Age and Gender Detection using Facial Images

Esmat Mohamed
Department of Information Technology
Misr University for Science and Technology
Giza, Egypt
esmat.mohamed@must.edu.eg

Ahmed Ashraf
Department of Artificial intelligence
Misr University for Science and Technology
Giza, Egypt
94142@must.edu.eg

Waleed Matar
Department of Artificial intelligence
Misr University for Science and Technology
Giza, Egypt
94184@must.edu.eg

Mohamed Tarek
Department of Artificial intelligence
Misr University for Science and Technology
Giza, Egypt
94148@must.edu.eg

***Abstract*—Fueled by advancements in Machine Learning and Computer Vision, this thesis explores the development of an automated age and gender detection system. It details how machines can be trained to "see" and interpret facial features to achieve human-like recognition accuracy. The applications range from security to marketing, and the core of the system relies on machine learning models that learn from vast datasets to make these predictions. By examining the ethical considerations and future prospects, this thesis dives into the complexities of facial recognition technology.**

Keywords—age and gender detection, computer vision, facial recognition, machine learning

I. INTRODUCTION

Artificial intelligence (AI) is a rapidly evolving field within computer science that focuses on building intelligent systems capable of performing tasks typically associated with human thinking. Machine learning (ML), a crucial subset of AI, empowers computers to learn and improve independently without being explicitly instructed. ML algorithms achieve this by discovering patterns within large datasets, allowing them to make predictions or decisions without strictly defined rules.

Age and gender detection utilizes the power of machine learning within computer vision. By analyzing facial images, these systems can estimate a person's probable age range and gender category. The core concept lies in training algorithms to recognize subtle visual cues that differentiate age groups and genders. Wrinkles, skin texture, and facial hair distribution are just a few of the key features analyzed by these models.

Fig.1. Age Detection. [1]

This research explores the application of advanced computer vision techniques for the precise identification of age and gender in various settings, aiming to enhance digital interactions and tailor experiences across diverse sectors. By analyzing demographic data of event attendees, this study offers event organizers a method to identify key demographic groups, optimizing engagement and service delivery. Additionally, the project extends its utility to security and access control, employing targeted approaches to ensure safety while respecting privacy. Incorporating multidimensional analysis, the research transcends traditional demographic methods, providing a richer, more integrated understanding of user profiles, which is crucial for businesses and establishments in optimizing interactions and operational strategies.

There is a strong push for research in age and gender detection using facial images due to its vast potential applications. This technology offers a non-invasive and automated way to gain insights about a person's demographics from mere images. This motivates research to improve the accuracy and efficiency of such systems. Furthermore, the challenges posed by factors like lighting variations, pose changes, and occlusions in real-world scenarios drive researchers to develop robust algorithms that can handle these complexities [2]. Additionally, the ethical considerations surrounding potential biases in these systems motivate research towards ensuring fairness and inclusivity in age and gender detection models. Overall, the research is driven by the desire to create a powerful tool with wide-ranging applications while ensuring responsible development and efficient deployment.

The project aims to develop an efficient age and gender detection system using computer vision and machine learning techniques. The primary objective is to create a robust algorithm capable of accurately estimating the age and gender of individuals from images or video inputs in real-time. Employing computer vision methodologies, specifically pre-trained CNNs trained on large datasets such as ImageNet. [3]

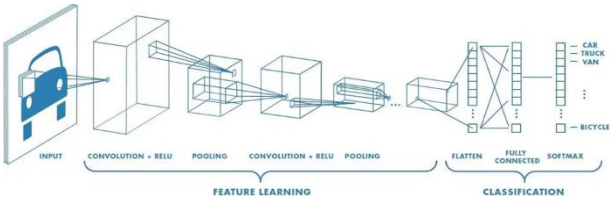

Fig.2. Example of CNN. [4]

The proposed project is a simple desktop application that utilizes the power of machine learning and artificial intelligence to offer an easy way to detect demographics from images. The application will feature a simple and user-friendly graphical user interface (GUI), enabling users to easily upload, analyze, and interpret image data from either uploaded images or live feed from camera, where a facial recognition algorithm will detect faces in each photo, passing them to our age model and gender model, then the predictions of the models will be stored in a database, allowing the user to extract meaningful information from their inputs.

II. Literature Survey and Related Works

A. Introduction

In recent years, machine learning has seen significant advances in computer vision and pattern recognition. One of the more intriguing applications is predicting age and gender from image data, In this Chapter, we start with the history of age and gender detection in machine learning, then conduct a literature review of two papers: “MiVOLO: Multi-input Transformer for Age and Gender Estimation” by Maksim Kuprashevich and Irina Tolstykh [5], and "Generalizing MLPs With Dropouts, Batch Normalization, and Skip Connections" by Taewoon Kim [6] and discuss their methodologies, strengths, weaknesses, and how they compare

B. Scientific Background

Understanding age and gender from images has many applications, including marketing and advertising, where gender and age prediction play a crucial role in targeted marketing campaigns. In Social Sciences and Demographics, researchers studying social dynamics, population trends, and migration patterns can benefit from accurate age and gender predictions. In Security and Surveillance, identifying potential threats based on age and gender can help improve security protocols like access control and surveillance.

In the early days of age and gender prediction, Classical machine learning algorithms were used, such as Linear classifiers, K nearest neighbors, Linear discriminant, Support vector machines [7], and Early neural networks such as a perceptron [8]. These methods achieved low accuracy, were complex to design, struggled with variations in lighting, facial appearance and pose, and didn't generalize well to unseen data.

Introducing Convolutional Neural Networks (CNNs) to age and gender detection marked a significant leap in the field. CNNs do not require manual feature engineering. Through their convolutional layers, CNNs learn relevant features for age and gender classification. Compared to traditional methods, CNNs were more tolerant of variations in pose, lighting, and image quality [9]. And with techniques such as Deep learning and Transfer learning, CNNs became dominant in the field of age and gender detection.

TABLE 1. Comparison between famous deep learning architectures. [9]

Feature Extractor	Age Estimation (MAE)	Gender Classification (Accuracy)
VGG_f	4.86	93.42
ResNet50_f	4.65	94.64
SENet50_f	**4.58**	**94.94**

C. Literature Review

The research introduces MiVOLO, a cutting-edge model for age and gender estimation utilizing both face and body features, achieving unparalleled accuracy across five benchmarks and showing robust performance even in the absence of visible faces [5]. MiVOLO is trained on a dataset of about 500,000 annotated images using a multi-input strategy that leverages a visual transformer model, VOLO [10], and a feature fusion module to integrate cross-view features. It uses a weighted MSE loss function [11] for age prediction and a binary cross-entropy loss for gender prediction. This methodology not only compares favorably against existing models and human-level accuracy but also facilitates rigorous model validation and evaluation through the newly introduced LAGENDA benchmark dataset.

