# PROJECT OVERVIEW: ADVANCED DEEPFAKE DETECTION WITH DEEP LEARNING MODELS

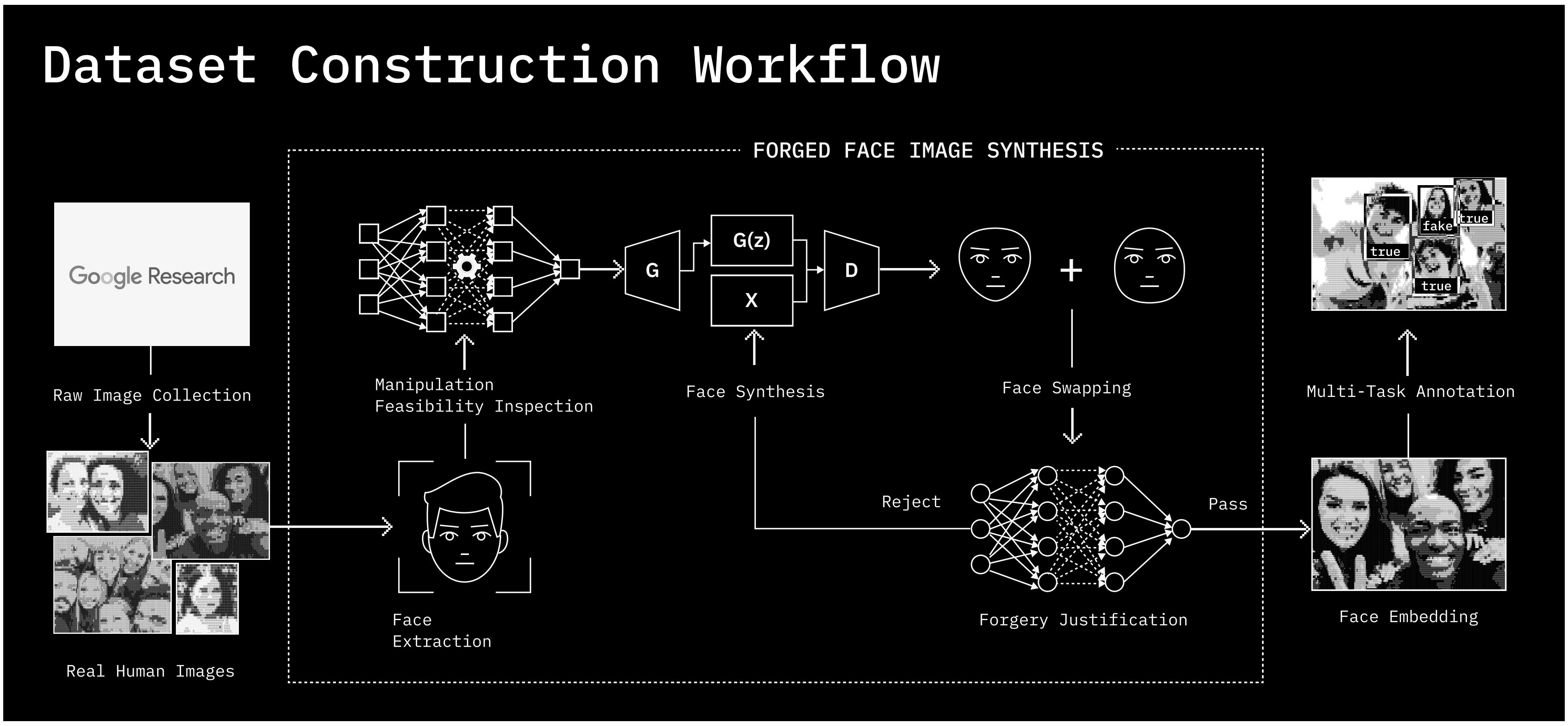
In this project, we developed a cutting-edge system for detecting deepfake using deep learning techniques. Below is a detailed description of the process and the impressive results we achieved

## 1. Dataset Construction:

The foundation of the project was a dataset designed by [1] specifically for face forgery detection. The dataset creation involved three key steps:

- **Raw Image Collection:** Gathered real human face images and manually selected high-quality examples for the training process.
- **Forgery Synthesis:** For each real face image, generated forged faces by swapping identities. This process was repeated until the forged faces could reliably fool a basic classifier.
- **Multi-task Annotation:** Comprehensive annotations were applied to the faces, labeling various attributes necessary for training the models effectively.

