

*Gender Classification Using Hybrid Deep Learning & Machine Learning Approach*

**A project work submitted to**

**Symbiosis Institute of Geoinformatics**

**Constituent of Symbiosis International (Deemed University), Pune**

**Submitted by: -** Nikhil Anil Patil

**PRN: -** 24070243038

**Course: -** Data Science and Spatial Analytics

**Batch: -** 2024-25

**Submitted to: -** Mr. Devawrat Bhave

**Index**

1. **Problem Statement**
2. **Objectives**
3. **Literature Review**
4. **Introduction to Dataset**
5. **Methode and Approach**
6. **Models**

* Convolutional Neural Network (CNN) using MobileNetV2
* Support Vector Machine (SVM)
* Random Forest Classifier
* XGBClassifier
* Ensemble model

1. **Analysis and Results**

* CNN
* Machine Learning

1. **Conclusion**
2. **Future work**
3. **References**

**Problem Statement: -**

In recent years, automatic facial analysis has become an essential component of various real-world applications such as surveillance systems, demographic studies, and human-computer interaction. One of the key tasks in facial analysis is gender classification, which aims to automatically identify the gender of an individual from their facial image. However, accurately predicting gender from facial images is a challenging task due to several factors such as pose variations, occlusions, lighting conditions, age-related facial changes, and low-resolution images. These challenges become even more significant when working with real-world datasets like UTKFace, which contains images captured in uncontrolled environments.

The core problem addressed in this project is to develop an efficient and reliable computer vision-based gender classification system that can accurately predict the gender of a person based on their facial image. The system should be capable of handling real-world challenges such as image noise, diversity in race and age, and variations in facial expressions. The solution aims to leverage deep learning techniques, specifically Convolutional Neural Networks (CNN), and implement them using OpenCV and other relevant libraries to ensure real-time and high-accuracy predictions.

**Objectives: -**

1. Clean, resize, and normalize the facial images to prepare them for model training and testing.
2. Utilize a pre-trained deep learning model based on Convolutional Neural Networks (CNN) to classify gender as either Male or Female.
3. Measure the accuracy and efficiency of the model on test data and analyse its capability to handle various real-world challenges such as pose variations and lighting conditions.
4. Assess the strengths and limitations of the implemented model and provide suggestions for enhancing its accuracy and performance in future work.

**Literature Review: -**

Several research studies have been conducted in the field of gender classification using facial images, focusing on various approaches ranging from traditional machine learning techniques to advanced deep learning models. Early research primarily utilized classical machine learning algorithms combined with handcrafted facial features for gender prediction. However, these approaches often struggled with complex real-world scenarios involving variations in facial pose, illumination, and occlusions.

With the advancement of deep learning, particularly Convolutional Neural Networks (CNNs), significant improvements have been achieved in facial analysis tasks. Recent studies have demonstrated that CNN-based models can automatically learn hierarchical and discriminative features from facial images, resulting in higher accuracy and better generalization compared to conventional methods.

Additionally, researchers have explored the impact of dataset diversity and demographic fairness on model performance. It has been observed that imbalanced datasets may introduce bias in gender classification models, affecting their accuracy across different demographic groups. Therefore, several works emphasize the need for balanced and diverse datasets to ensure fair and reliable predictions.

Moreover, many studies have addressed challenges related to real-world image conditions, such as variations in facial expressions, occlusions, lighting, and background clutter. These factors often affect the accuracy of gender classification models, leading to the development of robust preprocessing techniques and data augmentation methods to improve model performance.

The insights gained from previous research have guided the design and implementation of this project. This project incorporates CNN-based deep learning techniques and leverages a large, diverse dataset to build an efficient and reliable gender classification model capable of handling real-world complexities.

**Introduction to dataset: -**

The UTKFace dataset is a large-scale facial image dataset widely used in computer vision research, particularly for tasks like age estimation, gender classification, and ethnicity recognition. It consists of over 23,000 facial images collected from real-world environments under unconstrained conditions, meaning the images display a wide range of facial poses, lighting variations, occlusions, and expressions. Each image in the dataset is labelled with three key demographic attributes—age, gender, and race—which are embedded in the filename of the image itself. The labelling format follows the structure: [Age]\_[Gender]\_[Race]\_[Date Time].jpg, where gender is labelled as 0 for female and 1 for male. The age labels range from 0 to 116 years, and race is categorized into five groups: White, Black, Asian, Indian, and Others. In this project, only the gender labels are utilized for the classification task.

