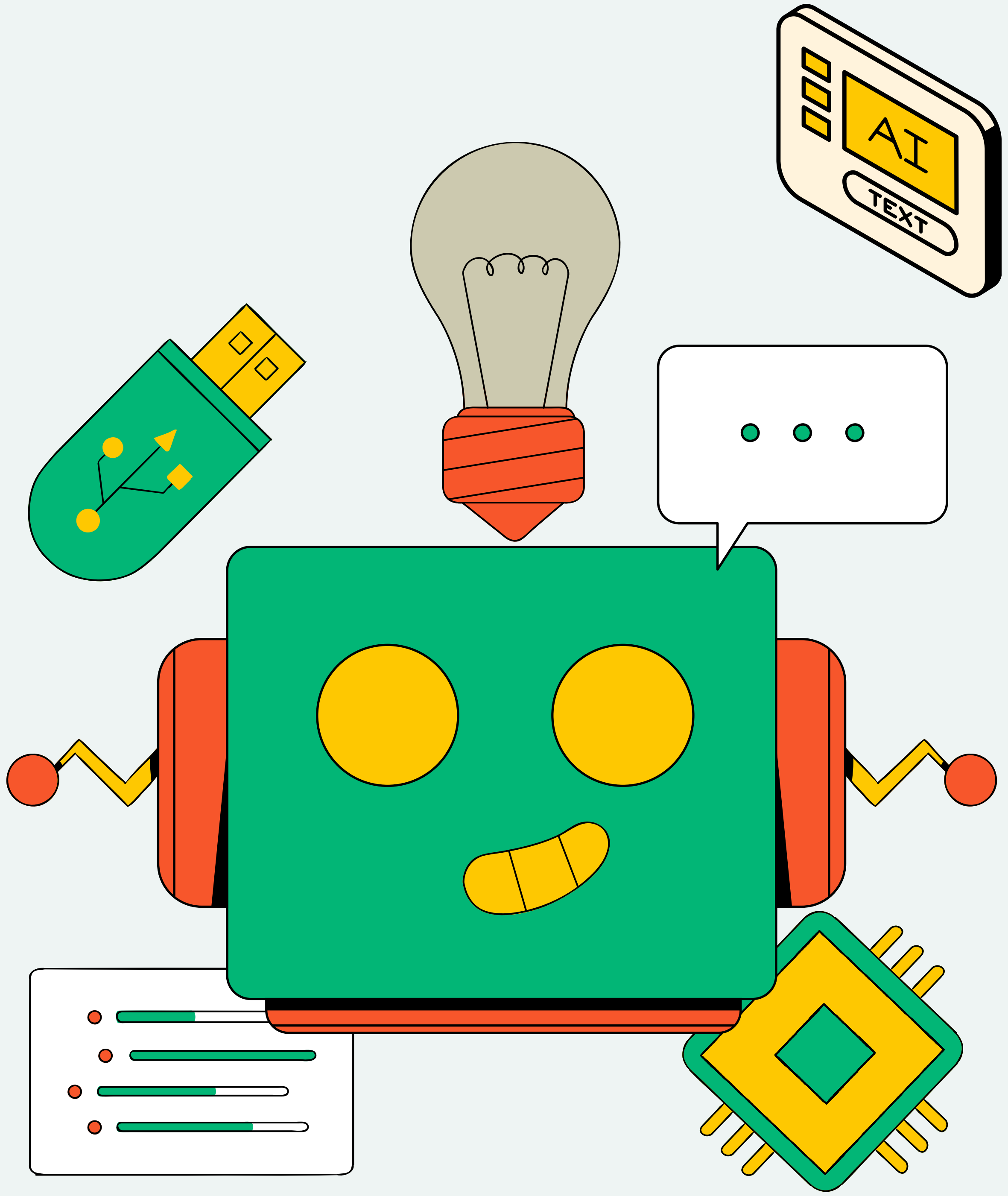
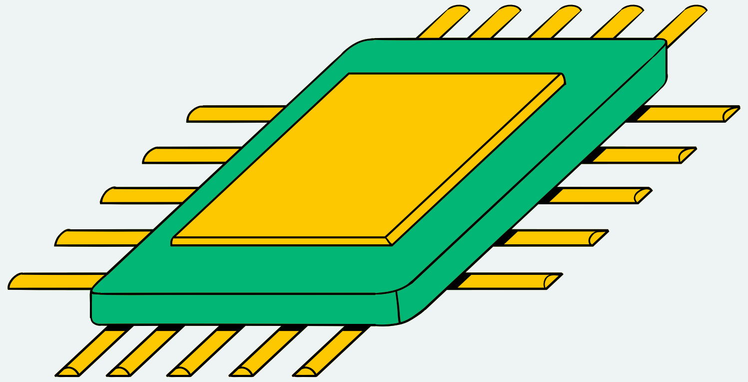


DEEPFAKE DETECTION USING A HYBRID MODEL

VGG16 AND BEIT VISION TRANSFORMER

PRESENTED BY:

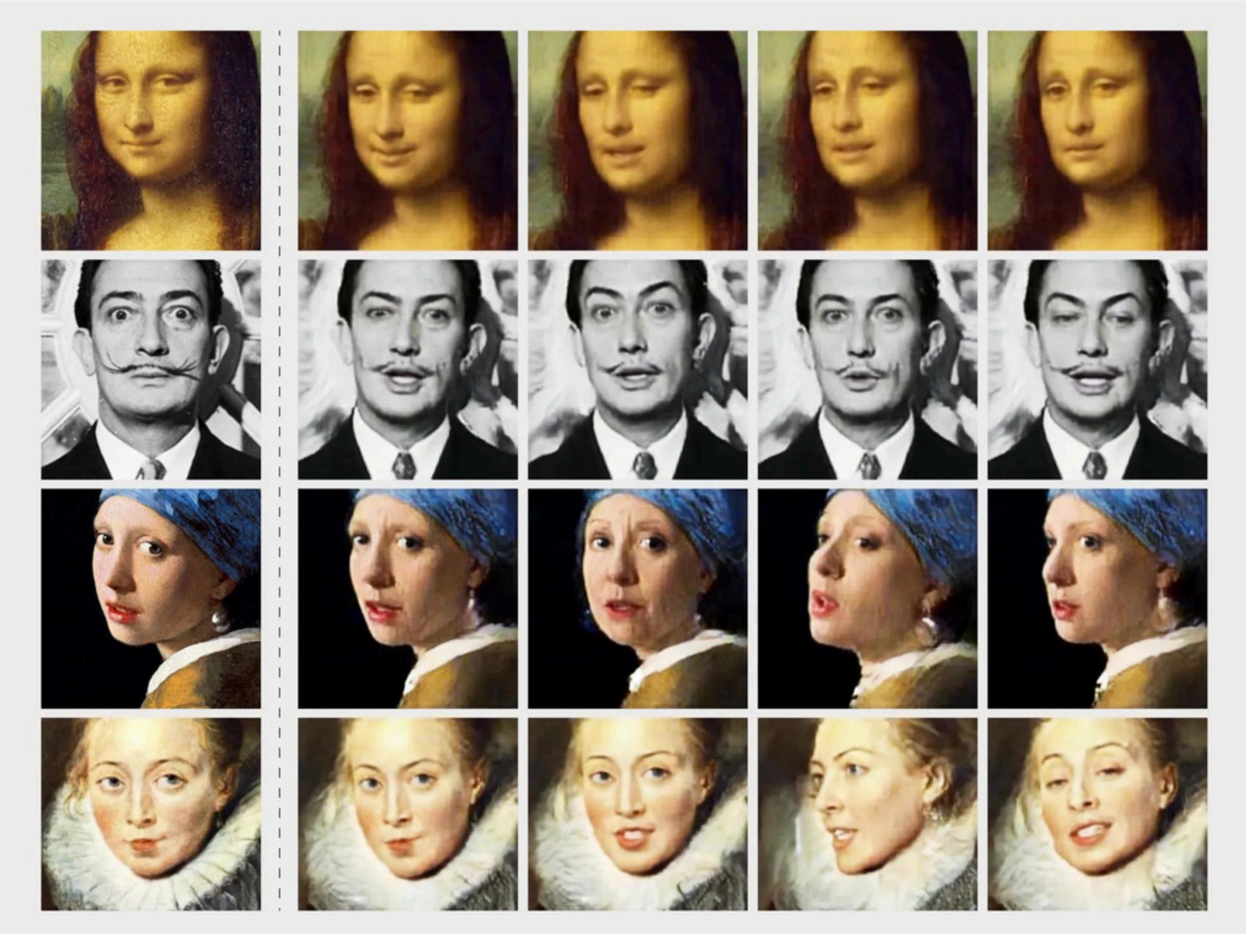
DANIELE PANCOTTINI



PRESENTATION OUTLINE

- Introduction
- Related works
- Proposed model
- Datasets and metrics
- Implementation details
- Experimental results
- Conclusion and future works

INTRODUCTION



- Deepfakes are AI-generated media that can manipulate reality, posing serious challenges to security and media integrity.

- This project introduces a hybrid deepfake detector combining **VGG16** and the **BEiT Vision Transformer** to extract complex visual features and classify images.

WHY A HYBRID MODEL?

- Complementary nature of CNNs and Transformers: VGG16 captures **local features**, while ViT captures **global context** (long-range relationships).

