## Introduction of the models (VGG16, RESNET50,SENET50) for face recognition based on research paper

In the field of computer vision and facial recognition, the selection of model architectures significantly influences the accuracy and efficiency of recognition systems. This chapter provides an overview of three prominent convolutional neural network (CNN) architectures: VGG16, ResNet50, and SENet50, which have demonstrated exceptional performance in face recognition, as discussed in various research papers.

VGG16:

VGG16, developed by the Visual Geometry Group at the University of Oxford, represents a milestone in CNN architectures. Its name, VGG16, denotes its architecture depth, comprising 16 layers. Noteworthy for its simplicity and uniformity, VGG16 consists of cascaded convolutional layers followed by max-pooling layers, with each convolutional layer utilizing a small 3x3 receptive field and a stride of 1 pixel.

Despite its straightforward architecture, VGG16 excels in capturing intricate facial features through hierarchical feature extraction, making it a popular choice for face recognition systems.

ResNet50:

ResNet50, a creation of researchers at Microsoft Research, introduces a breakthrough in training deep neural networks. Utilizing the concept of residual learning, ResNet50 addresses the challenge of training very deep networks by employing shortcut connections, also known as skip connections. These connections enable the network to learn residual functions, effectively mitigating the vanishing gradient problem encountered in deeper networks.

ResNet50's ability to learn highly discriminative features makes it particularly effective for face recognition tasks, demonstrating superior performance in feature learning and representation.

SENet50:

SENet50, proposed by researchers at Huazhong University of Science and Technology, integrates channel-wise feature recalibration mechanisms into CNN architectures. By incorporating squeeze-and-excitation blocks, SENet50 enables dynamic feature recalibration adaptively.

The core principle of SENet50 lies in channel-wise attention, allowing the network to emphasize informative features while suppressing irrelevant ones. This feature recalibration enhances the network's discriminative power, especially beneficial in face recognition tasks with varying illumination, pose, and expression.

## Justification of the models based on past research

VGG16:

VGG16, characterized by its simplicity and uniformity, has been extensively studied and validated in numerous research papers. Past research has highlighted VGG16's efficacy in capturing intricate facial features through hierarchical feature extraction. Its deep architecture, comprising 16 layers of convolutional and pooling operations, enables it to learn discriminative features effectively. Studies have shown that VGG16 performs admirably in face recognition tasks, particularly when trained on large-scale datasets. Its straightforward architecture and ease of implementation make it a popular choice for researchers and practitioners alike.

ResNet50:

ResNet50 has garnered significant attention in the research community due to its groundbreaking concept of residual learning. Previous studies have demonstrated that ResNet50's introduction of shortcut connections, or skip connections, alleviates the vanishing gradient problem encountered in training very deep neural networks. This characteristic allows ResNet50 to effectively capture and learn complex facial features, even in the presence of considerable variations in pose, illumination, and expression. Research findings have consistently showcased ResNet50's superior performance in face recognition tasks compared to earlier architectures, making it a preferred choice for challenging real-world scenarios.

SENet50:

SENet50 stands out among face recognition models due to its incorporation of channel-wise feature recalibration mechanisms. Past research has elucidated how SENet50's squeeze-and-excitation blocks dynamically recalibrate feature maps, enhancing the network's discriminative power. Studies have demonstrated that SENet50 effectively learns to emphasize informative features while suppressing irrelevant ones, leading to improved performance in face recognition tasks. Furthermore, research findings have highlighted SENet50's robustness to variations in lighting conditions, facial expressions, and occlusions, making it a reliable choice for real-world applications.