The UTKFace dataset is known for its diversity, including individuals from various age groups, ethnic backgrounds, and both genders. This diversity makes it an ideal choice for developing a gender classification model that can perform well across different demographic variations. However, the dataset also introduces several challenges such as pose variations, blurred or occluded faces, and possible label noise, which adds to the complexity and realism of the classification task. Despite these challenges, the UTKFace dataset is considered a benchmark dataset in facial analysis research and provides a solid foundation for training deep learning models for gender prediction. Its real-world variability ensures that the model trained on this dataset will generalize well when deployed in practical applications like surveillance, demographic analysis, and human-computer interaction systems.

Gender classification from facial images plays a crucial role in various real-world applications such as security surveillance, demographic analysis, targeted advertising, and human-computer interaction. Automatically identifying a person’s gender helps systems personalize services, improve user experience, and assist in crowd analysis without manual intervention. It also supports law enforcement agencies in suspect identification and missing person searches. In the field of computer vision, gender classification is considered a fundamental task that enhances the accuracy of higher-level facial analysis systems like age estimation and emotion recognition. Building an efficient gender classification model contributes to the development of smart, automated systems capable of making quick and accurate gender predictions in real-time environments.

**Method and Approach: -**

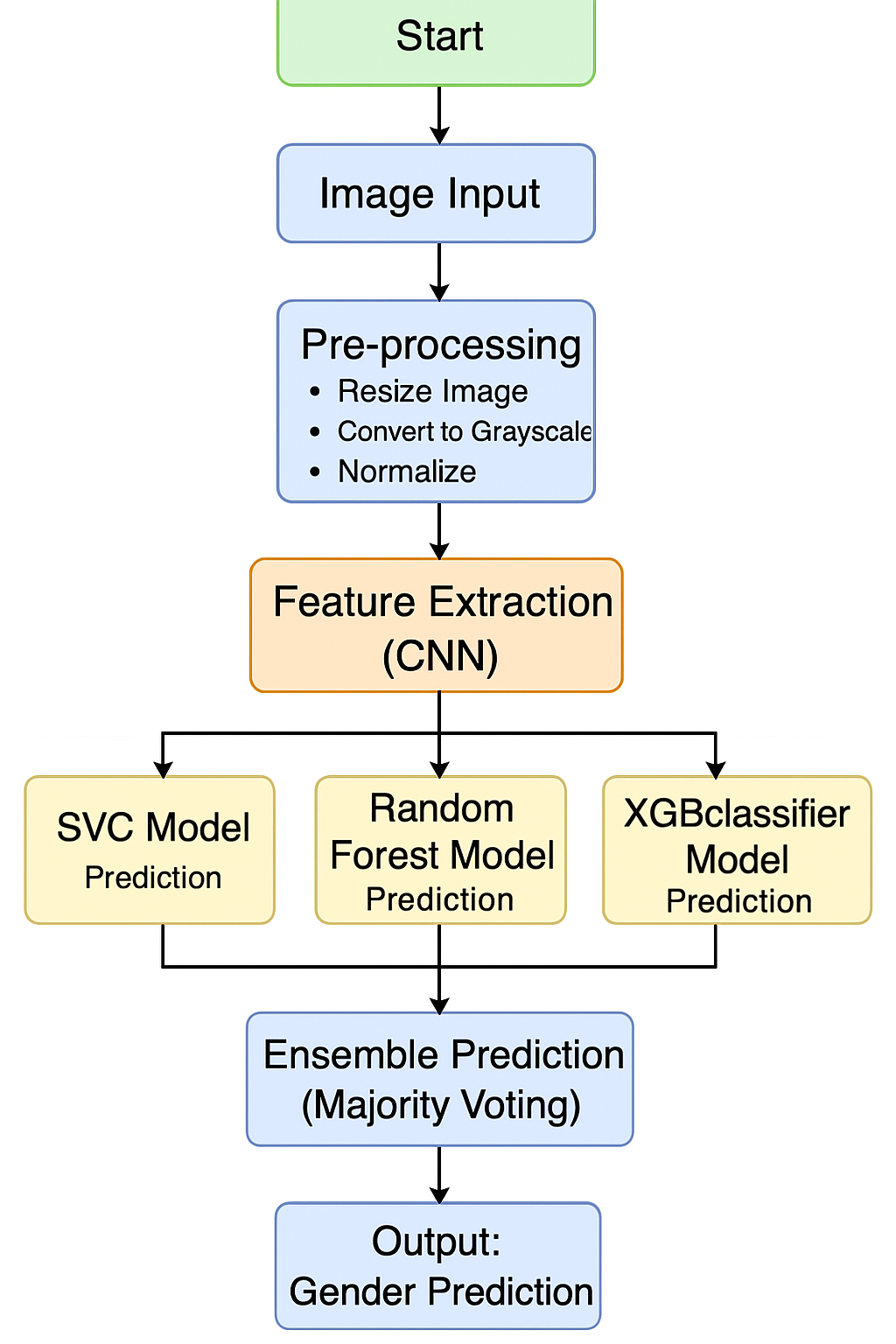
In this project, a comprehensive approach was implemented to classify gender from facial images using the UTKFace dataset, combining both deep learning and machine learning techniques. The process began with data collection and preprocessing, where the facial images were loaded in grayscale format to reduce computational complexity. Each image was resized to a uniform dimension of 128×128 pixels and normalized to scale the pixel values between 0 and 1. The corresponding gender labels were extracted from the filenames of the images. The dataset was then split into training and testing sets to evaluate the model’s performance effectively.

