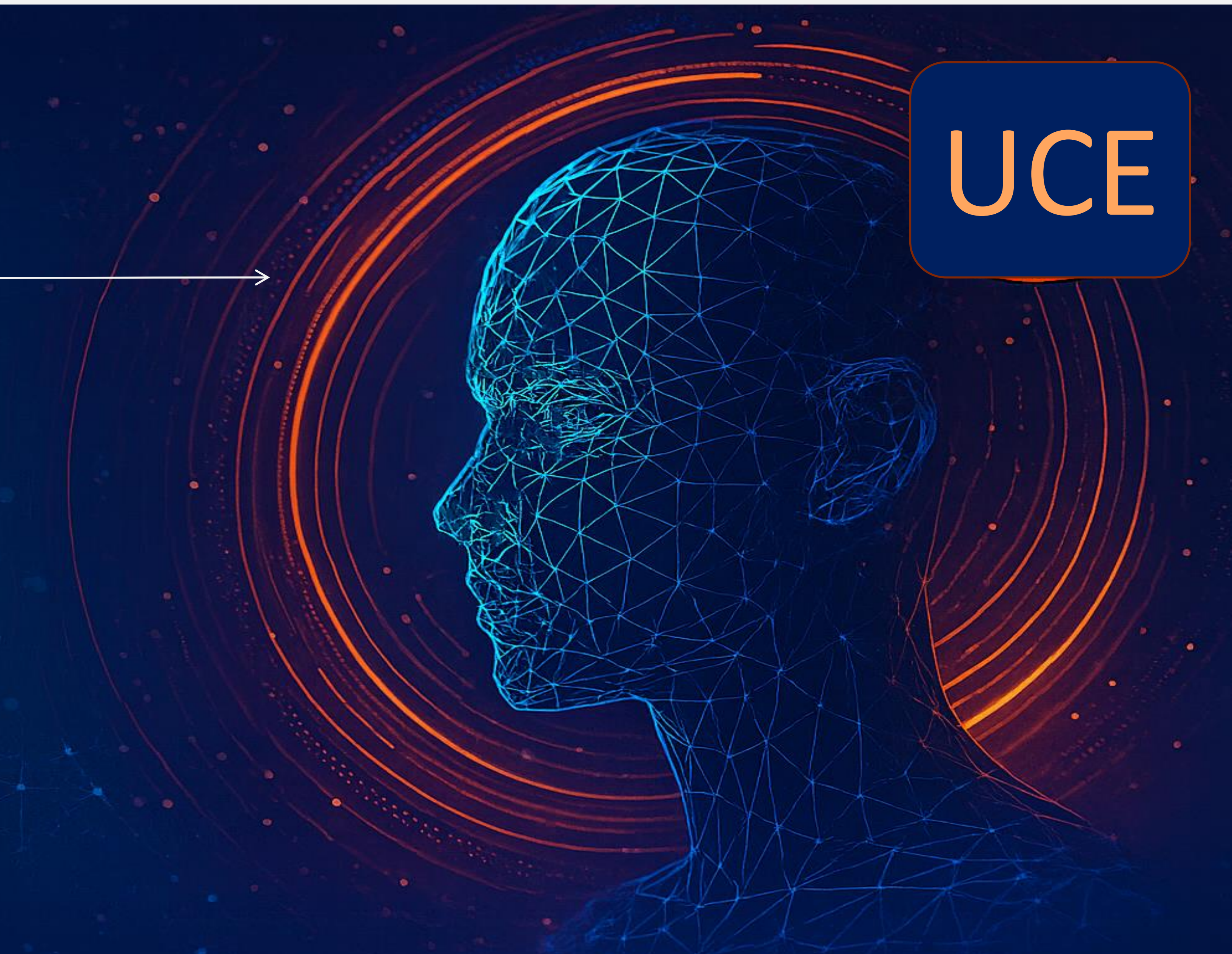
A model for Longitudinal Candidate Re-identification using Deep Learning:

a case study of UNEB’s PLE AND UCE

HENRY JOE OPIO
Department of ICT Services

UGANDA NATIONAL EXAMINATIONS BOARD



41st AEAA Conference 2025

Background

What is Person Re-Identification?

"Recognizing the Same Person Across Images, Places, and Time"

*It is the task of matching individuals across non-overlapping cameras or image datasets over time, despite changes in pose, illumination, or appearance. (*Zheng et al. 2016).

Core Challenge: Faces, clothes, and environmental conditions vary → the system must extract robust, discriminative features invariant to noise.

Why it Matters in APRE-ID: Goes beyond static verification → ensures continuity of identity across PLE → UCE exam cycles.

Background

The Growing Threat of Qualification Fraud

"Fraudulent Academic Credentials: A Global and Local Crisis"

Billions lost annually worldwide due to forged or misrepresented academic qualifications.

Uganda and Africa at large face rising cases of imposters obtaining genuine licenses with stolen identities.