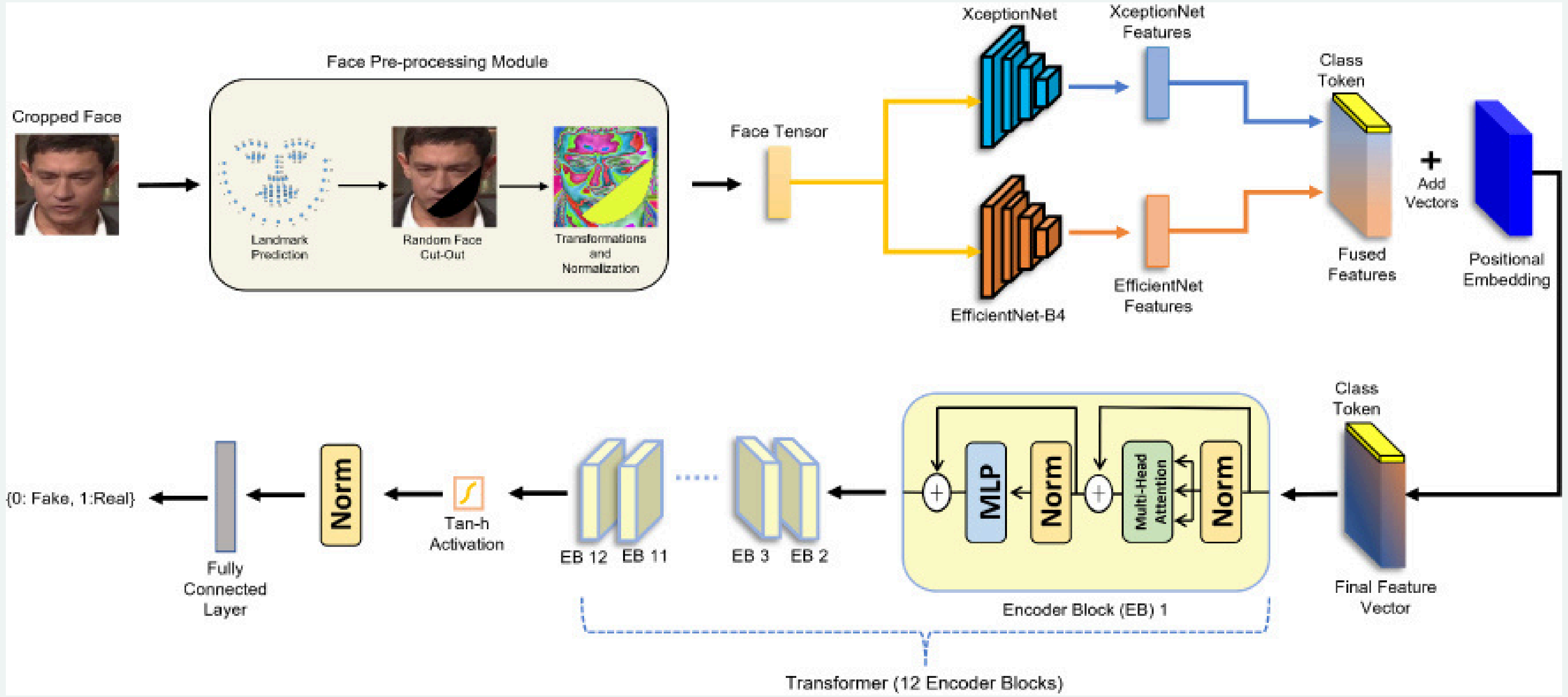
- Combining both models allows leveraging local and global features for **better accuracy**

RELATED WORKS

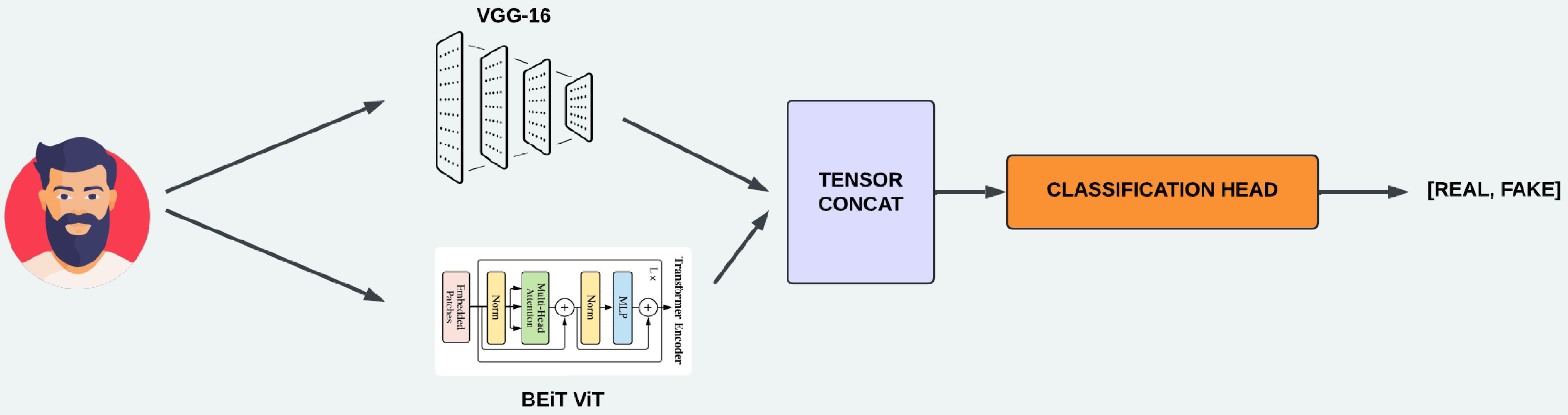
Hybrid models combining CNN and Vision Transformers:

- Wang, Y., Huang, L., & Zhu, X. (2022). **Improved Hybrid CNN-Transformer Network for Deepfake Detection**

- Luo, Y., Zhang, S., & Wang, Y. (2021). **Deepfake Detection with Temporal-Aware CNN and Vision Transformer**



PROPOSED MODEL



- The same image is fed into **both VGG16 and BEiT models**
- Features from VGG16 and BEiT are **concatenated**
- The concatenated features are passed through the **classification head**

IMPLEMENTATION DETAILS

Feature Extractors:

- **Pretrained** VGG16 and BEiT (ViT) are used as **feature extractors**, with their **parameters frozen** during training

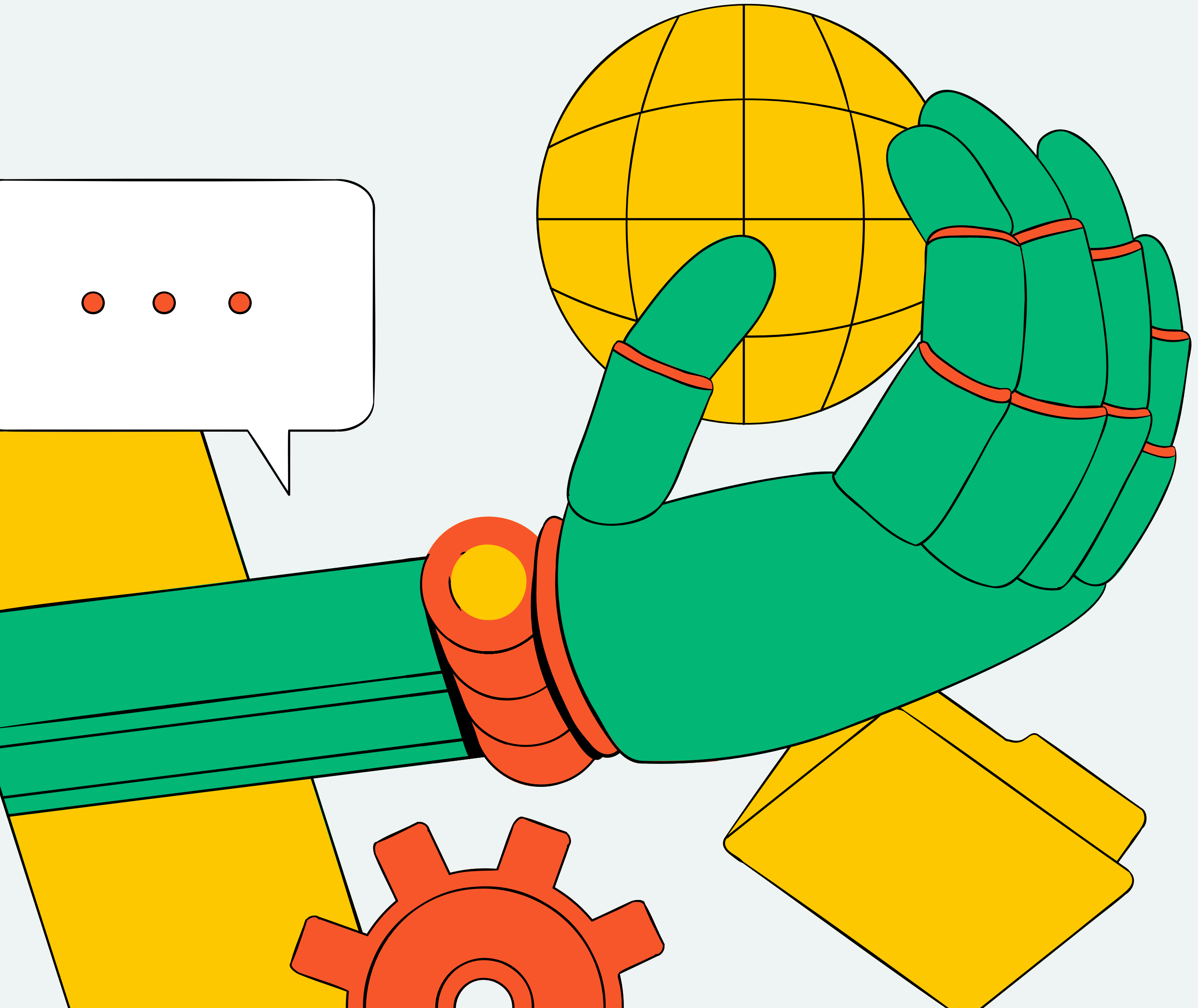
Classification Head:

- A **fully connected layer** (FCL) of size **1768 x 256** with ReLU activation function.
- A **fully connected layer** (FCL) of size **256 x 2**, returning **unsoftmaxed** prediction logits for use with the **CrossEntropy** loss function

Training:

- Training for **six epochs**, with a learning rate of **1e-4**, using the **Adam optimizer** and **CrossEntropy** loss function

