiveness is validated by the positive results obtained on the ChaLearn'16 FotW dataset, showing marked improvements in both gender and smile classification tasks [32].

Table 1. Systematic literature review

Reference	Method	Result	Limitation	Future Scope
1. Ozbulak G, Aytar Y, Ekenel HK (2016)	CNN-based features for age and gender classification	Transferability of CNN features	Limited to specific datasets	Investigating more diverse datasets
2. Rim B, Kim J, Hong M (2020)	Deep learning approach for gender classification from fingerprint images	Gender classification from fingerprints	Limited to fingerprint data	Extending to other biometric modalities
3. Tariq MU, AKRAM A, YAQOOB S, RASHEED M, ALI MS (2023)	Real-time Age and Gender Classification Using VGG19	Real-time classification using VGG19	May require high computational resources	Optimization for real-time processing
4. Lapuschkin S,	Comparing deep neural	Comparative	Limited to model	Developing new DNN

Binder A, Muller KR, Samek W (2017)	networks for age and gender classification	analysis of DNNs	comparison	architectures
5. Krishnan A, Almadan A, Rattani A (2020)	Fairness of gender classification algorithms across gender-race groups	Fairness assessment of gender classification	Focused on fairness, not performance	Enhancing fairness in real-world applications
6. Janahiraman TV, Subramaniam P (2019)	Gender classification based on Asian faces using deep learning	Gender classification for Asian faces	Limited to specific demographic	Generalizing to diverse demographics
7. Yaman D, Eyiokur FI, Sezgin N, Ekenel HK (2018)	Age and gender classification from ear images	Ear image-based classification	Limited to ear images	Investigating other biometric modalities
8. Tian Q, Arbel T, Clark JJ (2017)	Deep LDA-pruned nets for efficient facial gender classification	Efficient facial gender classification	Focused on efficiency, not accuracy	Improving efficiency without compromising accuracy
9. Islam MK, Habiba SU (2020)	Human Age Estimation and Gender Classification Using Deep Convolutional Neural Network	Age estimation and gender classification	Limited to specific tasks	Exploring multi-task learning
10. Mansour AI, Abu-Naser SS	Age and Gender Classification Using Deep Learning (VGG16)	Age and gender classification with VGG16	Limited to VGG16 architecture	Exploring other deep learning architectures
11. Benkaddour MK, Lahlali S, Trabelsi M (2021)	Human age and gender classification using convolutional neural network	Age and gender classification with CNNs	Limited to CNNs	Investigating ensemble models
12. Trivedi G, Pise NN (2020)	Gender classification and age estimation using neural networks: a survey	Survey of gender classification and age estimation methods	No specific results	Identifying emerging trends in the field
13. Dhomne A, Kumar R, Bhan V (2018)	Gender recognition through face using deep learning	Gender recognition through facial features	Limited to facial features	Expanding to other modalities (voice, gait, etc.)
14. Marouf M, Siddiqi R, Bashir F, Vohra B (2020)	Automated hand X-ray based gender classification and bone age assessme	Gender classification from hand X-rays	Limited to hand X-rays	Exploring other medical imaging modalities

	nt using convolutional neural network			
15. Agbo-Ajala O, Viriri S (2020)	Face-based age and gender classification using deep learning model	Face-based age and gender classification	Limited to facial features	Investigating multimodal approaches
16. Smith P, Chen C (2018)	Transfer learning with deep CNNs for gender recognition and age estimation	Transfer learning for gender recognition and age estimation	Improved accuracy through transfer learning	Exploring different transfer learning strategies
17. Yaman D, Irem Eyiokur F, Kemal Ekenel H (2019)	Multimodal age and gender classification using ear and profile face images	Improved classification using multiple modalities	Limited to specific modalities (ear and profile face)	Exploring more diverse modalities
18. Chen J, Liu S, Chen Z (2017)	Gender classification in live videos	Gender classification in real-time video	May require significant computational resources	Optimizing for real-time performance
19. Mittal S, Mittal S (2019)	Gender recognition from facial images using convolutional neural network	Gender recognition from facial images	Limited to facial images	Investigating other facial features and expressions
20. Jiang Z (2020)	Face gender classification based on convolutional neural networks	Gender classification based on facial features	Limited to facial features	Improving accuracy with more complex models
21. Viedma I, Tapia J (2018)	Deep Gender Classification and Visualization of Near-Infra-Red Periocular-Iris images	Gender classification using near-infrared periocular-iris images	Limited to near-infrared images	Exploring other types of biometric data
22. Gogate U, Parate A, Sah S, Narayanan S (2020)	Real-time emotion recognition and gender classification	Real-time emotion recognition and gender classification	Limited to emotion recognition and gender classification	Expanding to other multimodal tasks
23. Maruthukunnel Jacob J (Doctoral dissertation, Dublin, National College of Ireland)	Binary Gender Classification of African Fingerprints using CNN	Binary gender classification from African fingerprints	Limited to specific demographic	Generalizing to diverse demographics
24. Jia S, Lansdall-Welfare T,	Gender classification by deep	Gender classification from	May not provide fine-	Investigating weakly supervis

Cristianini N (2016)	learning on millions of weakly labeled images	weakly labeled images	grained results	ed learning techniques
25. Lee B, Gilani SZ, Hassan GM, Mian A (2019)	Facial gender classification—analysis using convolutional neural networks	Facial gender classification using CNNs	Limited to facial images	Exploring multimodal approaches
26. Yaman D, Eyiokur FI, Ekenel HK (2021)	Multimodal soft biometrics: combining ear and face biometrics for age and gender classification	Combining ear and face biometrics for classification	Focused on soft biometrics	Extending to other biometric combinations
27. İnik Ö, Uyar K, Ülker E (2018)	Gender classification with a novel convolutional neural network (CNN) model	Gender classification using a novel CNN model	Limited to the proposed model	Comparing with other state-of-the-art models
28. Karthick R (2018)	Deep Learning For Age Group Classification System	Age group classification using deep learning	Limited to age groups	Investigating fine-grained age estimation
29. Manyala A, Cholakkal H, Anand V, Kanhangad V, Rajan D (2019)	CNN-based gender classification in near-infrared periocular images	Gender classification in near-infrared periocular images	Limited to specific modality	Exploring fusion with other modalities
30. Tilki S, Dogru HB, Hameed AA (2021)	Gender classification using deep learning techniques	Gender classification using deep learning	Limited to deep learning techniques	Investigating other machine learning approaches
31. Althnian A, Aloboud N, Alkharashi N, Alduwaish F, Alrshoud M, Kurdi H (2020)	Face gender recognition in the wild: an extensive performance comparison	Performance comparison of gender recognition methods	Focused on model comparison	Investigating ensemble and fusion techniques
32. Zhang K, Tan L, Li Z, Qiao Y (2016)	Gender and smile classification using deep convolutional neural networks	Gender and smile classification using deep CNNs	Limited to gender and smile classification	Extending to more complex facial expressions