For the deep learning component, a Convolutional Neural Network (CNN) model was developed using the MobileNetV2 architecture through transfer learning. Since MobileNetV2 requires three-channel input images and the dataset images were in grayscale, an additional convolutional layer was introduced to convert grayscale images into three channels. The pre-trained MobileNetV2 model, initially trained on the ImageNet dataset, was employed as a feature extractor, followed by additional layers including a global average pooling layer, a dense layer with ReLU activation, a dropout layer to reduce overfitting, and a final output layer with sigmoid activation for binary classification (male or female). The CNN model was fine-tuned by unfreezing certain layers of MobileNetV2 to adapt it to the UTKFace dataset.

However, rather than using the CNN model solely for classification, the project adopted a hybrid modelling approach. After training the CNN, the deep features were extracted from the intermediate Global Average Pooling layer. These high-level feature representations captured critical facial patterns and were then used as input for classical machine learning classifiers, including Support Vector Machine (SVM), Random Forest Classifier, and XGBoost Classifier. This two-stage strategy leveraged the CNN's ability to learn complex feature representations and the effectiveness of classical classifiers in performing the final gender classification.

Additionally, to provide a comparative analysis, the project also implemented a classical machine learning approach based on handcrafted feature extraction. Local Binary Pattern (LBP) features were extracted from the facial images, and the same set of classifiers—SVM, Random Forest, and XGBoost—were applied to these features. This enabled a thorough evaluation of how handcrafted features compare to deep features in gender classification tasks.

The performance of all models—CNN-based, hybrid (CNN features + ML classifiers), and classical ML models (LBP features)—was evaluated using standard metrics such as accuracy, precision, recall, F1-score, and confusion matrix. This comparative analysis highlighted the superior performance and generalization capability of the hybrid deep feature-based models over classical approaches, emphasizing the effectiveness of combining deep learning feature extraction with machine learning classifiers in gender prediction tasks.



**Models: -**

* **Convolutional Neural Network (CNN) using MobileNetV2: -**

It is a type of deep learning model commonly used in image classification, object detection, and other computer vision tasks. CNNs are designed to automatically and adaptively learn spatial hierarchies of features from input images through the use of multiple layers such as convolutional layers, pooling layers, and fully connected layers. The main strength of CNNs lies in their ability to extract local patterns (such as edges, textures, and shapes) and combine them into more complex patterns for accurate predictions.

MobileNetV2 is an advanced CNN architecture introduced by Google that is specifically designed to be lightweight and efficient, making it ideal for mobile and real-time applications. It is an improvement over the original MobileNet architecture and is based on two main concepts namely Depthwise Separable Convolutions and Inverted Residuals with Linear Bottlenecks.

* **Support Vector Machine (SVM): -**

It is a popular supervised machine learning algorithm used primarily for classification tasks. The key idea behind SVM is to find the optimal hyperplane that best separates the data points of different classes in a high-dimensional space. It aims to maximize the margin, which is the distance between the hyperplane and the nearest data points from each class, known as support vectors. This maximization improves the generalization ability of the model.

SVM can efficiently perform linear and non-linear classification by using kernel functions (such as linear, polynomial, or radial basis function (RBF) kernels) to transform the data into a higher dimension where it becomes linearly separable. SVM is effective in high-dimensional spaces and is robust against overfitting, especially in cases where the number of features exceeds the number of samples.