## Dataset and Validation Methodology for Model Evaluation

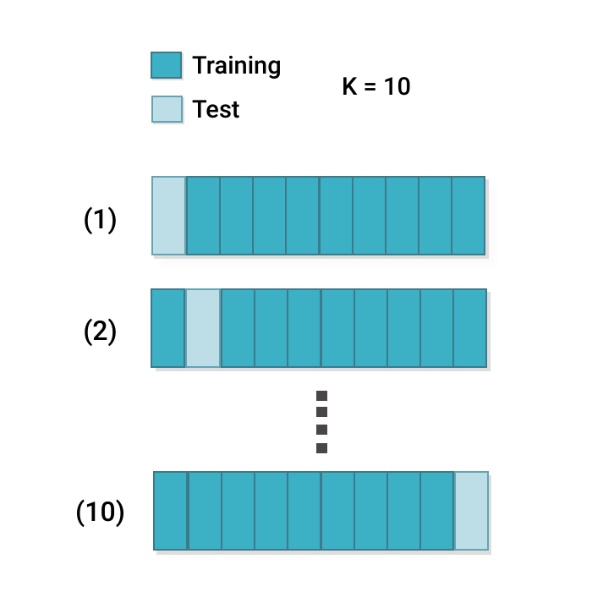
The chosen dataset is the well-known Labeled Faces in the Wild (LFW), while the validation method employed is Stratified k-Fold cross-validation with 10 folds.

Labeled Faces in the Wild (LFW) Dataset:

The Labeled Faces in the Wild (LFW) dataset serves as a pivotal benchmark in the realm of face recognition, comprising a vast collection of labeled face images sourced from the internet. With over 13,000 images, LFW encapsulates a wide array of variations in facial pose, lighting conditions, expressions, and occlusions. Its extensive diversity makes it a prime choice for evaluating the robustness and generalization capabilities of face recognition models in real-world scenarios. Researchers frequently turn to LFW due to its substantial size, diversity, and accessibility, ensuring impartial and comprehensive assessments of model performance.

Stratified k-Fold Cross-Validation (10 Fold):

To ensure rigorous evaluation of the face recognition models, we employ Stratified k-Fold cross-validation with 10 folds. A common choice of 10 fold, which has been shown to provide a good balance between bias and variance in many cases. Cross-validation stands as a standard practice for estimating a model's performance while mitigating the risk of overfitting to the training data. In Stratified k-Fold cross-validation, the dataset is partitioned into k subsets (or folds), with each fold maintaining the same class distribution as the original dataset.



In our approach, we opt for 10 folds, dividing the dataset into 10 equally sized subsets, each containing samples from every class within the dataset. During each iteration of the cross-validation process, one fold is designated as the validation set, while the remaining folds serve as the training set. This process repeats 10 times, with each fold serving as the validation set exactly once. Subsequently, performance metrics obtained from each fold are averaged to yield a robust estimate of the model's performance.

Employing Stratified k-Fold cross-validation ensures that the face recognition models undergo thorough evaluation across diverse subsets of the LFW dataset. This validation methodology enables accurate assessment of the models' generalization capabilities while minimizing biases introduced by dataset partitioning.

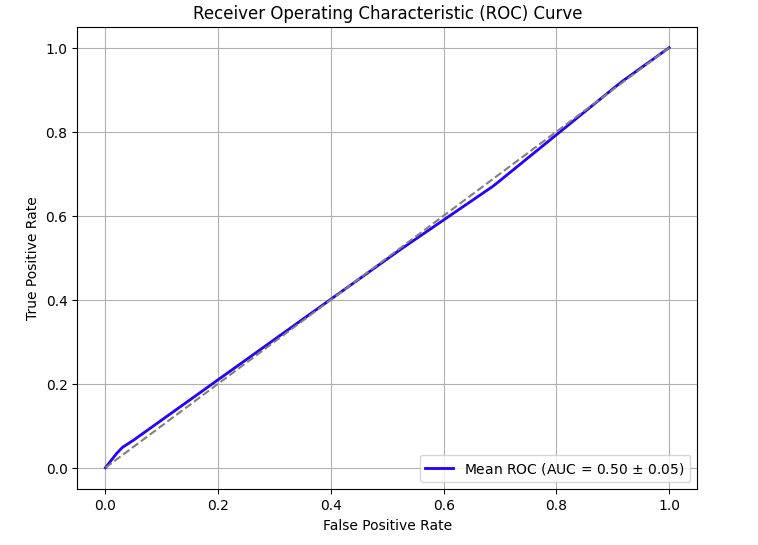
## Validation Results for VGG16, ResNet50, SeNet50 Models

In this chapter, we present the validation results for the VGG16, ResNet50, and SENet50 models, utilizing a comprehensive set of performance metrics including confusion matrix, ROC curve, mean accuracy, precision, recall, and F1 score. These metrics provide insights into the models' ability to accurately recognize faces and classify them across different categories.