TABLE 2. FairFace, Adience, AgeDB Test results using MiVOLO-D1 trained on LAGENDA train set. [5]

Method	Test set	Age Acc	Age MAE	Gender Acc
FairFace	FairFace	59.70	-	94.20
MiVOLO-D1 Face&Body	FairFace	61.07	-	95.73
DEX	AgeDB	-	13.1	-
MiVOLO-D1 Face	AgeDB	-	5.55	98.3
MWR	Adience	62.60	-	-
AL-ResNets-34	Adience	67.47	-	-
Compacting	Adience	-	-	89.66
Gen MLP	Adience	-	-	90.66
MiVOLO-D1Face	Adience	68.69	-	96.51

One of the strengths of the paper is its innovative approach. Where most studies still use a CNN [9], the study uses a transformer, leveraging advancements in computer vision technology. MiVOLO also achieves state-of-the-art performance on five benchmarks, IMDB-Clean [12], UTKFace [13], FairFace [14], Adience [15], and AgeDB [16], demonstrating high accuracy and robustness in challenging scenarios. Even when faced with images where faces are not visible, the model shows strong generalization capabilities, which highlights its versatility.

Limitations of the Study include a data imbalance for ages above 60, as the number of images begins to decrease rapidly, which may introduce bias towards the other age groups. The study also primarily focuses on accuracy metrics for model evaluation and includes a more in-depth analysis of other performance metrics or qualitative assessments that could provide a comprehensive evaluation.

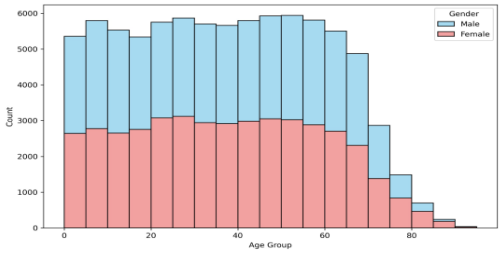

Fig.3. Age and gender distributions with bin steps of 5 in the LAGENDA dataset. [5]

The Second research paper explores different MLP [17] architectures and their performance on age and gender datasets. The study focuses on enhancing MLPs by whitening [18] inputs before linear layers and incorporating skip connections [19] to improve convergence and performance. The research introduces a structured approach to testing MLP architectures and demonstrates the benefits of these modifications through empirical experiments. The paper also discusses dropouts [20] for approximating Bayesian inference and batch normalization [21], which allows higher learning rates to be used, thus leading to faster convergence.

Methodologies include the study testing various MLP architectures on age and gender datasets to evaluate their performance, Whitening the input data before linear layers to improve convergence and generalization, and adding skip connections to the MLP architecture to facilitate better information flow and training stability, and adding dropouts which are used for regularization. The study conducts empirical experiments to compare the performance of different MLP architectures with and without the proposed enhancements.

By systematically testing and incorporating these enhancements, the study aims to demonstrate the effectiveness of these modifications in improving the convergence and performance of MLP architectures for age and gender classification tasks.

The study finds that incorporating techniques such as whitening inputs, skip connections, and dropouts in MLP architectures improves performance on age and gender classification tasks and shows better generalization to other datasets, indicating the effectiveness of these enhancements in improving model robustness. It also demonstrates a structured approach to testing different MLP architectures, highlighting the importance of systematic evaluation to understand model performance better.

TABLE 3. COMPARISON OF MOST SUCCESSFUL TRIAL IN EACH MLP ARCHITECTURE. [6]

Model	Test Loss	Test Acc
2-class gender classification, random init, no dropout	0.4047	0.8413
2-class gender classification, pretrained, no IC	0.2679	0.9066
8-class age classification, random init	1.6409	0.5482
8-class age classification, pretrained	1.3567	0.6086
modified 8-class age classification, random init	1.6319	0.5477
modified 8-class age classification, pretrained, no IC	1.3880	0.6030

The strength of the paper lies in applying techniques originally designed for CNNs to MLPs. The paper introduces a novel perspective on enhancing MLP architectures for improved convergence and performance. Through empirical experiments, the paper provides concrete evidence of the effectiveness of incorporating whitening inputs, skip connections, and dropouts in improving model performance.

A limitation of the paper is using two unbalanced datasets, IMDB-Wiki and Adience, which creates an imbalance in the gender data, where only 41.5% of the data is female, as well as an imbalance in the age data, where most of the images lies in the 20 to 60 years old, making ages under 20 and above 60 under-represented in the data. Another limitation is primarily. focusing on performance metrics such as accuracy and cross-entropy loss, potentially overlooking other essential evaluation criteria.

TABLE 3. COMPARISON OF MOST SUCCESSFUL TRIAL IN EACH MLP ARCHITECTURE. [6]

Dataset	Number of images before removal	Number of images after removal	Gender	
			Female	Male
Adience	19,370	17,055 (88.04% of original)	9,103	7,952
IMBD-WIKI	523,051	398, 251 (76.14% of original)	163,228	253,023

D. Synthesis and Comparison

Both papers delve into enhancing models for age and gender estimation, tackling the inherent challenges and complexities of accurately determining these demographics. The "MiVOLO" paper leverages both facial and body information to improve model generalization and introduces a multi-input transformer model tailored for this task. It performs extensive experiments across multiple datasets such as LAGENDA, UTKFace, IMDB-clean, Adience, FairFace, and AgeDb, demonstrating robust performance across these diverse benchmarks. Conversely, the

"Generalizing MLPs" paper emphasizes the limitations of treating age estimation purely as a classification problem and discusses enhancements to multilayer perceptron models. This includes the implementation of techniques such as dropouts, batch normalization, and skip connections, specifically testing these improved MLP architectures on the Adience and IMDB-Wiki datasets only.

In terms of results, the "MiVOLO" paper highlights that the MiVOLO model consistently outperforms human capabilities in age recognition tasks, showing superior accuracy across various age groups and details the specific architecture of the transformer model used. Meanwhile, the "Generalizing MLPs" paper illustrates that MLP architectures enhanced with dropouts, whitening, skip connections, and batch normalization can yield better performance on age and gender classification tasks. Both papers, however, identify a significant gap in their research due to the imbalanced data problem, noting an insufficient representation of images for individuals under 20 and above 60 years old. Additionally, while both papers measure accuracy and mean absolute error (MAE), they lack other important evaluation metrics such as precision, recall, and area under the curve (AUC), which could provide a more comprehensive understanding of model performance.