* **Random Forest Classifier: -**

It is an ensemble learning technique based on decision trees. It operates by constructing a large number of individual decision trees during training and outputs the class that is the mode of the classes (majority vote) predicted by the individual trees. The randomness in the algorithm is introduced by selecting random subsets of the dataset and random subsets of features at each split in the trees.

Random Forest is highly effective for both classification and regression tasks due to its ability to handle large datasets with higher dimensionality, reduce overfitting compared to individual decision trees, and provide good accuracy even when a large proportion of the data is missing. It is also less sensitive to noise and outliers.

* **XGBoost Classifier: -**

It is a highly optimized, scalable, and efficient implementation of the gradient boosting framework. It is an ensemble technique that builds a sequence of weak learners (typically decision trees), where each new tree attempts to correct the errors made by the previous trees. XGBoost uses a technique called gradient boosting, where the model minimizes a loss function by adding new trees that predict the residual errors of previous trees.

One of the key advantages of XGBoost is its ability to handle missing data, regularization to prevent overfitting, and parallel computation for faster training. It is widely used in machine learning competitions and real-world applications due to its high predictive power and efficiency.

* **Ensemble Modelling: -**

It is a powerful machine learning technique that combines predictions from multiple individual models to produce a more accurate and robust final prediction. The core idea behind ensemble methods is that by aggregating the strengths of multiple models, the overall performance improves compared to relying on a single model.

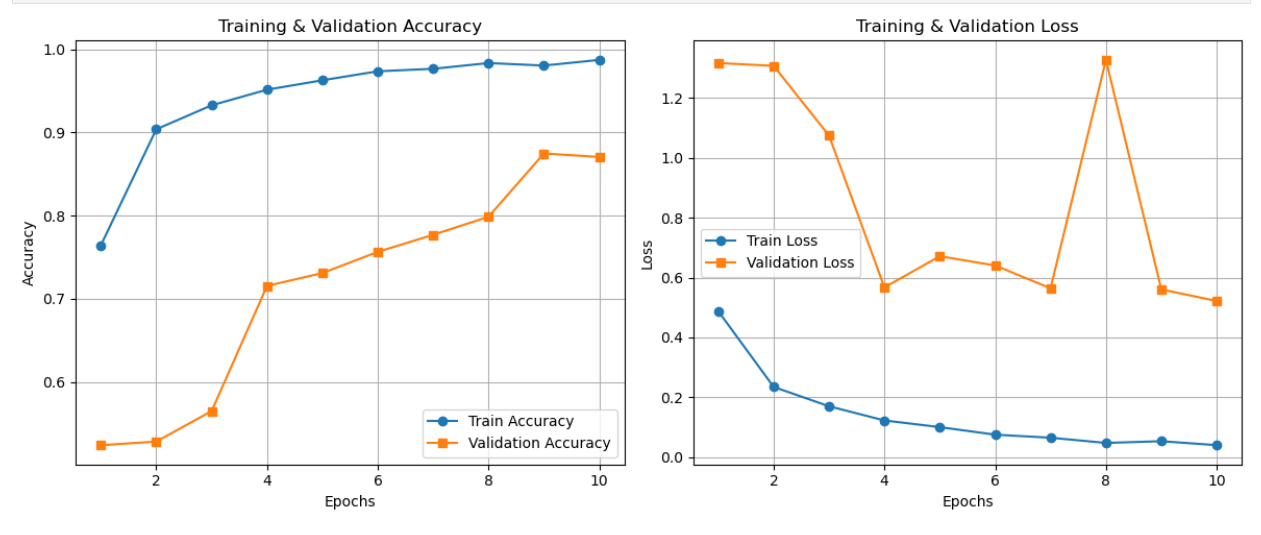
There are two main types of ensemble strategies:

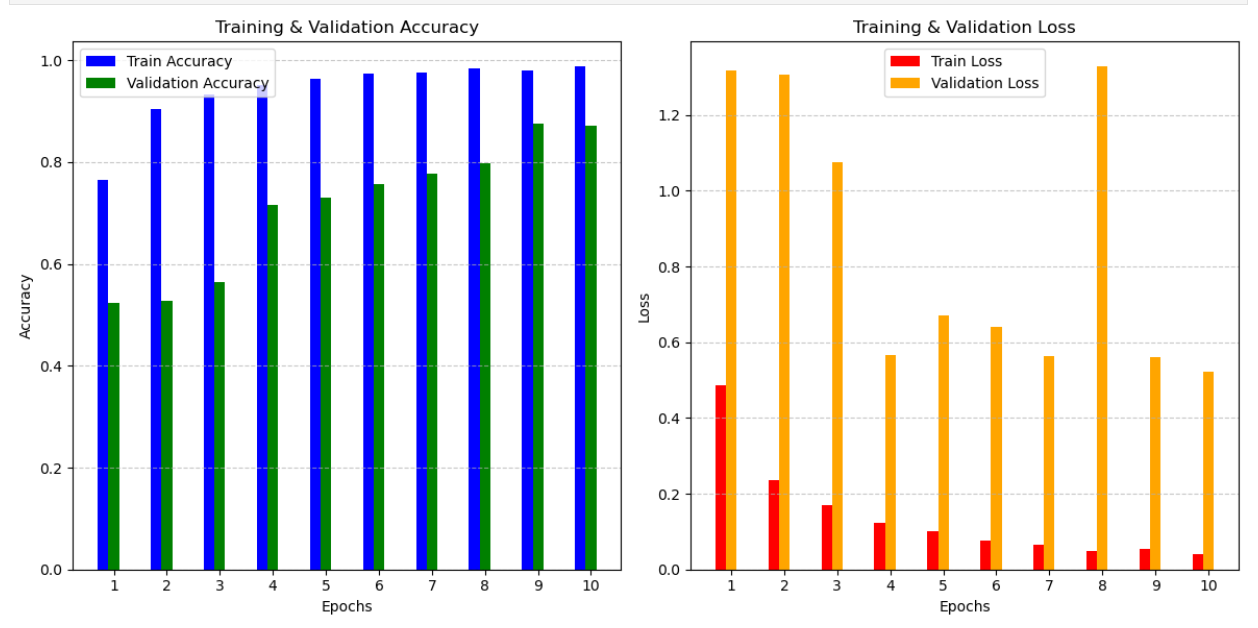
1. Bagging (Bootstrap Aggregating) – Multiple models are trained independently, often on different subsets of the training data, and their predictions are combined through methods like majority voting (for classification) or averaging (for regression). Random Forest is a classic example of bagging.
2. Boosting – Models are trained sequentially, where each model tries to correct the errors of the previous one. XGBoost is a popular boosting algorithm.
3. Stacking – Stacking is another ensemble technique where the predictions of multiple base models (level-0 models) are used as input features for a higher-level meta-model (level-1 model). The meta-model learns how to best combine the base model predictions to improve final prediction performance.

**Analysis and result: -**

**CNN model: -**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Epoch | Time (sec) | Accuracy | Loss | Val\_accuracy | Val\_loss |
| 1 | 656 | 0.7643 | 0.4876 | 0.5240 | 1.3169 |
| 2 | 605 | 0.9038 | 0.2347 | 0.5283 | 1.3075 |
| 3 | 661 | 0.9326 | 0.1702 | 0.5650 | 1.0744 |
| 4 | 615 | 0.9514 | 0.1228 | 0.7155 | 0.5672 |
| 5 | 605 | 0.9627 | 0.1005 | 0.7309 | 0.6716 |
| 6 | 601 | 0.9735 | 0.0753 | 0.7564 | 0.6405 |
| 7 | 607 | 0.9765 | 0.0651 | 0.7769 | 0.6543 |
| 8 | 587 | 0.9834 | 0.0477 | 0.7986 | 1.3274 |
| 9 | 583 | 0.9804 | 0.0535 | 0.8747 | 0.5606 |
| 10 | 568 | 0.9872 | 0.0402 | 0.8705 | 0.5219 |