**Fig1. Dataset construction workflow**



## 2. Image Preprocessing:

Prior to training, the images underwent a series of preprocessing steps aimed at improving model performance and generalization. The images were first resized to a uniform dimension to standardize the input for the models. To further enhance the diversity of the training data and reduce overfitting, we applied several data augmentation techniques. These included random horizontal flips, slight rotations, and variations in brightness, contrast, and saturation. By introducing these random transformations, we ensured that the model would learn to be robust to different orientations and lighting conditions.

Additionally, we normalized the images to align with the pre-training statistics of popular deep learning models, allowing for more efficient and stable training. Finally, a technique known as **Random Erasing** was applied to simulate occlusions and improve the model's ability to handle partially obscured faces in real-world scenarios

## 3. Model Training:

For the face deepfake detection task, we experimented with a variety of deep learning models:

- **Pre-trained Models:** We utilized advanced pre-trained models, such as Xception and other state-of-the-art architectures, known for their exceptional feature extraction capabilities. These models, pre-trained on large image datasets, were fine-tuned on our face forgery dataset to leverage their powerful visual recognition features.
- **Custom CNN Models:** In addition to pre-trained networks, we designed custom convolutional neural networks (CNNs) to compare their performance with the pre-trained models. This approach allowed us to test different architectures and determine the most effective solution.

After training these models, we achieved prediction accuracies ranging from **94% to 96%** on the test dataset.