The introduction of the MiVOLO multi-input transformer model and enhancements to traditional multilayer perceptron in the "Generalizing MLPs" paper demonstrate progress in improving model accuracy and robustness. These studies emphasize the importance of model enhancements, such as dropouts, batch normalization, and skip connections, in advancing the convergence and generalization of models for age and gender estimation tasks. The state-of-the-art performance achieved by the MiVOLO model on various benchmarks showcases the effectiveness of multi-input transformer models in handling diverse datasets and challenging scenarios. Furthermore, the comparison of model performance with human-level accuracy suggests the potential of AI models to outperform humans in age recognition tasks.

Some areas warrant further investigation. Addressing dataset bias and class imbalance, as highlighted in the discussion of existing benchmarks, is crucial for improving model fairness and accuracy. As well as providing other metrics to evaluate the models.

III. Proposed Systems and Algorithms

This chapter presents a detailed examination of the tools and algorithms proposed for age and gender detection systems. We begin by outlining the core components of the system, Then, the chapter dives into the selected algorithms. This discussion also includes details on the datasets and tools used to build the proposed system.

A. Methodologies

.Face detection is a computer vision technology used to identify and locate human faces within digital images or video frames. Face detection algorithms search for specific patterns within an image that resemble facial features, such as eyes, nose, and mouth. Once potential face regions are located, the algorithm verifies them using more advanced techniques to confirm the presence of a true face. Models Testing in the system include:

- Haar cascades[22] use a series of simple rectangular features resembling light and dark patterns within an image. These features are compared to potential face regions in a cascade-like fashion, quickly eliminating areas that do not resemble a face and focusing on promising candidates.

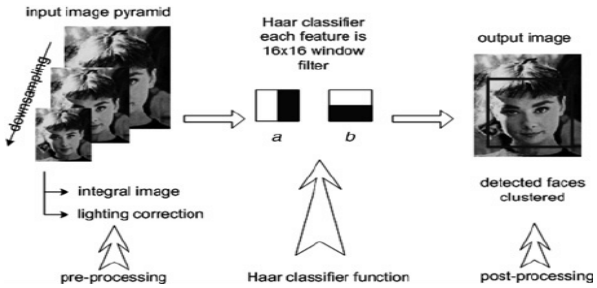

Fig.4. Haar Cascade example. [23]

- MTCNN[24] is a deep learning-based face detector that uses three cascading stages (networks) for efficiency and accuracy. The initial network proposes rough candidate regions, the second stage refines potential face boxes, and the final network provides fine-grained facial bounding boxes along with the location of key facial landmarks (eyes, nose, mouth).
- BlazeFace[25] is a lightweight and ultrafast face detector designed for mobile devices, leveraging a Google AI research model as an efficient feature extractor. It employs a novel anchor strategy and predicts both facial landmarks and bounding boxes, using single-stage and double-stage non-maximum suppression for accuracy.

Data augmentation refers to techniques that artificially expand the size and diversity of an image dataset. This helps prevent overfitting and improves the robustness of machine learning models designed for tasks in computer vision. Common augmentation techniques include flipping images horizontally (to handle mirrored faces), applying random rotations (to address varying head poses), and changing brightness or contrast levels (to simulate different lighting conditions). Additionally, techniques like adding noise or blurring can make the model more resilient to image imperfections.

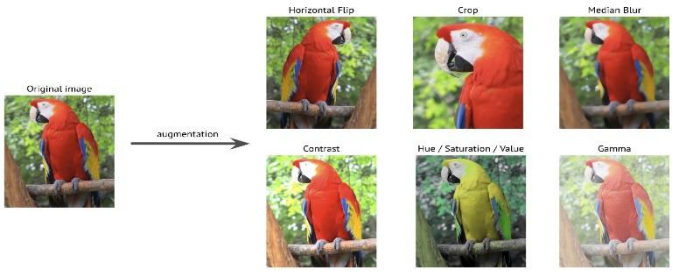

Fig.5. Augmentation Example. [26]

Feature extraction is a process in machine learning where raw data is transformed into a set of numerical representations

called features. These features capture the essential characteristics of the data, making it easier for machine learning algorithms to learn and make predictions. The goal of feature extraction is to reduce the dimensionality of the data while preserving the most important information. Algorithms tried were as followed:

- Histogram of Oriented Gradients [27] is a traditional feature extraction method used in computer vision. It calculates the distribution of edge directions or gradient orientations within localized portions of an image. HOG features are often combined with machine learning classifiers, like Support Vector Machines (SVMs), to capture distinctive patterns relevant to object detection or image classification.
- Convolutional Neural Networks provide a powerful deep learning approach to feature extraction. CNNs learn hierarchical features through multiple layers, automatically discovering intricate patterns from image data that are highly discriminative for various classification tasks.

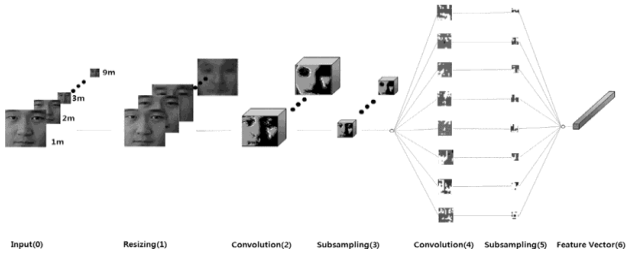

Fig.6. Feature extraction using CNN. [28]

Classification refers to the process of assigning data points to specific categories. For age and gender detection, classification algorithms analyze the extracted features from facial images and predict the most likely age group and gender of the individual in the image. Convolutional Neural Networks (CNNs) can be tailored for classification by adding final layers that predict the age group or gender based on the learned features. These deep learning models often demonstrate superior performance when trained on large datasets of labeled facial images.

B. Datasets Used

Training on a single dataset has limitations. Real-world faces come in all shapes, sizes, and ethnicity with various poses, lighting conditions, and accessories. A single dataset might not capture this variety, which can affect the model's accuracy when encountering faces from a different source.

- The IMDB-WIKI dataset is a large-scale collection of face images with associated metadata, including age and gender labels. It was primarily created by automatically scraping images from the IMDB and Wikipedia websites. The dataset contains over 500,000 images, offering significant diversity in terms of age, gender, and ethnicity. Despite containing noise and inaccuracies, the IMDB-WIKI dataset serves as a valuable benchmark for research in age and gender estimation from facial images.
- The FairFace dataset is a curated dataset of face images designed specifically to address the issue of bias in age, gender, and race classification models. It contains a balanced distribution of images across seven race categories and multiple age groups. The dataset aims to mitigate the tendency of models to be less accurate for certain demographic groups. By promoting fairness and reducing bias, FairFace facilitates the development of more inclusive and ethical facial analysis systems.