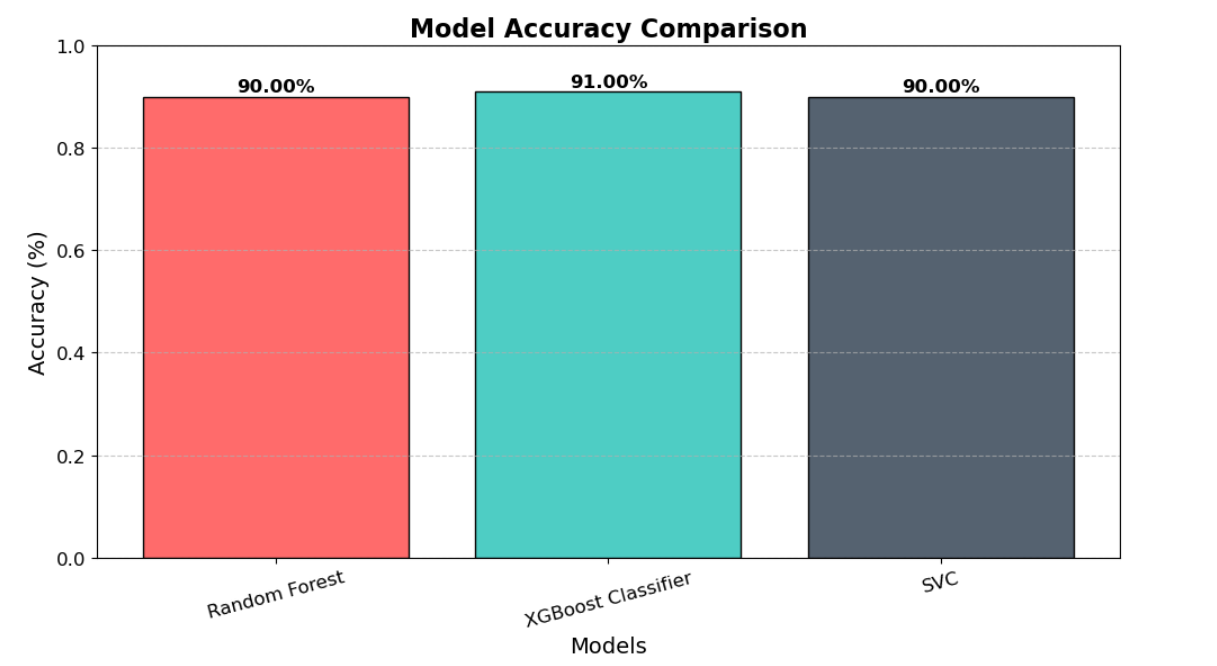


The training accuracy increases rapidly and stabilizes around 98-99% after a few epochs, meaning the model is learning very well on the training dataset. The validation accuracy improves steadily and reaches around 88-89% but does not match the training accuracy. There is a visible gap between training and validation accuracy, which indicates slight overfitting—the model performs better on training data than on validation data.

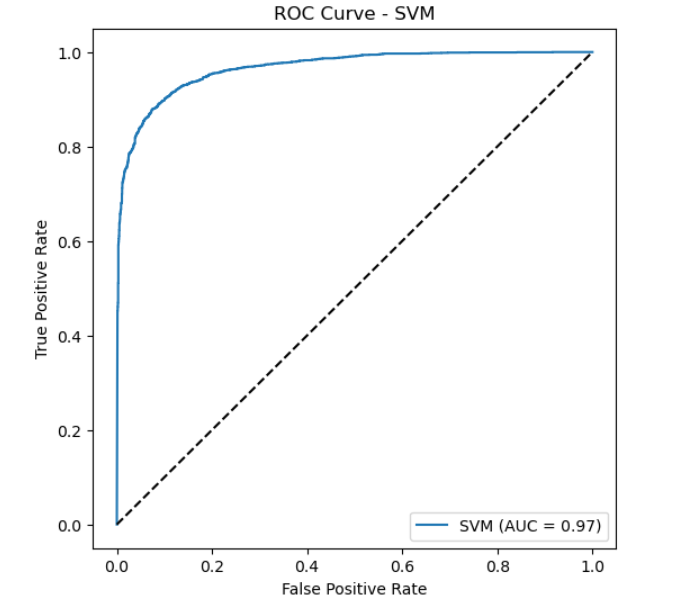
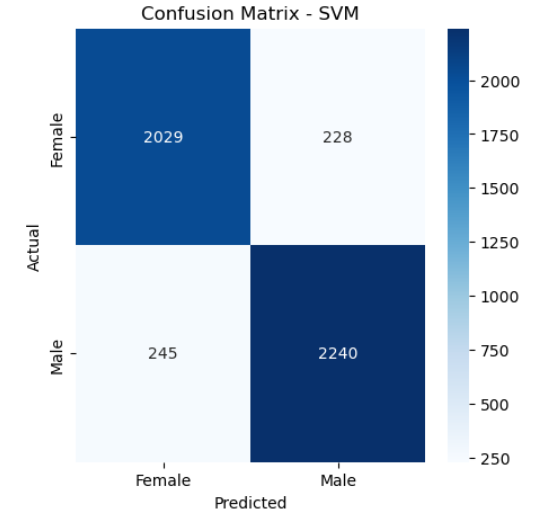
The training loss decreases smoothly and consistently, meaning the model is minimizing error on training data effectively. The validation loss is a bit unstable and fluctuates (you can see spikes in some epochs, especially around epoch 8). This fluctuation in validation loss is a clear sign of overfitting or noise in the validation data. It means the model is learning the training data very well but is struggling to generalize to unseen validation data.