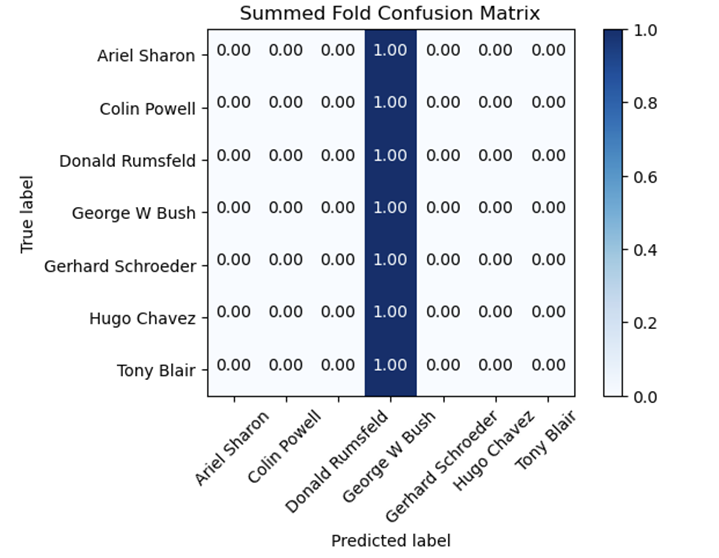
You can refer to the Jupyter Notebook .ipynb files to see more details for the result.