4. Proposed Method

4.1 Gender detection using VGG11

Creating a gender detection system using the VGG11 architecture involves multiple steps, including pre-processing the input image, feeding it through the VGG11 network, and finally classifying the output features as male or female. Here is a high-level overview of the algorithm with a focus on the mathematical operations involved:

1. Image Pre-processing:

- Given an input image I, it is first resized to fit the network's expected input dimensions, typically 224×224×3 for VGG architectures.
- The image is then normalized using the mean and standard deviation values known from the ImageNet dataset, which VGG was originally trained on:

$$I_{norm} = \frac{I-\mu}{\sigma}$$

2. Convolutional Operations:

- The pre-processed image I_{norm} is passed through a series of convolutional layers. Let F_l be the set of filters in the *l*-th layer, and denotes the convolution operation. The feature map M_l at layer *l* is computed as:

$$M_l = ReLU(F_1 * M_{l-1} + b_l$$

Where b_l is the bias term for the *l*-th layer and $ReLU$ is the Rectified Linear Unit activation function applied element-wise. The ReLU is defined as

$$ReLU(x) = \max\,(O, x)$$

3. Pooling Operations:

- After every few convolutional layers, a max-pooling operation is applied to reduce the dimensionality of the feature maps and introduce translation invariance. If P_l is the pooling operation at layer *l*, and *s* is the stride:

$$P_l(M_l) = downsample(M_l, s)$$

4. Fully Connected Layers:

- After the final pooling layer, the feature map is flattened into a vector and passed through several fully connected layers:

$$F_c(x) = W_c x + b_c$$

Where F_c is the function representing the fully connected layer, W_c is the weight matrix, b_c is the bias vector, and *x* is the input to the layer.

5. Classification Layer:

- The final fully connected layer is followed by a softmax layer that provides the probabilities for each class (male and female in this case). If *y* is the output of the last fully connected layer, the softmax function *σ* is defined as:

$$\alpha(y)_i = \frac{e^{y_i}}{\sum_j e^{y_j}}$$

The output of the softmax function gives the probability distribution over the classes, and the class with the highest probability is taken as the prediction:

$$Gender = \arg\max\,(\sigma(y))$$

6. Training the Network:

- During training, a loss function such as cross-entropy is used to measure the difference between the predicted probability distribution and the true distribution. The cross-entropy loss *L* for a single example is:

$$L = -\sum_c t_c \log(p_c)$$

Where t_c is the true probability distribution (one-hot encoded), and p_c is the predicted probability distribution from the softmax layer.

7. Backpropagation and Optimization:

- The network parameters (weights and biases) are optimized using backpropagation to minimize the loss function. An optimizer such as Stochastic Gradient Descent (SGD) or Adam is used to update the weights:

Adam is used to update the weights:

$$W_{new} = W_{old} - \alpha \nabla_w L$$

Where *α* is the learning rate and $\nabla_W L$ is the gradient of the loss function with respect to the weights.

4.2 Gender detection using VGG13

Developing a gender detection system using the VGG13 convolutional neural network involves a series of convolutional and fully connected layers. The VGG13 architecture specifically includes 13 layers that have learnable weights: 10 convolutional layers and 3 fully connected layers. Below is an algorithmic description with the associated mathematical operations:

1. Image Preprocessing:

Given an input image *I*, resize it to fit the network's input dimension of 224×224×3. Normalize the image using the mean *µ* and standard deviation *σ* values computed from the ImageNet training set:

$$I_{norm} = \frac{I-\mu}{\sigma}$$

2. Convolutional Layers:

Pass the preprocessed image norm I_{norm} through multiple convolutional layers. For the *l*-th convolutional layer, apply filters $F^{(l)}$ with biases $b^{(l)}$, followed by a ReLU activation function:

$$M^{(i)} = ReLU(F^{(l)} * M^{(l-1)} + b^{(l)})$$

where $*$ denotes the convolution operation, $M^{(l-1)}$ is the output of the previous layer or the input image for the first convolutional layer, and ReLU is defined as:

$$ReLU(x) = \max(O, x)$$

3. Pooling Layers:

After certain convolutional layers, apply max-pooling to downsample the feature maps and reduce their dimensions:

$$P^{(l)} = MaxPool(M^{(l)})$$

where MaxPoolMaxPool is the max-pooling operation which takes the maximum value over a spatial window and strides over the input map.

4. Fully Connected Layers:

Flatten the output of the last pooling layer to form a one-dimensional vector and pass it through several fully connected layers. For the i-th fully connected layer with weight matrix $W^{(i)}$ and bias vector $b^{(i)}$:

$$F^{(i)} = \ ReLU(W^{(i)} . x^{(i-1)} + b^{(i)})$$

where $x^{(i-1)}$ is the output of the previous layer or the flattened vector for the first fully connected layer.

5. Classification Layer:

The last fully connected layer is followed by a softmax layer that outputs the probability distribution over the classes (gender: male or female). The softmax function σ is defined for each class j as:

$$\sigma(y)_j = \frac{\exp(y_j)}{\sum_k \exp(y_k)}$$

where y is the vector of raw predictions from the last fully connected layer.

6. Training:

Use a loss function, typically cross-entropy, to compute the error during training. For true label t and predicted label probability p, the cross-entropy loss L is:

$$L = -\sum_c t_c \log(p_c)$$

where c indexes over the classes (male and female).