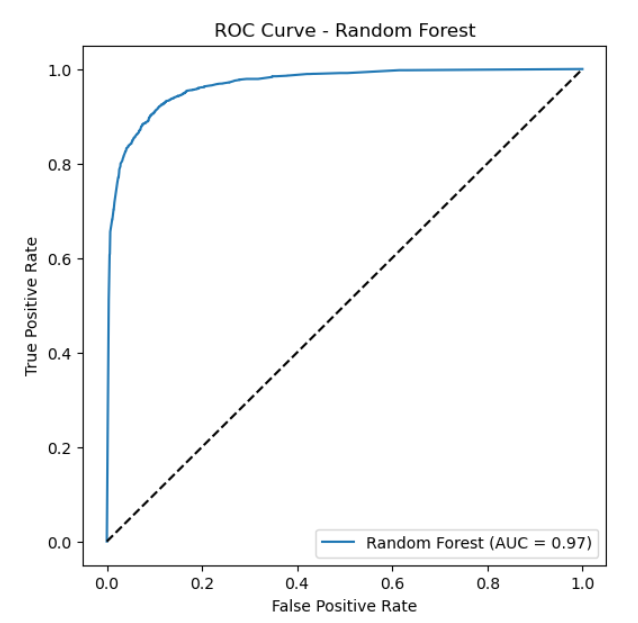
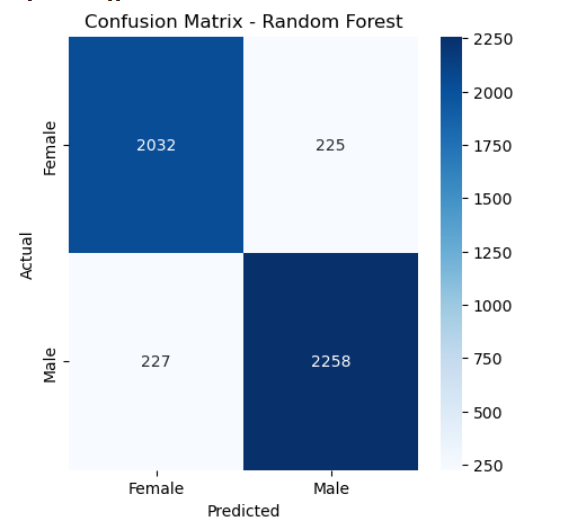
**Machine Learning model: -**

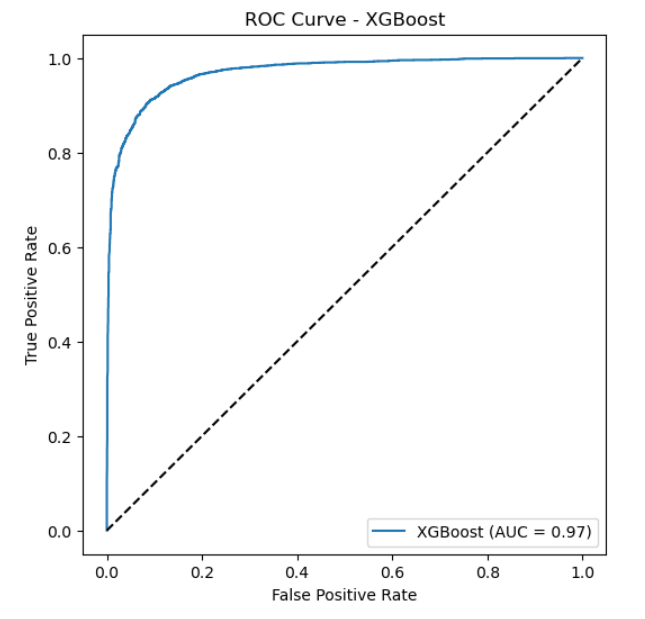
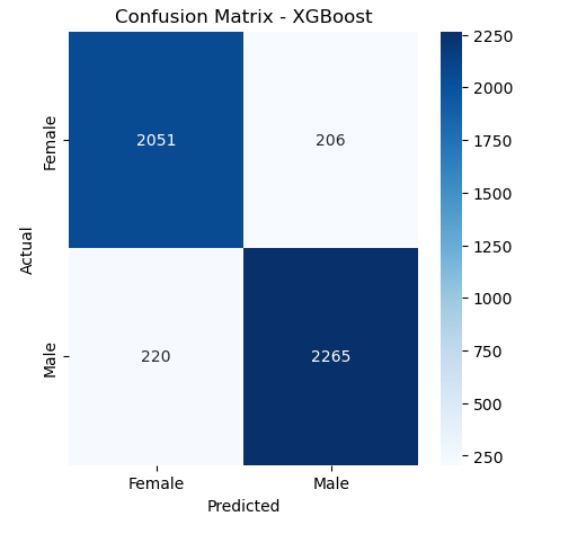
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1-score | ROC curve |
| Random Forest | 0.90 | 0.91 | 0.91 | 0.91 | 0.97 |
| SVC | 0.90 | 0.91 | 0.90 | 0.90 | 0.97 |
| XGBClassifier | 0.91 | 0.92 | 0.91 | 0.91 | 0.97 |