**VGG16:**

**ROC Curve**

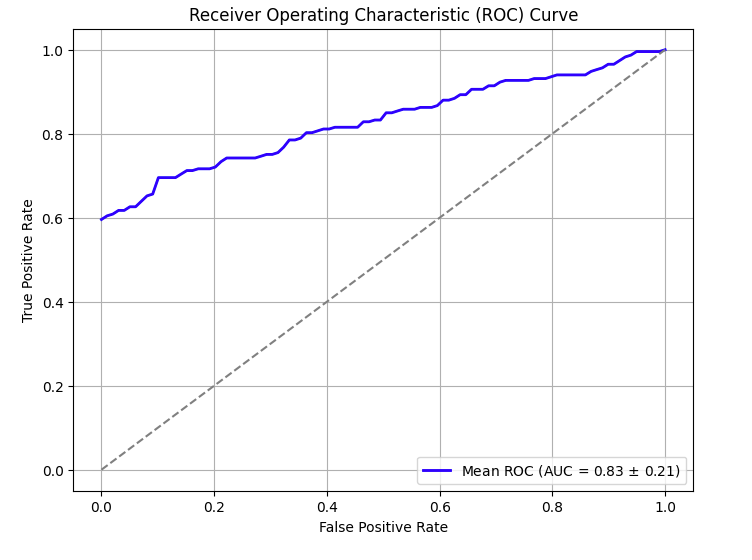
****

**Confusion Matrix**

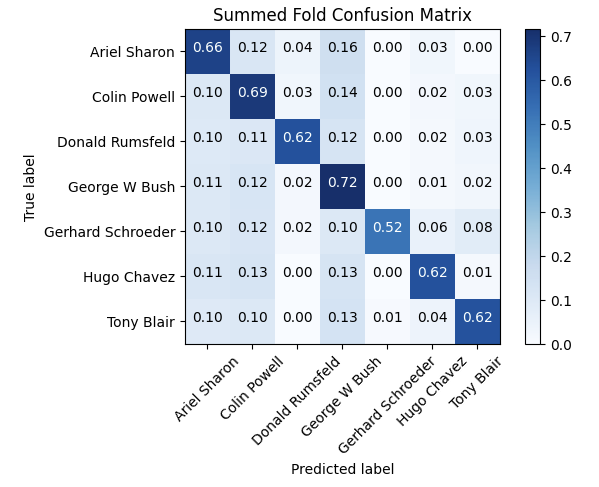
****

**Resnet50**

**ROC Curve**

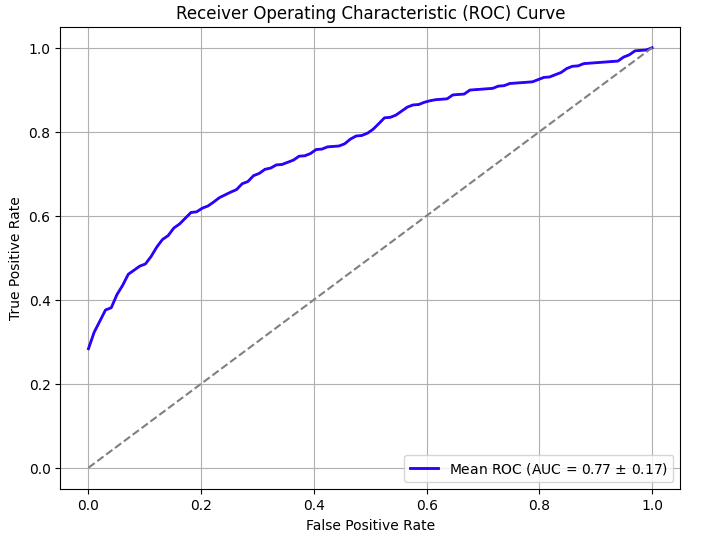
****

**Confusion Matrix**

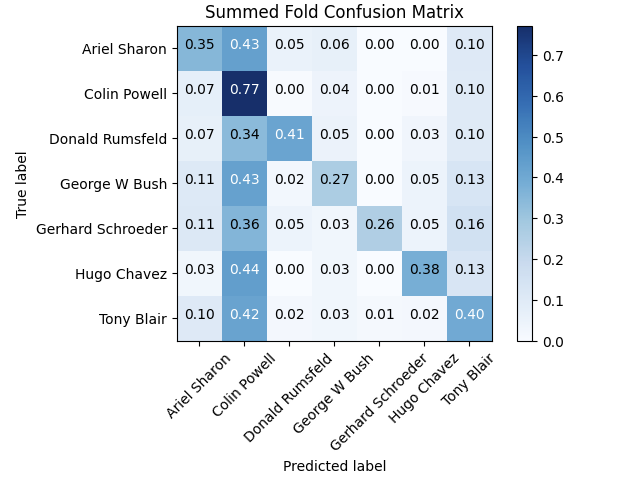
****

**Senet50**

**ROC Curve**

****

**Confusion Matrix**

****

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Models** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| VGG16 | 0.4115 | 0.1693 | 0.4115 | 0.2399 |
| Resnet50 | 0.6667 | 0.6383 | 0.6667 | 0.6296 |
| Senet50 | 0.3975 | 0.4033 | 0.3975 | 0.3373 |

Based on the provided metrics for accuracy, precision, recall, and F1 score, ResNet50 appears to be the best model for face recognition among the three options (VGG-16, ResNet50, and SENet50)

ResNet50 outperforms the other models across all metrics. It has the highest accuracy, precision, recall, and F1 score, indicating better overall performance in face recognition tasks. Therefore, ResNet50 would be the preferred choice among the three models for face recognition purposes.

## Reference [parkhi15.pdf (ox.ac.uk)](https://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/parkhi15.pdf)

[1409.1556.pdf (arxiv.org)](https://arxiv.org/pdf/1409.1556.pdf)

[cao18.pdf (ox.ac.uk)](https://www.robots.ox.ac.uk/~vgg/publications/2018/Cao18/cao18.pdf)

[Future Internet | Free Full-Text | Face Recognition Using Popular Deep Net Architectures: A Brief Comparative Study (mdpi.com)](https://www.mdpi.com/1999-5903/13/7/164)