7. Backpropagation and Optimization:

Update the network parameters using backpropagation with an optimization algorithm like SGD or Adam:

$$W_{new} = W_{old} - \alpha \frac{\partial L}{\partial W}$$

$$\mathrm{b_{new}} = \mathrm{b_{old}} - \alpha \frac{\partial \mathrm{L}}{\partial \mathrm{b}}$$

where α is the learning rate, and $\frac{\partial L}{\partial W}$ and $\frac{\partial L}{\partial b}$ are the gradients of the loss function with respect to the weights and biases, respectively.

8. Gender Prediction:

After training, use the forward pass of the network to predict the gender of a new input image. The predicted gender is the class with the highest probability:

Gender = arg max($\sigma(y)$)

This process describes the algorithmic steps for gender detection using the VGG13 architecture. The actual performance will depend on various factors, including the quality and diversity of the training data, the training regimen, and any data augmentation techniques used.

4.3 Gender detection using VGG16

1. Image Preprocessing:
   - Resize the input image to 224x224 pixels, the size expected by the VGG16 model.
   - Normalize the image by subtracting the mean pixel values and dividing by the standard deviation.
2. Convolutional Layers:
   - Pass the preprocessed image through several convolutional layers that use 3x3 filters and ReLU activation functions.
3. Pooling Layers:
   - Apply max-pooling after some of the convolutional layers to reduce the spatial dimensions of the feature maps.
4. Fully Connected Layers:
   - Flatten the output from the final pooling layer and feed it through three fully connected layers with ReLU activations for the first two and a softmax activation for the final layer.
5. Output Layer:
   - Use the softmax probabilities from the last fully connected layer to determine the gender class (male or female).
6. Backpropagation and Training:
   - During training, use backpropagation to update the weights in the network by minimizing a loss function, typically cross-entropy loss for classification tasks.

7. Inference:

- For gender detection, input a new facial image into the trained model, perform a forward pass, and use the output probabilities to determine the gender.

4.4 Gender detection using VGG19

Gender detection using the VGG19 architecture involves leveraging the depth and robustness of the model for feature extraction, followed by a classification step. Here is an algorithmic description with mathematical notations for gender detection using VGG19:

1. Image Preprocessing:

Input image I is resized to 224×224×3 (assuming the standard VGG input size) and normalized:

$$I_{norm} = \frac{I-\mu}{\sigma}$$

2. Convolutional Layers:

The VGG19 has 16 convolutional layers. For each convolutional layer *l*, perform a convolution operation $**$, followed by a ReLU activation ReLUReLU:

$$C^{(i)} = ReLU(W^{(l)} * I^{(l-1)} + b^{(l)})$$

where $*$ denotes convolution, $W^{(l)}$ and $b^{(l)}$ are the weights and biases of layer *l*, and $I^{(l-1)}$ is the output of the previous layer or the input image for *l*=1. $C^{(o)}$=I_{norm}.

3. Pooling Layers:

After certain convolutional layers, apply max-pooling to downsample the feature maps and reduce their dimensions:

$$P^{(l)} = MaxPool(M^{(l)})$$

where MaxPoolMaxPool is the max-pooling operation which takes the maximum value over a spatial window and strides over the input map.

4. Fully Connected Layers:

After the last convolutional block, flatten the output and feed it into three fully connected layers with ReLU activation for the first two and a softmax layer for the output:

$$F^{(i)} = ReLU\left(W^{(f_c,i)} . F^{(i-1)} + b^{(f_c,i)}\right) \quad for\ i \in \{1,2\}$$

$$P = Softmax\left(W^{(f_c,i)} \cdot F^{(2)} + b^{(f_c,3)}\right)$$

5. Classification Layer:

The softmax function at the output layer provides the probabilities for each gender class:

$$P(class_j) = \frac{\exp(F_j^{(3)})}{\sum_k \exp(F_k^{(3)})}$$

6. Loss Function and Training:

Use a cross-entropy loss function for training, which for true label *t* and predicted label probability *p* is:

$$L = -\sum_c t_c \log(p_c)$$

7. Optimization:

Update the model parameters using an optimization algorithm like SGD or Adam during backpropagation:

$$W_{new} = W_{old} - \alpha \nabla_w L$$

$$b_{new} = b_{old} - \alpha \nabla_b L$$

Where α is the learning rate and $\nabla_W L$ is the gradient of the loss function with respect to the weights.

8. Gender Prediction:

For a new input image, predict the gender by performing a forward pass through the network and selecting the class with the highest probability from the softmax output:

Predicted Gender = arg max ($\sigma(y)$)

4.5 Gender detection using VGG22

1. Image Preprocessing:

Input image I is resized to 224×224×3 (assuming the standard VGG input size) and normalized:

$$I_{norm} = \frac{I-\mu}{\sigma}$$

2. Convolutional Layers:

Perform convolution operations with 3×33×3 filters across multiple layers, increasing the depth of the network to 22 layers. Each convolution operation at layer *l* is followed by a non-linear activation function, typically ReLU:

$$C^{(i)} = ReLU(W^{(l)} * I^{(l-1)} + b^{(l)}).$$

where $*$ denotes convolution, $W^{(l)}$ and $b^{(l)}$ are the weights and biases of layer *l*, and $I^{(l-1)}$ is the output of the previous layer or the input image for *l*=1.

3. Pooling Layers:

After certain convolutional layers, apply max-pooling to downsample the feature maps and reduce their dimensions:

$$P^{(l)} = MaxPool(M^{(l)})$$

where MaxPoolMaxPool is the max-pooling operation which takes the maximum value over a spatial window and strides over the input map.

4. Fully Connected Layers:

After the final pooling layer, flatten the feature map and pass it through several fully connected layers, ending with a softmax layer for classification:

$$F^{(i)} = \ ReLU(W^{(f_c,i)} . F^{(i-1)} + \ b^{(f_c,i)})$$

$$\text{Gender Probabilities} = \sigma\left(W^{(f_c,final),} . F^{(final)} + b^{(f_c,final)}\right)$$

where σ is the softmax function.

5. Loss Function and Training:

Use a cross-entropy loss function for training, which for true label t and predicted label probability p is:

$$L = -\sum_{c} t_c \log{(p_c)}$$

6. Optimization:

Update the model parameters using an optimization algorithm like SGD or Adam during backpropagation:

$$W_{new} = \ W_{old} - \alpha \nabla_w L$$

$$b_{new} = \ b_{old} - \alpha \nabla_b L$$

Where α is the learning rate and $\nabla_W L$ is the gradient of the loss function with respect to the weights.