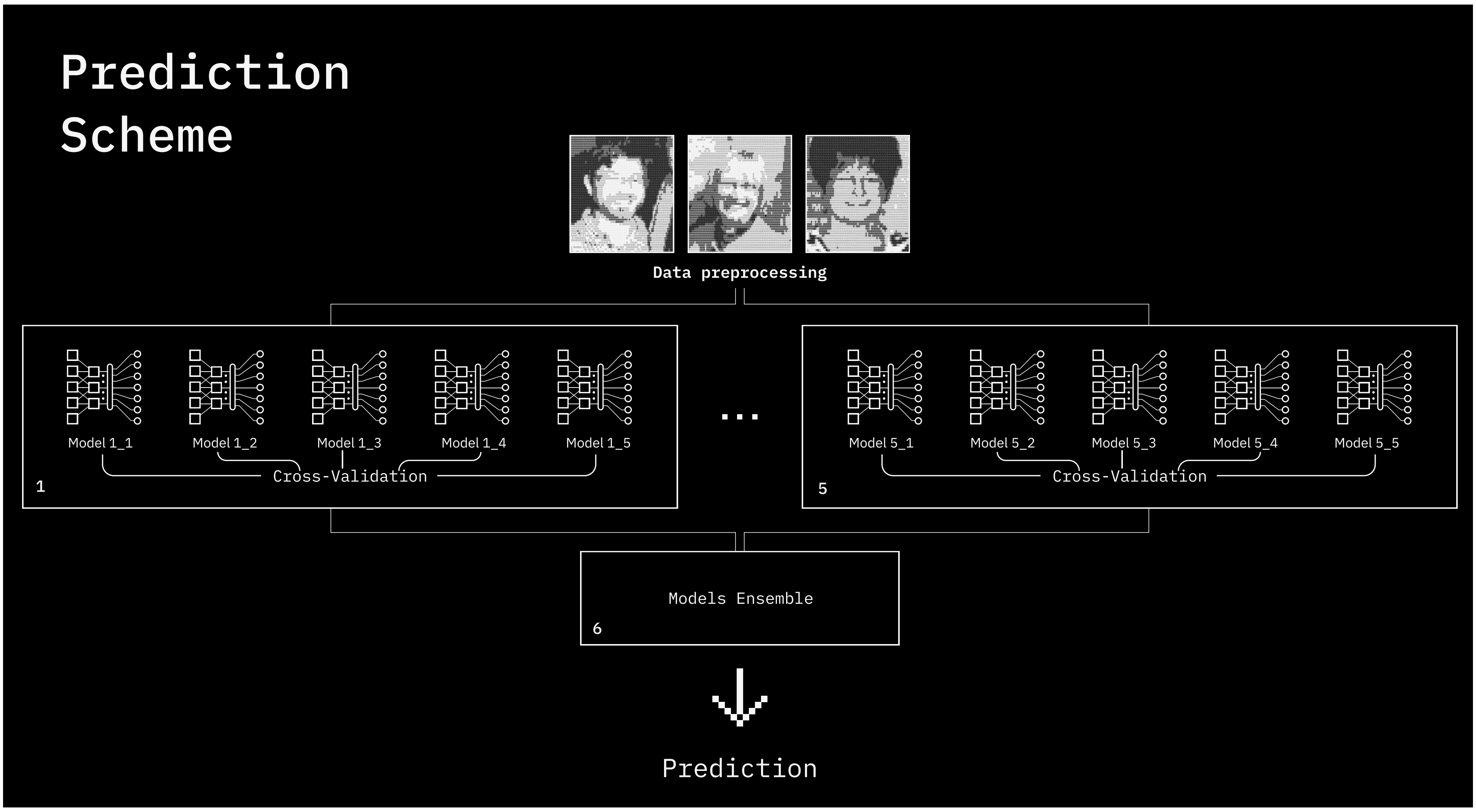
## 4. Cross-validation & Model Stability:

To ensure the models were stable and generalized well, we implemented **Stratified K-Fold Cross-Validation.** Unlike standard cross-validation, stratified folding ensures that the data is split in a way that maintains the same distribution of classes (real vs. forged) across all folds. This approach helped reduce any potential bias in the data splits, ensuring the model was tested on a diverse and representative set of images, improving its robustness against noise.

## 5. Model Ensembling:

To further enhance performance, we employed an **ensemble learning** approach, combining multiple trained models. By aggregating the predictions from each individual model, the ensemble method produced a more accurate and reliable outcome. This technique capitalized on the strengths of each model and improved overall performance.