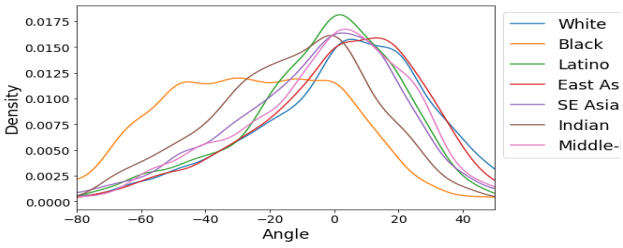

Fig.7. Distribution of Race in Fairface Dataset. [14]

- The UTKFace dataset is a large-scale facial image dataset containing over 20,000 images with annotations for age, gender, and ethnicity. The images span a wide age range (from 0 to 116 years old) and exhibit variations in pose, facial expression, and lighting conditions. UTKFace is a popular choice for training and evaluating age estimation models. Notably, the dataset includes both aligned and cropped faces as well as the original images, providing flexibility for research purposes.
- The AgeDB dataset is a focused collection of face images specifically designed for age estimation research. It contains over 16,000 images with detailed age annotations ranging from 1 to 101 years old. Images in the AgeDB dataset include variations in pose, lighting, and expression, making it a challenging but realistic benchmark. The dataset is widely used for evaluating and comparing the performance of age estimation algorithms.

To create a diverse and robust training dataset, we combined elements from AgeDB, UTKFace, FairFace, and IMDB-WIKI. Careful pre-processing was applied to ensure compatibility, including selecting one image per person from IMDB-WIKI, as well as applying the same Face detection algorithm on all four datasets to ensure similarity in size. This integrated process yielded a large-scale dataset with enhanced age, gender, and ethnic diversity, ideal for training comprehensive age and gender detection models.

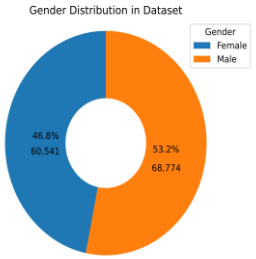

Fig.8. Gender distribution in our dataset

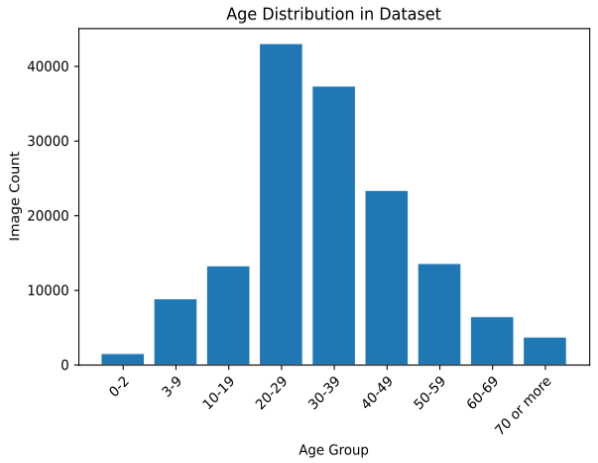

Fig.9. Age distribution in our dataset

C. Algorithms Used

- Logistic regression [] is a statistical classification model used to predict the probability of a categorical outcome (E.g.: male/female). It models the relationship between a set of predictor variables and the probability of a binary outcome using a logistic function. Logistic regression is often used when the dependent variable is dichotomous (having two possible values), however, it can be used for multi-class classification using the One-vs-Rest (OvR) approach.
- Support Vector Machines [29] are versatile machine learning algorithms used for classification tasks. SVMs aim to find the optimal hyperplane (sometimes called a decision boundary) that maximizes the separation between classes in a high-dimensional feature space. To handle non-linearly separable data, SVMs employ the "kernel trick" which implicitly maps data into higher dimensions where linear separation may be possible. SVMs are known for their robustness and ability to work well with high-dimensional data.
- Decision trees[30] are flowchart-like machine learning models that create a series of hierarchical rules to classify data. Each node in the tree represents a test on a feature (E.g. a pixel value in an image), and branches represent possible outcomes of that test. The process continues, with subsequent nodes further splitting the data, until a leaf node is reached, representing the final classification. Decision trees are known for being easy to interpret and visualize. However, they can be prone to overfitting if not carefully controlled.
- MobileNetV2[31] is a convolutional neural network architecture designed specifically for mobile and resource-constrained devices. It builds upon the original MobileNet architecture, introducing a key innovation called inverted residual blocks with linear bottlenecks. These blocks significantly reduce the number of computations and parameters required, making the model smaller and faster. MobileNetV2 achieves a great balance between accuracy and efficiency, making it ideal for image classification, object detection, and other tasks on mobile phones or embedded systems. Additionally, it serves as a strong foundation for transfer learning, where it can be adapted to new tasks with minimal fine-tuning.
- EfficientNetB4[32] is a member of the EfficientNet family of convolutional neural network architectures. It was developed through a systematic process called compound scaling, which carefully balances network depth, width, and resolution to achieve excellent accuracy and efficiency trade-offs. EfficientNetB4 outperforms many larger models while having fewer parameters and faster inference times. This makes it attractive for image classification and other vision tasks where computational resources are limited.
- ResNet50V2[33] is an evolution of the popular ResNet50 deep convolutional neural network architecture. It introduces improvements to the residual block design, including pre-activation and modifications to the convolutional layers within the block. These changes improve gradient flow during training, allowing for the construction of even deeper networks. ResNet50V2 is widely used for image classification, object detection, and as a feature extractor in various computer vision tasks.
- YOLOv8[34] is a state-of-the-art real-time object detection framework developed by Ultralytics renowned for its speed and accuracy. While primarily focused on object detection and localization using bounding boxes, YOLOv8 can be effectively adapted for image classification tasks. This involves utilizing the backbone network for feature extraction and appending a classification head. YOLOv8's backbone leverages efficient architectures and novel techniques for enhanced feature representation. With its speed advantages and potential strong classification performance where rapid inference is crucial.

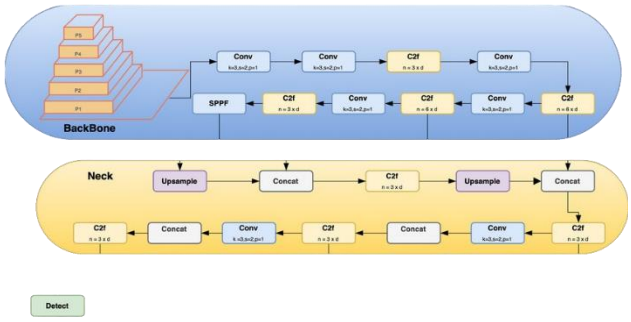

Fig.10. YOLOV8 architecture. [35]

- ViT[36] represents a recent paradigm shift in computer vision. Unlike traditional CNNs that rely solely on convolutional operations, ViTs directly process image patches using transformer architectures commonly used in Natural Language Processing (NLP). This allows ViTs to capture long-range dependencies and global context within images, potentially leading to superior performance on various tasks. ViT architectures often leverage techniques

like positional encoding to inject spatial information that is lost during patch processing. Training these models can be computationally expensive, but pre-training on large image datasets like ImageNet and fine-tuning on specific tasks like age and gender detection shows promising results.