DATASETS

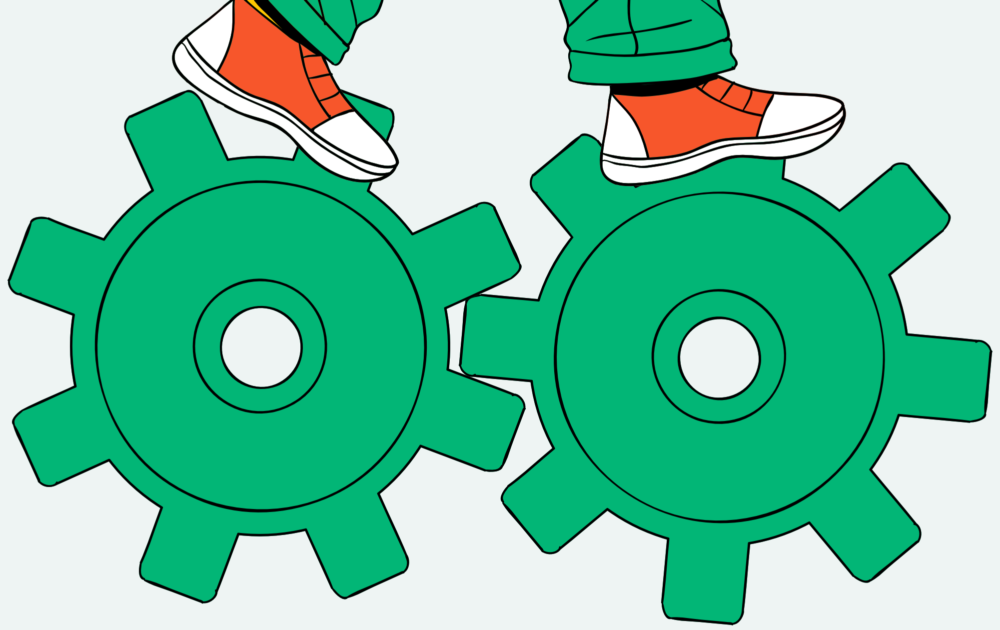


Training Dataset:

- UNICT **Deepfake Detection Challenge Training** set

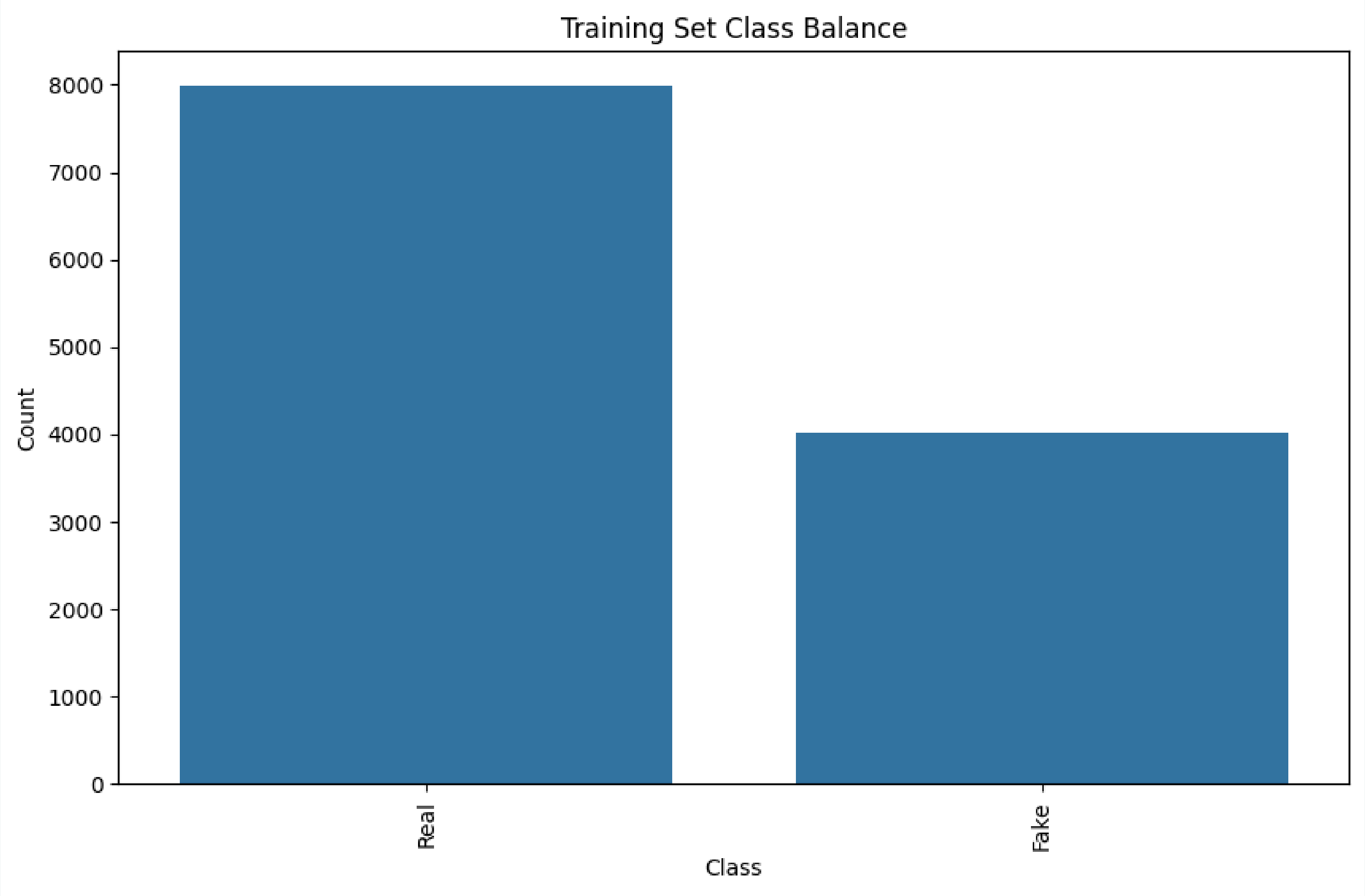
Evaluation Datasets:

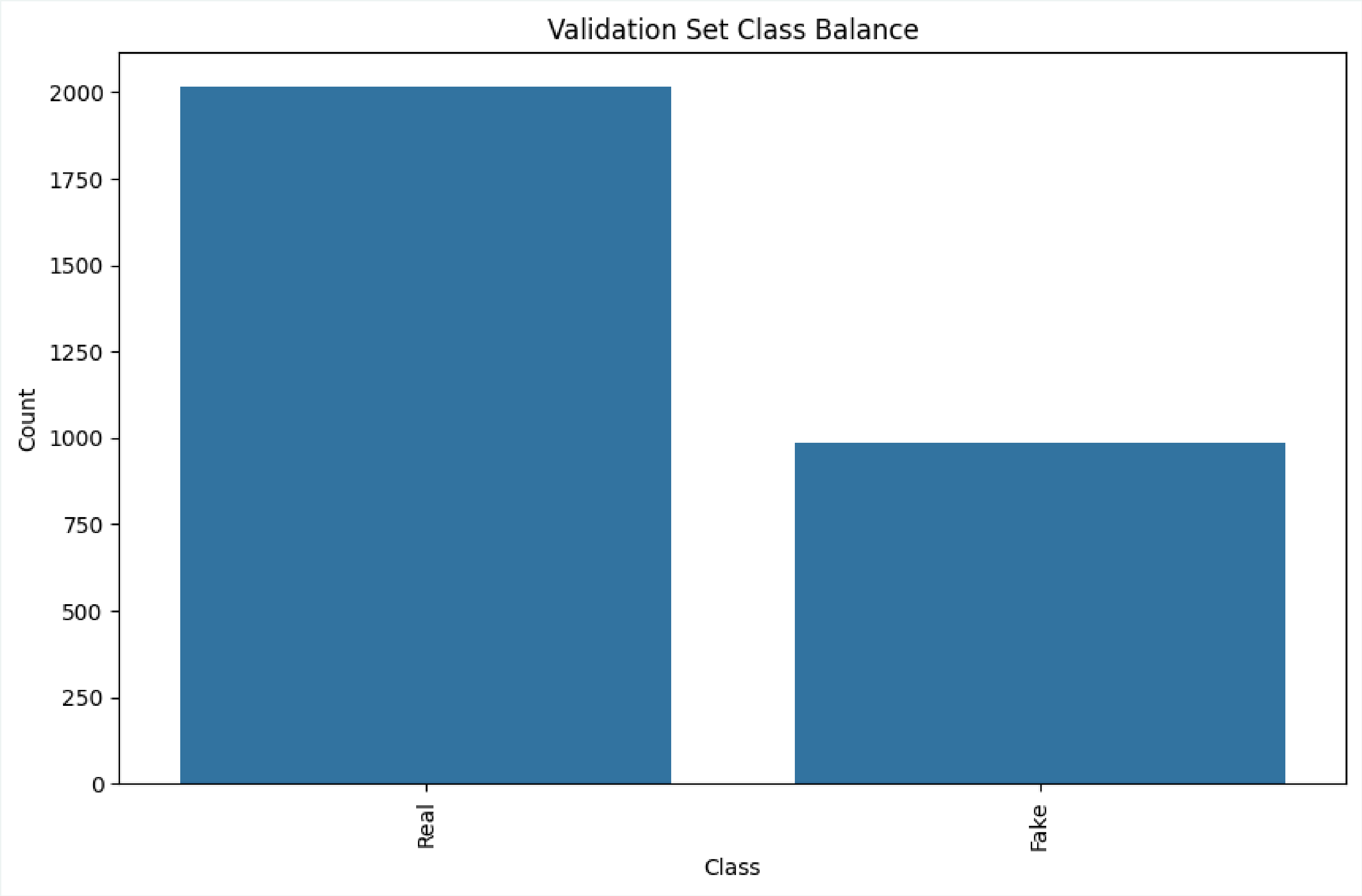
- **FFHQ** (Flickr-Faces-HQ) Dataset
- UNICT **Deepfake Detection Challenge Test** set



TRAINING DATASET

- Dataset containing **15.000 images: 10.000 real** images and **5.000 fake** images
- Random training split of **80%** of the entire dataset and **20%** for the validation set





EVALUATION DATASETS

- Randomly Subsampled **FFHQ Dataset**: Approximately **35.000 real images**
- **UNICT Deepfake Detection Challenge Test Set**: Approximately **7,000 images** (5,000 fake and 2,000 real).