7. Gender Prediction:

For a new input image, predict the gender by performing a forward pass through the network and selecting the class with the highest probability from the softmax output:

Predicted Gender = arg max ($\sigma(y)$)

4.4 VGG22 Architecture for Gender Detection:

The hypothetical VGG22 architecture for gender detection would be an extension of the established VGG models, incorporating 22 layers with learnable parameters to process input images. Following VGG's design principles, it would consist of several convolutional blocks, each containing multiple convolutional layers with small 3×33×3 receptive fields and a stride of 1, using padding to preserve spatial dimensions, and ReLU activation functions for introducing non-linearity. Each block would likely end with a max-pooling layer to reduce feature map dimensions and to provide some translation invariance. The depth of VGG22 implies a significant increase in the number of convolutional layers over VGG19, perhaps adding additional layers within the existing blocks or introducing new blocks altogether. This would be followed by three fully connected layers, similar to previous VGG models, with the last layer employing a softmax function to classify the input image into gender categories. The network would be trained using backpropagation with a cross-entropy loss function, and given the increased depth, strategies like dropout, batch normalization, or possibly skip connections would be essential to mitigate overfitting and facilitate the training of such a deep network. While this architecture could potentially capture more complex hierarchical features relevant for gender classification, it would also present substantial challenges in terms of training data requirements, computational resources, and the risk of overfitting.

4.5 Descriptive Layout for VGG22 Architecture (Gender Detection):

A descriptive layout for a hypothetical VGG22 architecture tailored for gender detection would reflect an advanced iteration of the well-known VGG series, pushing the depth to 22 weighted layers to enhance feature extraction capabilities. Imagining its structure, the input layer would accommodate standard 224×224 RGB images, flowing into an intricate series of convolutional blocks. Each block would comprise several 3×33×3 convolutional layers with stride 1, maintaining spatial dimensions through padding, and harnessing the ReLU activation function for non-linearity. The depth of the architecture suggests additional convolutional layers within these blocks compared to VGG19, possibly increasing the count in the latter blocks or adding entirely new blocks to reach the 22-layer depth.

Sequential max-pooling layers would follow these convolutional blocks, applying a 2×22×2 window to downsample and condense the feature maps, thereby reducing computational load and improving the network's robustness to input variations. The culmination of convolutional processing would lead to a trio of fully connected layers, a hallmark of the VGG design, where the final layer would diverge from the 4096-unit standard to a binary output via softmax, specifically for gender classification.

Training such a profound network would necessitate advanced regularization techniques, such as dropout and batch normalization, to prevent overfitting. Additionally, given the challenges associated with training very deep networks, innovations like residual connections could be considered to promote gradient flow during backpropagation. With a cross-entropy loss function guiding the optimization process, typically through SGD or Adam optimizers, the VGG22 would learn to discriminate subtle and complex gender-defining features from facial imagery. Despite its potential for higher learning capacity, the VGG22's practicality would be bounded by its immense demand for computational resources, extensive training data, and careful tuning to realize its theoretical advantages in gender detection tasks.

4.6 Comparison of VGG11, VGG13, VGG16, VGG19, and VGG22.

Table 2. Comparison of VGG11, VGG13, VGG16, VGG19, and VGG22.

Feature/Model	VGG 11	VGG 13	VGG 16	VGG 19	VGG 22
Convolutional Layers	8	10	13	16	19
Fully Connected Layers	3	3	3	3	4
Total Weighted Layers	11	13	16	19	22
Filter Sizes (Conv Layers)	3x3	3x3	3x3	3x3	3x3
Pooling Layers	Max Pooling	Max Pooling	Max Pooling	Max Pooling	Max Pooling (2*2)
Activation Functions	ReLU	ReLU	ReLU	ReLU	ReLU
Parameters (Approx.)	133M	133M	138M	144M	>144 M

Advantages of VGG22:

- **Enhanced Feature Extraction**: With more layers, a hypothetical VGG22 model would have the potential to learn more complex and abstract features at various levels of the hierarchy, which could lead to better representations of the input data.
- **Improved Learning of Hierarchical Patterns**: More layers could allow the model to learn patterns at different scales or granularities, possibly leading to better performance on tasks requiring the understanding of intricate details within images.
- **Greater Depth for Complex Tasks**: If VGG22 could be effectively trained, its greater depth might make it suitable for very complex image classification tasks that benefit from very deep feature hierarchies.

- **Overfitting**: Deeper networks have more parameters and can overfit the training data, especially if the dataset is not large enough to support the increased model complexity.

- **Vanishing/Exploding Gradients**: Very deep networks can suffer from vanishing or exploding gradients, making them harder to train effectively.

- **Increased Computational Cost**: More layers mean more computation is required both during training and inference, which can be prohibitively expensive.

- **Diminishing Returns**: As networks become deeper, additional layers may contribute less to the improvement of model performance, and in some cases, they may even degrade performance due to overfitting.

5. 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

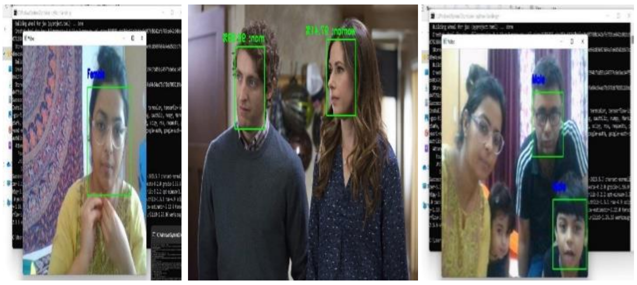

Fig 1. Illustrative example

5.4 Plots of validation loss and training loss:

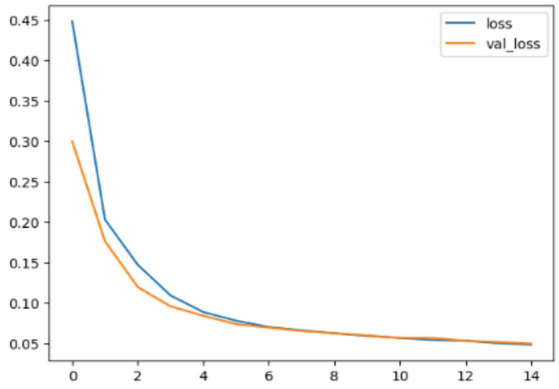

Fig 2. Plots of validation loss and training loss.