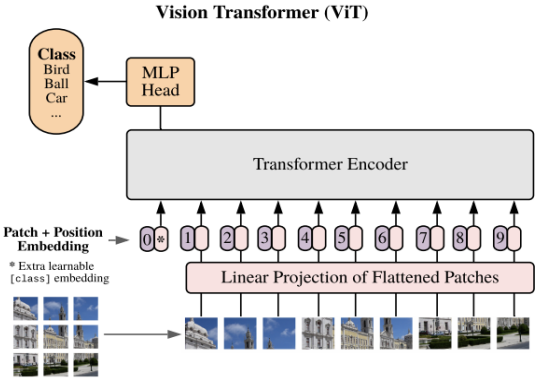

Fig.11. ViT architecture. [36]

D. Tools Used

In our project on Age and gender detection using face images, we have utilized various tools and technologies to perform data analysis, implement algorithms, and visualize results.

Here are the tools we have used:

- Python [37]-[39]: Python is a widely used programming language that provides a rich set of libraries and tools for data analysis, machine learning, and scientific computing. It serves as the primary language for developing the project.
- Scikit-learn [40],[41]: Scikit-learn is a popular machine learning library in Python that provides a wide range of algorithms and tools for tasks such as classification. It offers a unified interface and easy-to-use functions for training models and evaluating their performance.
- Tensorflow: TensorFlow is a popular open-source machine learning platform developed by Google. It provides a comprehensive suite of tools and libraries for building, training, and deploying machine learning models with a focus on deep neural networks.
- Pytorch: PyTorch is a popular open-source deep learning framework known for its flexibility and ease of use. It provides dynamic computational graphs for building neural networks, along with extensive tools and libraries for computer vision, natural language processing, and other machine learning domains.
- Tkinter: Tkinter is Python's built-in GUI (Graphical User Interface) toolkit. It provides a simple way to create desktop applications with elements like buttons, labels, text boxes, and more.
- PostgreSQL: PostgreSQL is a powerful, open-source object-relational database management system. It emphasizes extensibility, standards compliance, and reliability, making it a popular choice for demanding applications ranging from web development to data analysis.
- Kaggle: Kaggle is an online platform that offers a collaborative environment with access to powerful computing resources and a community of data scientists and machine learning practitioners.
- Visual Studio Code (VS Code): Visual Studio Code is a source code editor that provides a rich set of features and extensions for efficient coding and development. It offers a user-friendly interface, syntax highlighting, debugging capabilities, and integration with Git and other tools.

IV. Results and Conclusion

In this chapter, we present the results obtained from applying our proposed age and gender detection system on facial images. We discuss the effectiveness of each approach in achieving the project's objectives and explain why we chose YOLOv8. We analyze the accuracy, strengths, and limitations of the system's performance. Finally, we conclude the chapter by summarizing the key findings and outlining potential areas for future improvements.

A. Conditions of Trials

- We used MTCNN to detect faces in each of the datasets mentioned before and combine them into a new dataset that consists of 135,106 images for training, 15011 image for validation, and 8032 image for test in the Age dataset. And 116,383 for training, 12931 image for validation, and 5000 image for test in the Gender Dataset.
- Overall accuracy, as well as precision, recall, and F1-score were calculated for both age and gender detection. Results were broken down and analyzed per age group/gender to identify potential biases, then combined with weighted average to get a definite value for model's performance. Also, we included a confusion matrix and train and loss metrics for the chosen model.
- Training & testing were performed on a P100 GPU-powered server on Kaggle platform using Python and the PyTorch deep learning framework.

B. Results

For Gender Model:

TABLE 5. Performance of each model on gender dataset.

Model Name	Accuracy	Precision	Recall	F1 Score
Logistic Regression	81.7%	81.8%	81.7%	81.7%
SVM	64.1%	64.1%	64.1%	64.1%
Decision Tree	72%	72.1%	72%	72%
MobileNetV2	92.8%	95.4%	89.3%	92.2%
Efficientnet B4	94%	95.8%	91.7%	93.7%
ResNet 50V2	91.8%	93.5%	90%	91.7%
Vision Transformer	90.1%	89.4%	90.2%	90%
YOLOv8	**94.2%**	**93.8%**	**94.6%**	**94.2%**

- **Logistic Regression**: Logistic Regression provides a solid baseline performance. All its metrics are consistent, scoring around 81.7% for accuracy, precision, recall, and F1 score. indicating a balanced model without significant bias towards false positives or false negatives. It's a good starting point for classification problems.
- **SVM**: SVM demonstrates the lowest performance overall in this comparison. Scoring 64.1% for all metrics. Its identical scores across all metrics suggest a model that might be struggling to find effective decision boundaries to separate the data. Further tuning or a different kernel function might be needed.
- **Decision Tree:** Decision Tree offers a slight improvement over SVM. Its scores are still relatively modest with an accuracy of 72% and similar scores for precision, recall, and F1 score, hinting at potential overfitting or a need for pruning. However, decision trees are valued for their interpretability.
- **MobileNetV2:** MobileNetV2 exhibits very strong performance. It has high accuracy (92.8%) and an excellent F1 score (92.2%), demonstrating a good balance between precision and recall (95.4% and 89.3% respectively). This model is likely well-optimized for its task, making it suitable for deployment in resource-constrained environments.
- **EfficientNet B4:** EfficientNet B4 deliveries achieved the second-highest accuracy (94%) and excelled in precision (95.8%), However, like MobileNetV2, its recall (91.7%) was lower. This suggests exceptional ability to correctly classify data points while minimizing both false positives and false negatives. It likely comes with a slightly higher computational cost compared to models like MobileNetV2.
- **ResNet50V2:** ResNet50V2 good accuracy (91.8%) and precision (93.5%) but had a slightly lower recall (90%). While slightly less accurate than EfficientNet B4, it still offers an excellent balance of precision and recall. ResNets are known for their ability to handle complex image classification tasks.
- **Vision Transformer:** Vision Transformer shows competitive performance, but its accuracy (90.1%) lags slightly behind the top performer's accuracy with precision (89.4%) and recall (90.2%). Further fine-tuning might push its accuracy closer to the leading models.
- **YOLOv8:** YOLOv8 achieved the highest accuracy (94.2%) among all models. It also had strong precision (93.8%) and recall (94.6%), indicating it excelled at correctly classifying the data. Its high accuracy and F1 scores reflect its ability to both locate and correctly classify objects within images. YOLO models are known for their real-time detection capabilities.