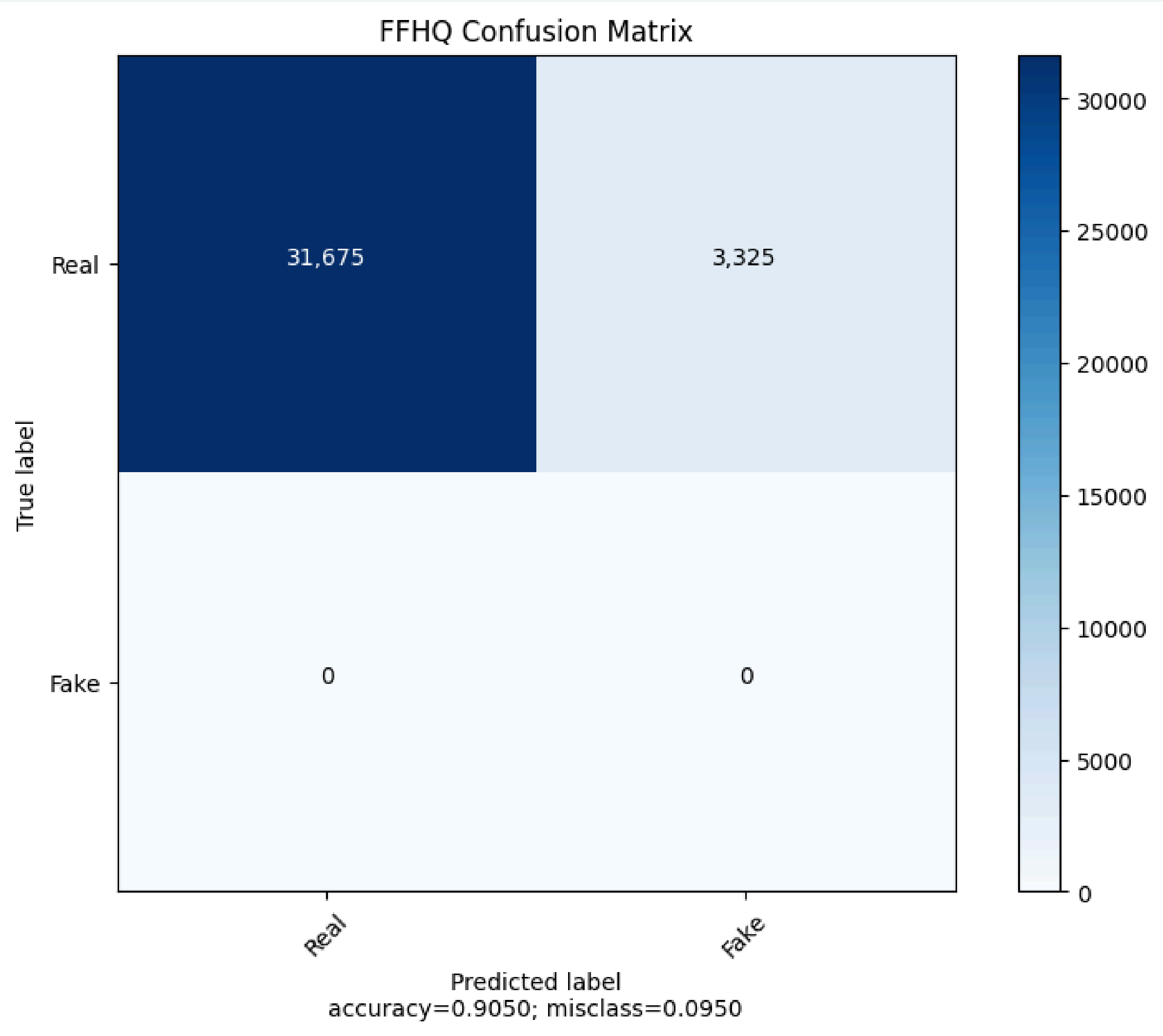
EVALUATION METRICS

- **Accuracy**: the ratio of correctly predicted instances (both true positives and true negatives) to the total instances.
- **Precision**: the proportion of true positive predictions out of all positive predictions made by the model.
- **Recall**: the proportion of true positive predictions out of all actual positive instances.
- **F1-score:** the harmonic mean of precision and recall, providing a balanced measure when precision and recall are uneven.