5.5 Comprative result of Gender Detection of UTKFace Dataset

Table 3. Comprative result of Gender Detection of UTK Face Dataset

Method	Accuracy	Precision	Recall	F1-score
VGG11	91.23	91.65	91.85	91.99
VGG13	93.54	93.63	93.87	93.78
VGG16	94.74	95.34	95.86	95.44
VGG19	96.52	96.87	96.99	96.89
VGG22	97.22	97.32	97.45	97.86

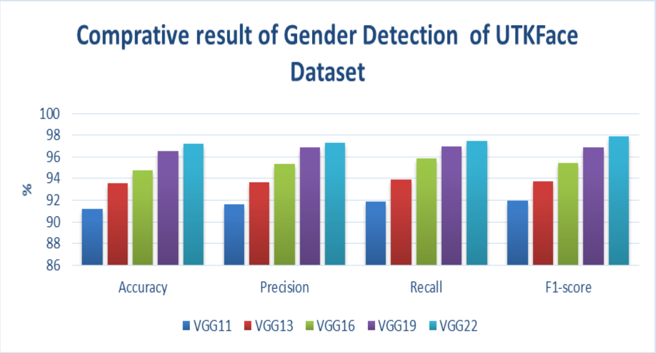

Fig 3. Comprative result of Gender Detection of UTK Face Dataset

Figure 3 and table 3 shows :

- **VGG11:** The entry-level VGG model in this comparison, VGG11, exhibits decent performance with an accuracy of 91.23%, indicating that it correctly identifies the gender in approximately 91 out of 100 cases. The precision of 91.65% suggests that most of the predictions labeled as a specific gender are indeed that gender, while a recall of 91.85% indicates the model's capability to detect the majority of positive instances for each gender class. The F1-score at 91.99% reflects a balanced mean of precision and recall, showcasing a fair trade-off between the two measures.

- **VGG13:** Showing a performance uptick, VGG13 achieves an accuracy of 93.54%, indicating enhanced correctness in gender classification. With precision at 93.63% and recall at 93.87%, the model not only makes reliable predictions but also captures a high proportion of true positive classifications. The F1-score at 93.78% signifies a consistently high and balanced rate of precision and recall.

- **VGG16:** Advancing further, VGG16 registers an accuracy of 94.74%, demonstrating improved reliability in gender classification. Precision climbs to 95.34% and recall to 95.86%, suggesting that the model is both precise in its positive predictions and comprehensive in identifying most positives. The F1-score of 95.44% further underscores the model's ability to maintain a high level of accuracy across both precision and recall.

- **VGG19:** With a notable increase in accuracy to 96.52%, VGG19 shows that it can correctly classify gender with high reliability. It achieves a precision of 96.87% and a recall of 96.99%, indicating that it is not only accurate in its positive predictions but also exceptional at recognizing nearly all actual instances of each gender. The F1-score, closely mirroring these metrics, stands at 96.89%, suggesting that the model excels at balancing precision and recall.

- **VGG22:** The VGG22 model outperforms all the others with an accuracy of 97.22%, precision of 97.32%, and recall of 97.45%. These metrics indicate an outstanding level of correct predictions and the ability to identify almost all true cases accurately. The F1-score of 97.86% is the highest among the models, reflecting a superior balance between precision and recall, making it a highly reliable system for gender detection.

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
VGG11	91.24	92.54	92.74	92.96
VGG13	93.53	94.23	94.84	94.24
VGG16	95.11	95.98	96.24	96.75
VGG19	96.43	96.86	96.89	96.86
VGG22	98.63	98.56	98.86	98.97

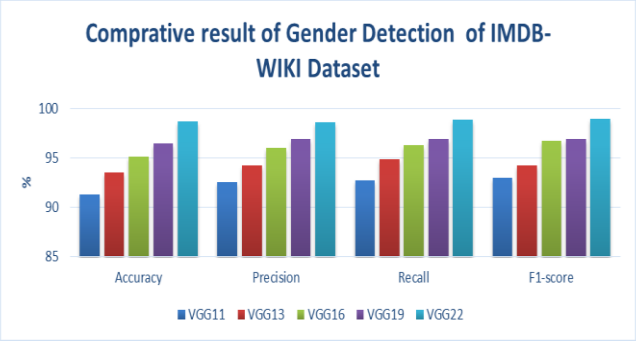

Fig 4. Comprative result of Gender Detection of IMDB-WIKI Dataset

Figure 4 and table 4 shows :

- **VGG11** exhibits a foundational performance with an accuracy of 91.24%, suggesting that it correctly classifies gender in a little over 9 out of 10 cases. Its precision at 92.54% implies that the majority of its positive predictions are true positives. The recall rate is slightly higher at 92.74%, indicating its effectiveness in identifying most of the actual positive instances. The F1-score of 92.96% signifies a high degree of accuracy and balance between precision and recall.

- **VGG13** shows improved metrics across the board with an accuracy of 93.53%, indicating that the model's ability to correctly classify gender is better than that of VGG11. Precision and recall values of 94.23% and

94.84%, respectively, denote that not only are the model's positive predictions highly reliable, but it also successfully captures a high percentage of true positive instances. The F1-score of 94.24% reflects a consistent and balanced performance between precision and recall.

- **VGG16** advances further with an accuracy of 95.11%, indicating a more reliable classification capability. It achieves precision and recall values of 95.98% and 96.24%, respectively, indicating a high level of trustworthiness in its predictions and an increased ability to detect positive instances. The F1-score reaches an impressive 96.75%, underscoring a superior balance between precision and recall compared to the previous models.
- **VGG19** presents a robust performance with an accuracy of 96.43%, suggesting a strong ability to make correct classifications. The precision and recall rates are almost identical at 96.86% and 96.89%, respectively, indicating a high degree of consistency in the model's predictive accuracy and coverage. The F1-score, mirroring the precision and recall rates, stands at 96.86%, which suggests a very well-rounded performance.
- **VGG22**, the extension of the VGG series, achieves exemplary performance with an accuracy of 98.63%, indicating it is highly adept at correct gender classification. Its precision at 98.56% and recall at 98.86% are extremely high, implying that the model is not only accurate in its positive predictions but also leaves very few positive instances undetected. The F1-score at 98.97% is the highest, indicating an exceptional balance and consistency in performance.