**Fig 2. Scheme of prediction**

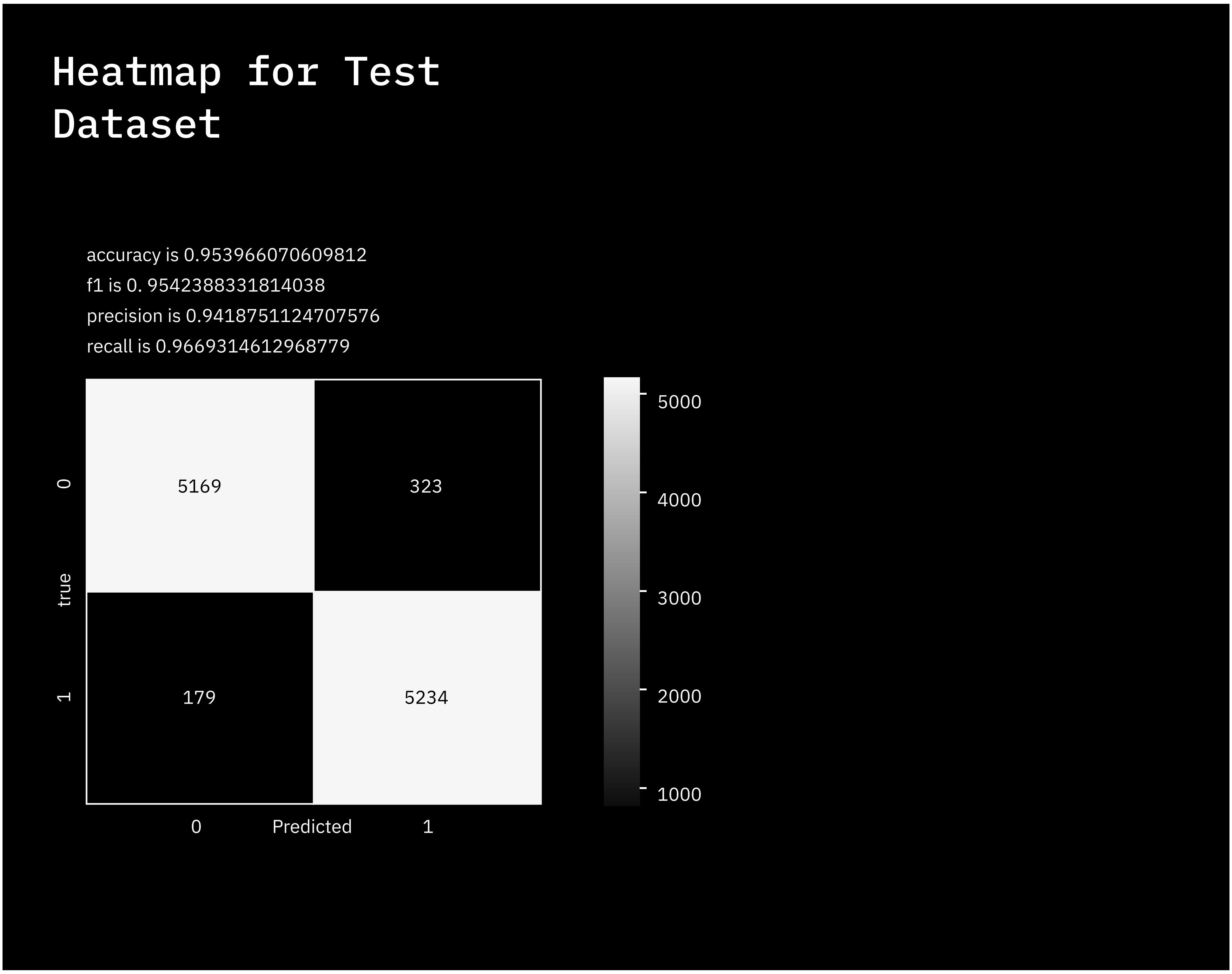


## 6. Results on Test Data:

The ensemble model achieved outstanding results on the test dataset:

| | |
|---|---|
| Accuracy | 95.40% |
| F1 Score: | 95.42% |
| Precision: | 94.19% |
| Recall: | 96.69% |
| AUC (Area Under the Curve): | 99.16% |