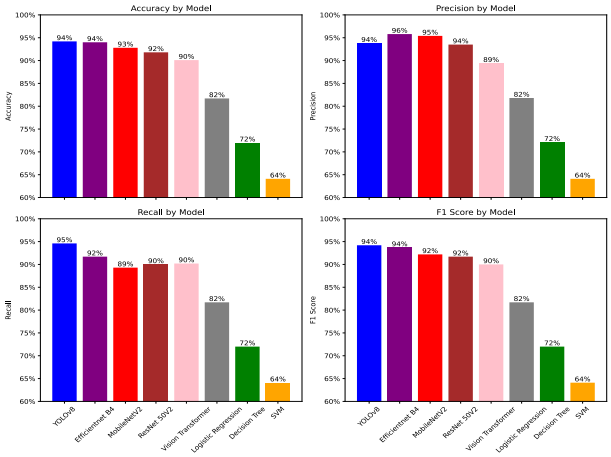

Fig.12. Performance of each model on gender dataset.

For Age Model:

TABLE 6. PERFORMANCE OF EACH MODEL ON AGE DATASET

Model Name	Accuracy	Precision	Recall	F1 Score
Logistic Regression	39.9%	39.5%	39.9%	39.7%
SVM	33.9%	35.5%	33.9%	34.7%
Decision Tree	33.8%	33.9%	33.8%	33.8%
MobileNetV2	59%	73.5%	39.2%	51.1%
Efficientnet B4	**64.4%**	**66.7%**	**61.5%**	**64%**
ResNet 50V2	49.9%	66.3%	28%	39.4%
Vision Transformer	59.2%	60.6%	61%	60.8%
YOLOv8	**62.5%**	**61.8%**	**62.2%**	**62%**

- **Logistic Regression:** This model demonstrates the lowest accuracy in this age classification task, achieving only 39.9%. Its linear nature likely limits its ability to handle the complex relationships between image features and age.
- **SVM:** The SVM performs slightly better than Logistic Regression but still struggles with accuracy, reaching 33.9%. This suggests that finding an optimal separating hyperplane for age classes within the image data is challenging.
- **Decision Tree:** The Decision Tree also exhibits relatively low accuracy at 33.8%. This implies that its rule-based structure may oversimplify the patterns needed to distinguish between different age groups.
- **MobileNetV2:** MobileNetV2 achieves a moderate accuracy level of 59%. Its focus on efficiency seems to come with some trade-off in its ability to capture the nuances of age in images, as evidenced by its lower precision (73.5%) and recall (39.2%).
- **EfficientNetB4:** EfficientNetB4 delivers the best performance among the models listed, reaching an accuracy of 64.4%. Its carefully scaled architecture appears well-suited to identifying the visual cues that represent age.
- **ResNet50V2:** ResNet50V2 delivers a middling performance in this task with an accuracy of 49.9%. Its

residual connections, while helpful for many image problems, might not provide the ideal feature representations for age classification, as seen in its lower recall rate (28%).

- **Vision Transformer:** The Vision Transformer shows a moderate level of accuracy at 59.2%. Its attention-based mechanism demonstrates potential for age classification, but further refinement might be needed.
- **YOLOv8:** YOLOv8 achieves a decent accuracy score of 62.5%. While primarily designed for object detection, it seems to adapt reasonably well to the task of age classification.

While EfficientNet B4 achieved the highest accuracy metrics in this age classification task, its large size (~750mb) and complexity likely lead to slower inference speeds. In contrast, YOLOv8, despite slightly lower accuracy, is preferred in situations where real time age classification is essential due to its faster inference potential. This scenario underscores the common trade-off in model selection, where one must balance the desire for the highest accuracy against the practical need for efficient and fast performance.

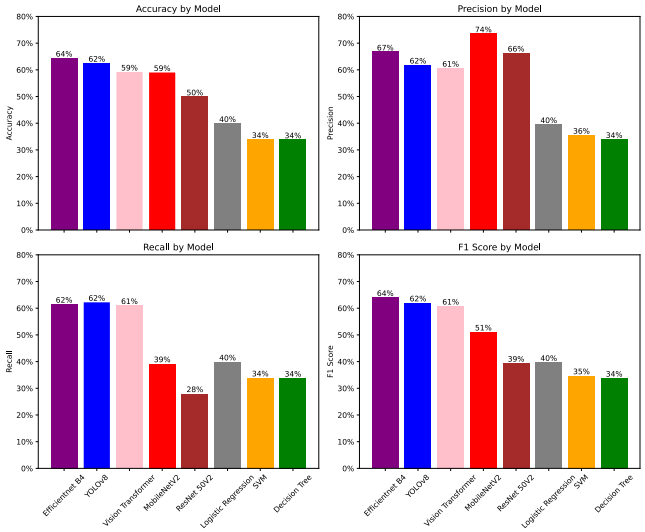

Fig.13. Performance of each model on Age dataset.

C. *Future Considerations*

Our current age and gender detection models demonstrate promising results, particularly in the domain of gender classification. To further refine the project, we propose a multi-pronged approach:

- Intelligent Dataset Fusion: We will investigate techniques such as weighted ensemble methods to strategically combine the strengths of our existing datasets. This will allow us to capitalize on the high accuracy of our gender model while improving the age detection model's performance.
- Dataset Expansion for Representativeness: We will actively seek more datasets that expand both the age range and ethnic diversity of our training data. Emphasis will be placed on acquiring datasets with controlled image quality factors (lighting, occlusions) to optimize the feature extraction process essential for reliable classification.
- Exploration of Techniques Beyond Classification: While traditional classification methods have formed the basis of our work, we will explore alternative approaches. This might include regression techniques for age estimation or investigating unsupervised learning methods to uncover patterns within the data without explicit labels or object detection instead of relying on face detection models.
- Evaluation of Advanced Techniques: We will explore the evaluation of Generative Adversarial Networks (GANs) to synthesize additional training data, potentially addressing imbalances in underrepresented demographics.
- Active Learning: We will explore implementing active learning techniques to identify the most informative and challenging data samples. These samples would then be prioritized for further labeling, leading to a more efficient use of your resources.