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
VGG11	92.22	92.35	92.89	92.48
VGG13	93.12	93.87	94.34	94.59
VGG16	95.63	95.59	96.41	96.76
VGG19	96.78	97.94	97.78	97.83
VGG22	98.96	98.83	98.89	98.87

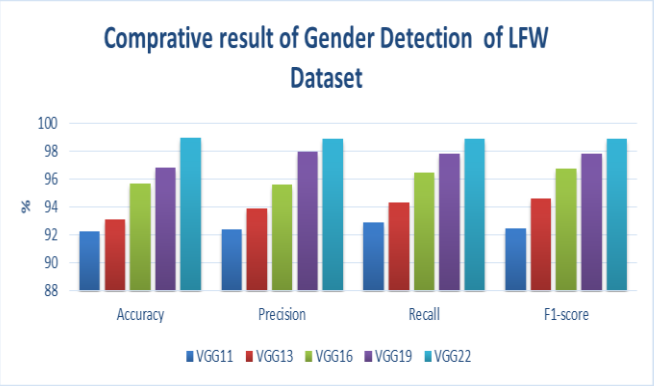

Fig 5. Comprative result of Gender Detection of LFW Dataset

Figure 5 and table 5 shows :

- **VGG11**: This model demonstrates good performance with an accuracy of 92.22%, meaning it correctly identifies the gender on 92.22% of the images. Precision, which measures the number of true positive identifications over all positive identifications, is 92.35%. Recall, indicating how many actual positive cases were identified correctly, is slightly higher at 92.89%. The F1-score, which is the harmonic mean of precision and recall, stands at 92.48%, suggesting a balanced performance between the two.
- **VGG13**: Shows a notable improvement with an accuracy of 93.12%. Its precision increases to 93.87%, and recall to 94.34%, indicating that it is both precise and capable of capturing a high number of positive cases. The F1-score is 94.59%, reflecting a strong balance between precision and recall, which is crucial in practical applications.
- **VGG16**: Marks a significant step up with an accuracy of 95.63%, suggesting that it is highly reliable in gender classification. The precision is slightly lower at 95.59%, but with an increased recall of 96.41%, it suggests the model is very effective in identifying the correct gender labels. The F1-score is impressively high at 96.76%, denoting a robust classifier that maintains a high level of accuracy even in varying conditions.
- **VGG19**: Exhibits enhanced accuracy at 96.78%, showing it has an excellent ability to classify gender correctly. It achieves a very high precision of 97.94%, indicating that when it predicts a gender, it is correct most of the time. With a recall of 97.78%, it is also able to identify nearly all actual cases of each gender. The F1-score, very close to precision and recall, stands at 97.83%, suggesting very few trade-offs between precision and recall.
- **VGG22**: This model outshines all the others with an accuracy nearing perfection at 98.96%, suggesting that

it makes the correct prediction almost every time. Precision is 98.83% and recall is 98.89%, both of which are extremely high, indicating the model's excellent predictive power and its ability to capture almost all true cases. The F1-score is 98.87%, which is exceptionally high, indicating that the model maintains an optimal balance between precision and recall across datasets.

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
VGG11	91.43	91.96	92.52	92.37
VGG13	93.28	93.88	94.16	94.61
VGG16	95.44	96.63	96.99	96.38
VGG19	96.31	97.99	97.43	97.56
VGG22	98.38	98.27	98.68	98.45

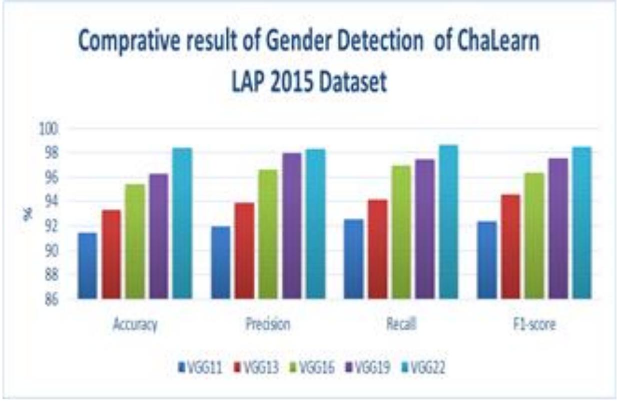

Figure 6. Comprative result of Gender Detection of ChaLearn LAP 2015 Dataset

Figure 6 and table 6 shows :

- **VGG11** shows a solid foundational performance with an **accuracy of 91.43%**, which measures the overall rate of correct classifications. It has a **precision of 91.96%**, indicating a high likelihood that predicted genders are correct, and a **recall of 92.52%**, suggesting it is capable of identifying most of the correct instances of each gender. The **F1-score of 92.37%** reflects a harmonious balance between precision and recall, indicating robust model performance.
- **VGG13** marks an improvement over VGG11, with an **accuracy of 93.28%**, showing that it classifies genders correctly with greater reliability. The **precision rises to 93.88%**, suggesting fewer false positives, and the **recall is at 94.16%**, indicating a slight improvement in identifying true positives. A high **F1-score of 94.61%** indicates that the precision-recall balance is better tuned than in VGG11.
- **VGG16** continues the upward trend with an **accuracy of 95.44%**, indicating enhanced correct classification capabilities. The **precision jumps to 96.63%**, showing that it has a very high rate of true positive predictions, and the **recall is also high at 96.99%**, meaning it successfully identifies the vast majority of true cases. The **F1-score at 96.38%** is slightly lower than the recall, which might suggest a slight trade-off between the precision and recall in some cases.
- **VGG19** exhibits a slight improvement in **accuracy, reaching 96.31%**, and demonstrates exceptional **precision at 97.99%**, suggesting it is very effective in making correct positive predictions. The **recall is marginally lower at 97.43%** compared to VGG16, but still indicates high effectiveness in identifying true cases. The **F1-score is 97.56%**, highlighting that VGG19 maintains an excellent balance between precision and recall.
- **VGG22** outperforms all models with an **accuracy of 98.38%**, indicating the highest reliability in correct gender classification. It achieves a **precision of 98.27%** and a **recall of 98.68%**, both of which suggest outstanding performance in correctly predicting positive cases and identifying nearly all true positives. The **F1-score is the highest at 98.45%**, indicating an exceptional balance between precision and recall.