MODEL RESULTS

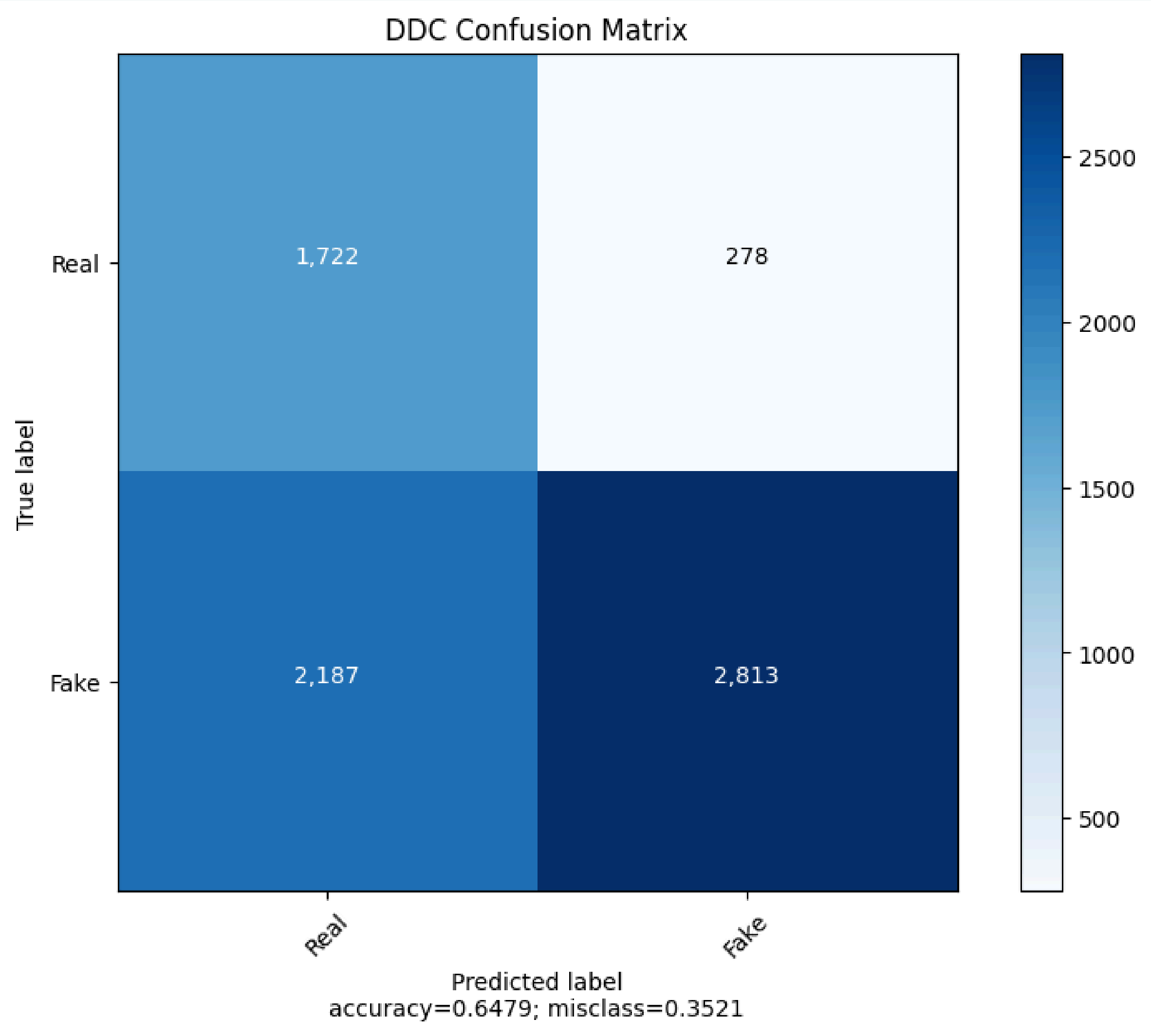
FFHQ Dataset

Accuracy: 90% **Precision**: 100%
Recall: 90.50% **F1-score**: 95.01%



DDC Dataset

Accuracy: 64.79% **Precision**: 77.59%
Recall: 64.79% **F1-score**: 66.32%



COMPARING RESULTS

Ranking list of the best seven models evaluated using the DDC dataset

Ranking	Team Name	Accuracy (%)
#1	VisionLabs	93.61%
#2	DC-GAN (Amped Team)	90.05%
#3	Team Nirma	75.38%
#4	AIMH Lab	72.62%
#5	PRA Lab—Div. Biometria	63.97%
#6	Team Wolfpack	40.61%
#7	SolveKaro	36.85%

- The hybrid approach achieved an **accuracy of 64.79%**, resulting in a **fifth-place position** overall

CONCLUSION AND FUTURE WORK

Future works about this project would involve performance improvements:

- **Expand Dataset**: increase the size of the training dataset beyond the current **15.000 samples**
- **Improve Data Quality:** address issues with **data balancing and distribution**
- **Longer Training Duration**: extend the **number of epochs** beyond the current six to improve model performance.
- **Optimize Classification Head:** implement changes to the classification head architecture to enhance accuracy