Random Forest, SVM, and XGBoost—shows that all models perform exceptionally well in gender classification, with accuracy scores of 90% to 91%. The XGBoost classifier outperforms the others slightly, achieving the highest accuracy (91%), precision (92%), recall (91%), and F1-score (91%), indicating it is the most balanced and reliable model in minimizing both false positives and false negatives. Both Random Forest and SVM also show strong results with 90% accuracy and 91% precision, but SVM has a slightly lower recall (90%). Additionally, all three models exhibit an excellent ROC-AUC score of 0.97, demonstrating their strong capability to distinguish between male and female classes effectively. Overall, the XGBoost model is the best-performing classifier among the three.







**Conclusion: -**

In this gender classification project, a deep learning-based Convolutional Neural Network (CNN) and three machine learning models—Random Forest, SVM, and XGBoost Classifier—were developed to predict gender from facial images. The CNN model achieved a high training accuracy of 98-99% but a lower validation accuracy of 88-89%, indicating slight overfitting, as seen from the fluctuating validation loss. In comparison, the machine learning models performed consistently well, achieving accuracies around 90-91%, with precision, recall, F1-score, and ROC-AUC values between 0.90-0.97. Among them, XGBoost Classifier showed the best performance with an accuracy and F1-score of 0.91. For real-time validation, 15 male and 15 female images were tested. The models correctly predicted 13 out of 15 male images and 11 out of 15 female images, with some misclassifications. Despite these, the models exhibited strong and reliable gender classification performance. Overall, this study concludes that combining CNN and ML models offers a robust and efficient solution for gender classification, achieving high accuracy while handling unseen data effectively, with minimal errors.

**Future work: -**

* **Data Augmentation & Regularization: -**  
  Apply data augmentation techniques (rotation, flipping, brightness adjustment) and regularization methods to reduce overfitting and improve model generalization on unseen data.
* **Hyperparameter Tuning & Model Optimization:**  
  Fine-tune the parameters of CNN, Random Forest, SVM, and XGBoost models using techniques like Grid Search or Random Search to achieve better accuracy and reduce misclassifications.
* **Real-Time Multi-Face Detection & Gender Counting:**  
  Integrate the gender classification model with a face detection system to automatically count and classify the number of males and females in a group photo or live camera feed.
* **Real-Time Video Stream Integration:**  
  Extend the system to perform **real-time gender classification** from live webcam / video streaming using OpenCV.
* **Explainability & Model Interpretation:**  
  Use Grad-CAM or SHAP to visualize which facial features the model is focusing on while predicting gender. This will make the model more transparent and explainable.

**References: -**

Adene, Gift, et al. "Detection and Classification of Human Gender into Binary (Male and Female) Using Convolutional Neural Network (CNN) Model." *Asian Journal of Research in Computer Science* 17.6 (2024): 135-144.

Fekri-Ershad, Shervan. "Developing a gender classification approach in human face images using modified local binary patterns and tani-moto based nearest neighbor algorithm." *arXiv preprint arXiv:2001.10966* (2020).

Muthukumar, Vidya, et al. "Understanding unequal gender classification accuracy from face images." *arXiv preprint arXiv:1812.00099* (2018).

Srinivasan, Sharadha, and Madan Musuvathi. "GenderRobustness: Robustness of Gender Detection in Facial Recognition Systems with variation in Image Properties." *arXiv preprint arXiv:2011.10472* (2020).

Wu, Wenying, et al. "Gender classification and bias mitigation in facial images." *Proceedings of the 12th ACM Conference on Web Science*. 2020.