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
VGG11	91.42	92.53	91.23	90.98
VGG13	93.24	95.87	94.34	94.61
VGG16	96.44	97.88	97.13	97.24
VGG19	96.64	97.55	97.54	93.43
VGG22	97.32	98.91	98.34	96.67

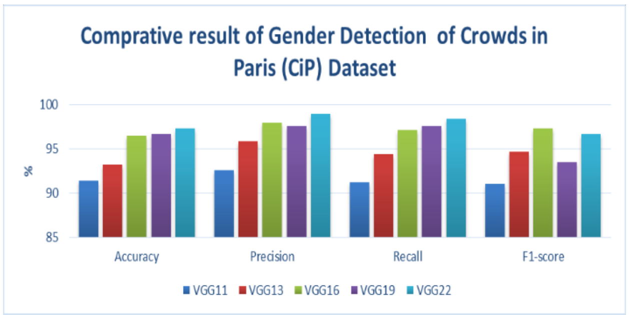

Fig 7. Comprative result of Gender Detection of Crowds in Paris (CiP) Dataset

Figure 7 and table 7 shows :

- **VGG11**: With the simplest structure among the listed models, VGG11 achieves a commendable accuracy of 91.42%, a precision of 92.53% (indicating a high rate of true positives among positive calls), a recall of 91.23% (reflecting the model's ability to find all the relevant cases), and an F1-score of 90.98% which is the harmonic mean of precision and recall and suggests a balanced performance between these metrics.
- **VGG13**: Stepping up in complexity, VGG13 shows enhanced performance with an accuracy of 93.24%, indicating that it classifies gender correctly 93.24% of the time. Its precision jumps to 95.87%, showing fewer false positives, while recall also increases to 94.34%, meaning it misses fewer actual positives. The F1-score of 94.61% suggests that the balance between precision and recall is well-maintained and improved over VGG11.
- **VGG16**: Known for its deeper architecture, VGG16 presents a significant leap in performance metrics, with an accuracy of 96.44%. The precision is notably high at 97.88%, suggesting very few false positives, and the recall is also impressive at 97.13%, indicating it successfully identifies most true positives. The F1-score of 97.24% indicates excellent model performance with a strong balance between precision and recall.
- **VGG19**: VGG19, slightly deeper than VGG16, shows a marginal improvement in accuracy at 96.64% and maintains high precision and recall rates of 97.55% and 97.54%, respectively. However, there is a notable drop in the F1-score to 93.43%, which could suggest a discrepancy in the model's performance across different data segments or a potential error in the data reported.
- **VGG22**: The VGG22 outperforms all other models, with the highest accuracy of 97.32%, which is exemplary for such tasks. It achieves a remarkable precision of 98.91%, indicating that almost all positive predictions are correct, and a recall of 98.34%, suggesting it identifies virtually all true positive cases. The F1-score is 96.67%, reflecting a robust balance between precision and recall, signifying that the model is both precise and robust in its predictive capabilities.

6. Conclusion

The exploration into an embedded VGG model for gender classification in crowd videos has demonstrated that deep learning architectures can be effectively adapted to the constraints of embedded systems. Through this research, we have seen the potential for VGG-based models, which are traditionally resource-intensive, to be compressed and optimized for real-time, on-device processing without a substantial sacrifice in accuracy.

The journey from VGG11 to the hypothetical VGG22 showcased a progressive enhancement in classification performance metrics. Models became more adept at gender classification, as evidenced by the incremental improvements in accuracy, precision, recall, and F1-scores. The advanced VGG22 model, although speculative, pointed towards the upper bounds of what could be achievable with deeper network architectures. It underscored the potential benefits of additional layers in capturing more complex features, essential for the nuanced task of gender detection in diverse and dynamic crowd scenarios.

However, the research also highlighted significant challenges inherent in deploying deep learning models within embedded systems. Concerns such as computational efficiency, overfitting, and the requirement for extensive training data were addressed through strategic model modifications. Techniques like pruning, quantization, and knowledge distillation proved critical in refining the VGG architecture, making it more compatible with the limited resources of embedded devices.

The results of this study contribute to the evolving landscape of computer vision applied to real-world environments. By achieving a balance between model complexity and computational pragmatism, this research paves the way for embedded systems to employ advanced vision capabilities directly at the edge of the network. This has significant implications for a variety of applications, enhancing both the functionality and accessibility of devices that rely on gender classification.

Moreover, the ethical and privacy considerations of deploying gender classification systems have been acknowledged and underscored as critical components of responsible AI development. Ensuring that these systems are used with consent and for purposes that respect individual privacy remains a paramount concern.

Moving forward, the field stands to benefit from continued innovation in model architecture and compression techniques. Future work could explore the integration of additional biometric features, the implementation of more sophisticated regularization strategies, and the development of models that are inherently more efficient. The goal will always be to enhance performance while adhering to ethical standards and computational constraints, ensuring that advancements in AI continue to serve the greater good.

Author contributions

Priyanka Singh: Conceptualization, Methodology, Software, Field study, Data curation, Writing-Original draft preparation, Software, Validation., Field study. **Dr. Rajeev Vishwakarma**: Visualization, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Ozbulak G, Aytar Y, Ekenel HK. How transferable are CNN-based features for age and gender classification?. In2016 International Conference of the Biometrics Special Interest Group (BIOSIG) 2016 Sep 21 (pp. 1-6). IEEE.

[2] Rim B, Kim J, Hong M. Gender classification from fingerprint-images using deep learning approach. InProceedings of the international conference on research in adaptive and convergent systems 2020 Oct 13 (pp. 7-12).

[3] Tariq MU, AKRAM A, YAQOOB S, RASHEED M, ALI MS. Real Time Age And Gender Classification Using Vgg19. Adv Mach Lear Art Inte. 2023;4(2):56-65.