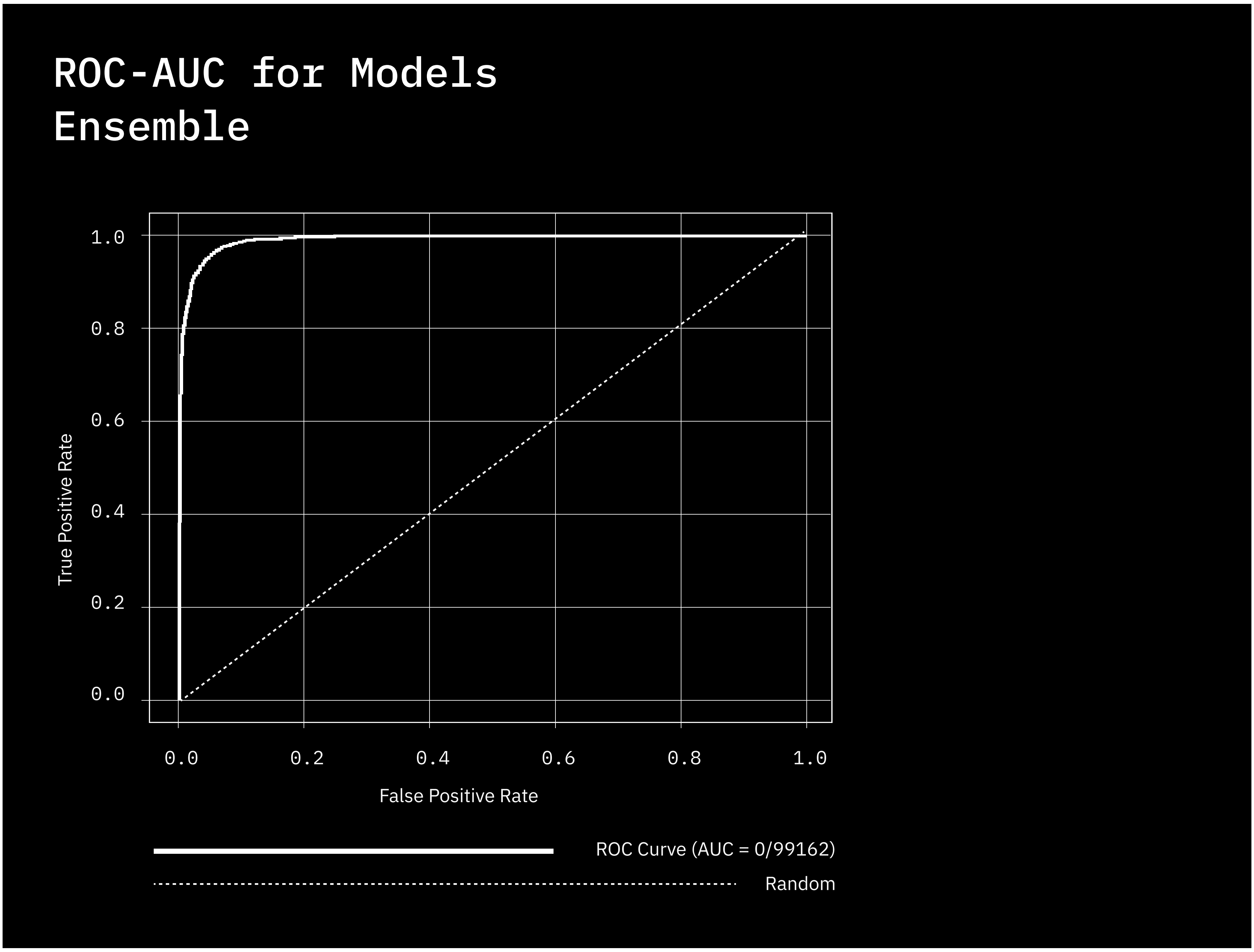
**Fig3. Heatmap for test dataset**



Heatmap for Test Dataset

```
accuracy is 0.953966070609812
f1 is 0. 9542388331814038
precision is 0.9418751124707576
recall is 0.9669314612968779
```

Fig 4. ROC-AUC for Models Ensemble



These results highlight the effectiveness of our approach in detecting deepfakes, achieving high accuracy, excellent recall, and an AUC score nearing perfection

## Conclusion:

Our face forgery detection system, built using a combination of advanced techniques and deep learning models, has proven to be highly effective and robust. With sophisticated image preprocessing, and a strategic ensemble approach, we've developed a solution that performs exceptionally well in detecting deepfakes with high accuracy and reliability. This system is ideal for applications in security, identity verification, and other fields where face is critical.

## Sources

1. Le T. N. et al. Openforensics: Large-scale challenging dataset for multi-face forgery detection and segmentation in-the-wild //Proceedings of the IEEE/CVF international conference on computer vision. – 2021. – C. 10117-10127.