D. *Business Value*

- **Cost:** Automated age and gender detection systems can significantly reduce costs for businesses. These systems replace manual labor involved in customer analysis and security, boosting efficiency. Additionally, this technology enables targeted marketing and better resource usage, improving decision-making and potentially increasing revenue.
- **Environmental Impact:** While age and gender detection have a smaller environmental impact than some industries, training the systems requires powerful computers that use a lot of energy. Luckily, researchers are making these systems more efficient, and the digital approach can even reduce waste in marketing materials.
- **Manufacturability:** Age and gender detection is software-based, making it easy to use on different devices without needing specialized hardware. This software approach avoids the environmental costs of traditional manufacturing, like materials and transportation.
- **Ethics:** Age and gender detection raises ethical concerns about privacy, user consent, and algorithmic bias. To address these, developers should prioritize strong data protection, transparency in data use, and actively reducing bias in the models themselves.
- **Social and Political Impact:** Age and gender detection can be a double-edged sword. While it raises privacy concerns due to potential mass surveillance, it can also improve services through personalization. To navigate this, governments will likely create new rules to ensure responsible use of this technology.
- **Health and Safety:** Age and gender detection can boost security in restricted areas and improve healthcare by personalizing communication and interventions. However,

accuracy is critical to avoid risks associated with misidentification.

- **Sustainability:** Age and gender detection promotes sustainability by minimizing physical resources. Targeted marketing with this tech cuts waste by reaching the right audience. Additionally, these software solutions are adaptable and easy to update, reducing the environmental impact compared to traditional methods. without the same level of resource use as physical products, supporting a more sustainable lifecycle.

E. Results of YOLOV8

Our YOLOV8 model on the Gender dataset achieved a Val Loss of 0.37, with 94.2% accuracy, exceeding some SOTA models. Predicting 4711 image correctly out of 5000 in our test set.

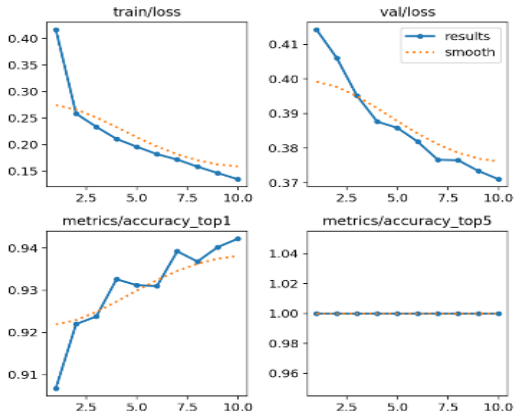

Fig.14. Training and Validation of YoloV8 Model on Gender Data.

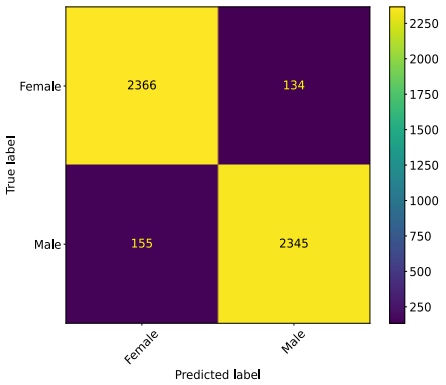

Fig.15. Confusion Matrix of YoloV8 Model on Gender Test Data.

For Age Detection, YOLOV8 achieved a Val Loss of 1.49, with 62.5% accuracy. Predicting 5005 images correctly out of 8000 in our test set.

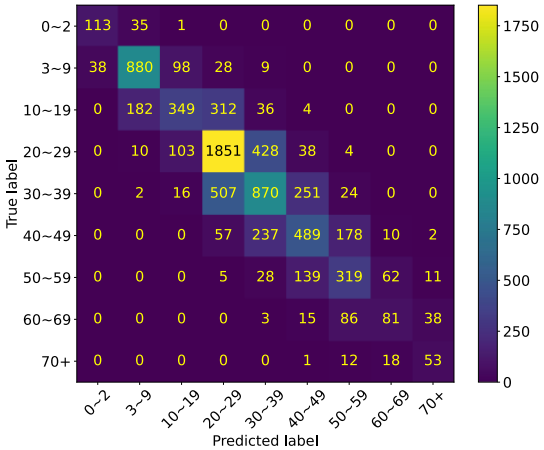

Fig.16. Training and Validation of YoloV8 Model on Age Data.

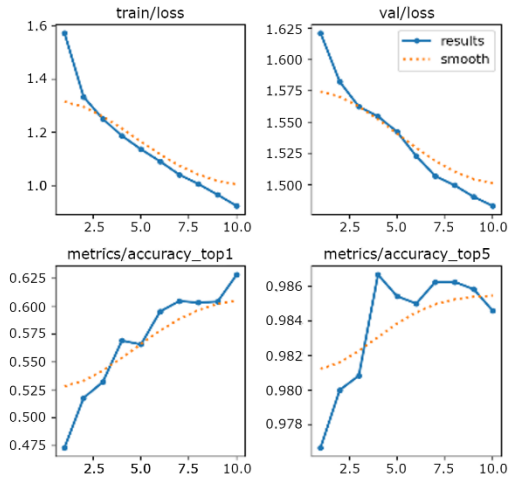

Fig.17. Confusion Matrix of YoloV8 Model on Age Test Data.

V. CONCLUSION

This thesis investigates the application of deep learning techniques for automatic age and gender detection from facial images. We explored the potential of different machine learning algorithms to extract and analyze subtle features within a human face, revealing valuable insights into a person's age and gender. This research hopes to contribute to the advancement of this rapidly growing field.

The significance of accurate age and gender detection extends across various domains, including demographics analysis, targeted advertising, and security applications. However, achieving accurate classification presents challenges due to inherent variations in facial appearance caused by ageing, diverse ethnicities, expressions, and external factors such as image quality, lighting, and camera angle.

To address these challenges, we conducted a comprehensive analysis of various machine learning models. Our methodology included careful data preprocessing, where images were collected from various datasets, mainly IMDB-Wiki, FairFace, AgeDB, and UTKFace, then face detection algorithms were applied to isolate facial regions, ensuring the models focused solely on relevant features.

Logistic Regression, Support Vector Machines, and Decision Trees were then compared against a range of CNN architectures, including MobileNetV2, EfficientNetB4, ResNet50V2, Vision Transformer, and YOLOv8. Our analysis revealed that EfficientNet B4 achieved the highest accuracy for age classification tasks, While YOLOV8 achieved the highest accuracy for Gender classification. This success can be attributed to its well-designed architecture, which effectively captures the intricate details that differentiate genders within facial images.

Beyond raw accuracy, real-world implementation necessitates consideration of computational efficiency. While EfficientNetB4 demonstrated superior performance in age detection, its larger size and slower inference speed may limit its applicability in real time scenarios. This highlights a critical trade-off between achieving the highest accuracy and ensuring efficient processing capabilities. For applications demanding real-time performance, models like YOLOv8, despite exhibiting slightly lower accuracy, could be more practical due to their faster processing speeds.