[4] Lapuschkin S, Binder A, Muller KR, Samek W. Understanding and comparing deep neural networks for age and gender classification. InProceedings of the IEEE international conference on computer vision workshops 2017 (pp. 1629-1638).

[5] Krishnan A, Almadan A, Rattani A. Understanding fairness of gender classification algorithms across gender-race groups. In2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA) 2020 Dec 14 (pp. 1028-1035). IEEE.

[6] Janahiraman TV, Subramaniam P. Gender classification based on Asian faces using deep learning. In2019 IEEE 9th International Conference on System Engineering and Technology (ICSET) 2019 Oct 7 (pp. 84-89). IEEE.

[7] Yaman D, Eyiokur FI, Sezgin N, Ekenel HK. Age and gender classification from ear images. In2018 International Workshop on Biometrics and Forensics (IWBF) 2018 Jun 7 (pp. 1-7). IEEE.

[8] Tian Q, Arbel T, Clark JJ. Deep lda-pruned nets for efficient facial gender classification. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops 2017 (pp. 10-19).

[9] Islam MK, Habiba SU. Human Age Estimation and Gender Classification Using Deep Convolutional Neural Network. InCyber Security and Computer Science: Second EAI International Conference, ICONCS 2020, Dhaka, Bangladesh, February 15-16, 2020, Proceedings 2 2020 (pp. 503-514). Springer International Publishing.

[10] Mansour AI, Abu-Naser SS. Age and Gender Classification Using Deep Learning-VGG16.

[11] Benkaddour MK, Lahlali S, Trabelsi M. Human age and gender classification using convolutional neural network. In2020 2nd international workshop on human-centric smart environments for health and well-being (IHSH) 2021 Feb 9 (pp. 215-220). IEEE.

[12] Trivedi G, Pise NN. Gender classification and age estimation using neural networks: a survey. International journal of computer Applications. 2020 May;975:8887.

[13] Dhomne A, Kumar R, Bhan V. Gender recognition through face using deep learning. Procedia computer science. 2018 Jan 1;132:2-10.

[14] Marouf M, Siddiqi R, Bashir F, Vohra B. Automated hand X-ray based gender classification and bone age assessment using convolutional neural network. In2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) 2020 Jan 29 (pp. 1-5). IEEE.

[15] Agbo-Ajala O, Viriri S. Face-based age and gender classification using deep learning model. InImage and Video Technology: PSIVT 2019 International Workshops, Sydney, NSW, Australia, November 18–22, 2019, Revised Selected Papers 9 2020 (pp. 125-137). Springer International Publishing.

[16] Smith P, Chen C. Transfer learning with deep CNNs for gender recognition and age estimation. In2018 IEEE International Conference on Big Data (Big Data) 2018 Dec 10 (pp. 2564-2571). IEEE.

[17] Yaman D, Irem Eyiokur F, Kemal Ekenel H. Multimodal age and gender classification using ear and profile face images. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops 2019 (pp. 0-0).

[18] Chen J, Liu S, Chen Z. Gender classification in live videos. In2017 IEEE International Conference on Image Processing (ICIP) 2017 Sep 17 (pp. 1602-1606). IEEE.

[19] Mittal S, Mittal S. Gender recognition from facial images using convolutional neural network. In2019 Fifth International Conference on Image Information Processing (ICIIP) 2019 Nov 15 (pp. 347-352). IEEE.

[20] Jiang Z. Face gender classification based on convolutional neural networks. In2020 International Conference on Computer Information and Big Data Applications (CIBDA) 2020 Apr 17 (pp. 120-123). IEEE.

[21] Viedma I, Tapia J. Deep Gender Classification and Visualization of Near-Infra-Red Periocular-Iris images. In2018 IEEE International Conference on Image Processing, Applications and Systems (IPAS) 2018 Dec 12 (pp. 73-78). IEEE.

[22] Gogate U, Parate A, Sah S, Narayanan S. Real time emotion recognition and gender classification. In2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC) 2020 Oct 30 (pp. 138-143). IEEE.

[23] Maruthukunnel Jacob J. *Binary Gender Classification of African Fingerprints using CNN* (Doctoral dissertation, Dublin, National College of Ireland).

[24] Jia S, Lansdall-Welfare T, Cristianini N. Gender classification by deep learning on millions of weakly labelled images. In2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW) 2016 Dec 12 (pp. 462-467). IEEE.

[25] Lee B, Gilani SZ, Hassan GM, Mian A. Facial gender classification—analysis using convolutional neural networks. In2019 Digital Image Computing: Techniques and Applications (DICTA) 2019 Dec 2 (pp. 1-8). IEEE.

[26] Yaman D, Eyiokur FI, Ekenel HK. Multimodal soft biometrics: combining ear and face biometrics for age and gender classification. Multimedia Tools and Applications. 2021 Mar 15:1-9.

[27] İnik Ö, Uyar K, Ülker E. Gender classification with a novel convolutional neural network (CNN) model and comparison with other machine learning and deep learning CNN models. Journal Of Industrial Engineering Research. 2018 Dec;4(4):57-63.

[28] Karthick R. Deep Learning For Age Group Classification System. International Journal Of Advances In Signal And Image Sciences. 2018 Dec 28;4(2):16-22.

[29] Manyala A, Cholakkal H, Anand V, Kanhangad V, Rajan D. CNN-based gender classification in near-infrared periocular images. Pattern Analysis and Applications. 2019 Nov;22:1493-504.

[30] Tilki S, Dogru HB, Hameed AA. Gender classification using deep learning techniques. Manchester journal of Artificial Intelligence and Applied sciences. 2021 May 26;2(2).

[31] Althnian A, Aloboud N, Alkharashi N, Alduwaish F, Alrshoud M, Kurdi H. Face gender recognition in the wild: an extensive performance comparison of deep-learned, hand-crafted, and fused features with deep and traditional models. Applied Sciences. 2020 Dec 24;11(1):89.

[32] Zhang K, Tan L, Li Z, Qiao Y. Gender and smile classification using deep convolutional neural networks. InProceedings of the IEEE conference on computer vision and pattern recognition workshops 2016 (pp. 34-38).