ACKNOWLEDGMENT

We would like to extend our deepest gratitude to the Dean of our faculty, Prof. Rania Elgohary, for her dedication to our academic growth and success. To our instructor and TA, Dr. Esmat Mohamed and T.A. Fares Emad, your guidance, expertise, and commitment have been invaluable in shaping our learning journey. Your passion for the subject matter has inspired us to dive deeper and strive for excellence. Their leadership has fostered a supportive learning environment that has empowered us as students.

REFERENCES

[1] "The ESRB wants to start using facial scanning technology to check people's ages," Yahoo! News, https://sg.news.yahoo.com/esrb-wants-start-using-facial-215241547.html?guccounter=1.

[2] Georgopoulos, Markos & Oldfield, James & Nicolaou, Mihalis & Panagakis, Yannis & Pantic, Maja. (2021). Mitigating Demographic Bias in Facial Datasets with Style-Based Multi-attribute Transfer. International Journal of Computer Vision. 129. 10.1007/s11263-021-01448-w.

[3] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248–255).

[4] S. Saha, "A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way | Saturn Cloud Blog," saturncloud.io, Dec. 15, 2018. https://saturncloud.io/blog/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way/

[5] M. Kuprashevich and I. Tolstykh, "MiVOLO: Multi-input Transformer for Age and Gender Estimation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2023, pp. 1-13.

[6] T. Kim, "Generalizing MLPs With Dropouts, Batch Normalization, and Skip Connections", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021

[7] B. Moghaddam and Ming-Hsuan Yang, "Gender classification with support vector machines," Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), Grenoble, France, 2000, pp. 306-311, doi: 10.1109/AFGR.2000.840651.

[8] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. SEXNET: A neural network identifies sex from human faces. In Advances in Neural Information Processing Systems, pages 572–577, 1991

[9] Sheoran, Vikas & Joshi, Shreyansh & Bhayani, Tanisha. (2021). Age and Gender Prediction using Deep CNNs and Transfer Learning.

[10] Li Yuan, Qibin Hou, Zihang Jiang, Jiashi Feng, and Shuicheng Yan. Volo: Vision outlooker for visual recognition, 2021.

[11] Yuzhe Yang, Kaiwen Zha, Ying-Cong Chen, Hao Wang, and Dina Katabi. Delving into deep imbalanced regression, 2021.

[12] Yiming Lin, Jie Shen, Yujiang Wang, and Maja Pantic. Fpage: Leveraging face parsing attention for facial age estimation in the wild. arXiv, 2021.

[13] Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017.

[14] Kimmo Karkkainen and Jungseock Joo. Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 1548–1558, 2021

[15] Eran Eidinger, Roee Enbar, and Tal Hassner. Age and gender estimation of unfiltered faces. IEEE Transactions on Information Forensics and Security, 9(12):2170–2179, 2014

[16] Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. Agedb: the first manually collected, in-the-wild age database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop, volume 2, page 5, 2017.

[17] F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review, 65 6:386–408, 1958.

[18] A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. Neural Networks, 13(4):411–430, 2000.

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 770–778, 2016.

[20] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(56):1929–1958, 2014.

[21] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML'15, page 448–456.JMLR.org, 2015

[22] Wilson, Phillip & Fernandez, Dr. (2006). Facial feature detection using Haar classifiers. Journal of Computing Sciences in Colleges. 21.

[23] Field programmable gate array-based Haar classifier for accelerating face detection algorithm - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Face-detection-flow-based-on-the-Haar-classifier_fig1_224141453

[24] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, 'Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks', 2016.

[25] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, and M. Grundmann, 'BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs'. 2019.

[26] "Albumentations documentation - what is image augmentation," What is image augmentation-AlbumentationsDocumentation, https://albumentations.ai/docs/introduction/image_augmentation/

[27] Zhou, Wei & Shengyu, Gao & Zhang, Ling & Lou, Xin. (2020). Histogram of Oriented Gradients Feature Extraction from Raw Bayer Pattern Images. IEEE Transactions on Circuits and Systems II: Express Briefs. PP. 1-1. 10.1109/TCSII.2020.2980557.

[28] A face recognition system based on convolution neural network using multiple distance face - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Example-of-extracting-feature-vector-of-the-face-image-by-CNN-for-training_fig5_296693092

[29] Xie, G., Attar, H., Alrosan, A., Abdelaliem, S. M. F., Alabdullah, A. A. S., & Deif, M. (2024). Enhanced diagnosing patients suspected of sarcoidosis using a hybrid support vector regression model with bald eagle and chimp optimizers. PeerJ Computer Science, 10, e2455.

[30] L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. Wadsworth Int. Group, 1984

[31] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, 'MobileNetV2: Inverted Residuals and Linear Bottlenecks', 2018.

[32] M. Tan and Q. V. Le, 'EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks', 2019.

[33] K. He, X. Zhang, S. Ren, and J. Sun, 'Identity Mappings in Deep Residual Networks'. 2016.

[34] Deif, M. A., Attar, H., Alrosan, A., Solyman, A. A., & Abdelaliem, S. M. F. (2024). Design and development of an intelligent neck and head support system based on eye blink recognition for cervical dystonia. Discover Applied Sciences, 6(11), 602.

[35] Parking Time Violation Tracking Using YOLOv8 and Tracking Algorithms - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/YOLOv8-architecture_fig5_371888779

[36] A. Dosovitskiy et al., 'An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale'. 20.

[37] Deif, M. A., Attar, H., Hafez, M. A., Alrosan, A., & Sharfo, S. M. (2024). Hybrids of Support Vector Regression with Bald Eagle Search Optimizer for Diagnosing Patients Suspected of Sarcoidosis. International Journal of Intelligent Engineering & Systems, 17(6).

[38] Deif, M. A., Attar, H., Hafez, M. A., Alomoush, W., & Al-Faiz, H. (2025). Automatic Sarcoidosis Stage Classification Based on Gray Level Co-occurrence Matrix Features. Appl. Math, 19(1), 197-208.

[39] Ahmed, F. R., Alsenany, S. A., Abdelaliem, S. M. F., & Deif, M. A. (2023). Development of a hybrid LSTM with chimp optimization algorithm for the pressure ventilator prediction. Scientific Reports, 13(1), 20927.

[40] Alomoush, W., Khashan, O. A., Alrosan, A., Damseh, R., Alshinwan, M., Abd-Alrazaq, A. A., & Deif, M. A. (2024). Improved security of medical images using DWT–SVD watermarking mechanisms based on firefly Photinus search algorithm. Discover Applied Sciences, 6(7), 366.

[41] Youssef, M., Sharfo, S. M., Attar, H., Deif, M. A., Hafez, M., & Solyman, A. (2023, December). Sand Cat Swarm Optimizer with CatBoost for Sarcoidosis Diagnosis. In 2023 2nd International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI) (pp. 1-7). IEEE.