Background

Identity Theft as the Weakest Link

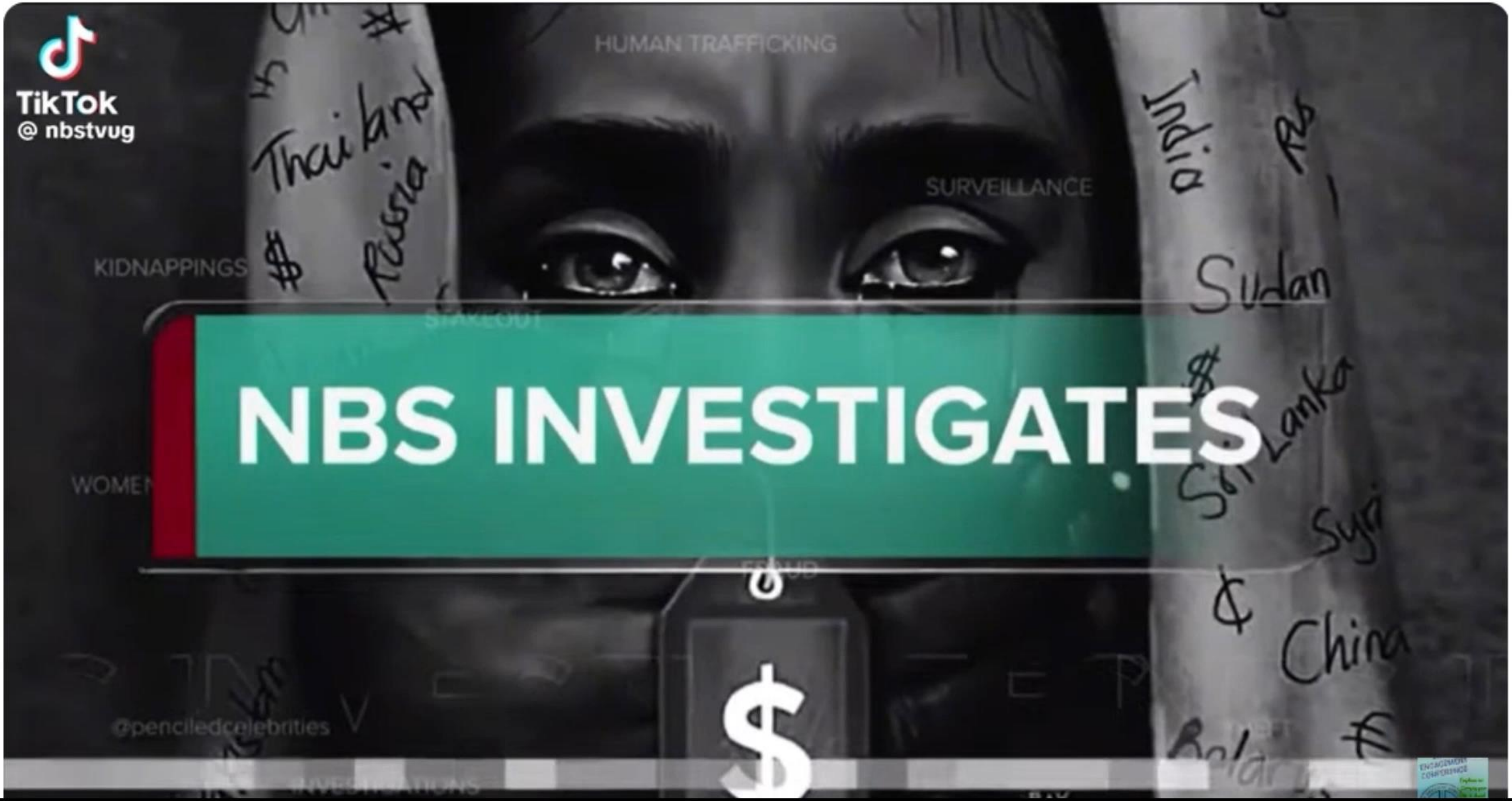
"When the Person and the Grades Do Not Match"

Traditional qualification verification relies heavily on paper documents. These are easily forged.

Imposters exploit identity theft, acquiring **genuine** grades tied to the wrong **face**.

Consequences:

- Unqualified individuals getting undeserved roles in society.
- Loss of credibility for authentic graduates.
- Massive cost to regulatory and examination bodies.


TikTok
@ nbstvug
HUMAN TRAFFICKING
SURVEILLANCE
KIDNAPPINGS
STAKEOUT
WOMEN
NBS INVESTIGATES
FRAUD
Thailand
Russia
India
Sudan
Sri Lanka
China
@penciledcelebrities
INVESTIGATIONS
$

The Problem

Given two temporally disjoint but demographically linked image databases—how can we accurately determine whether a UCE registrant is who they claim to be in the PLE database; using face-based AI, **irrespective of** age, lighting, and domain variation?

*“**Demographic linkage**” is a term used in **data integration, identity management, and re-identification tasks** to describe the process of connecting or matching records across datasets by relying on **demographic attributes.***

Our Objectives

"Beyond Documents: Verifying the Person, Not Just the Paper"

1 DEVELOP THE EMBEDDING MODEL

To develop a face-based identity embedding model robust to age progression and registration noise.

2 DEVELOP THE MATCHING SYSTEM

To design a scalable candidate matching system using a hybrid of deep metric learning and structured metadata filtering.

3 EVALUATE THE MODEL

To rigorously evaluate the system on real-world exam registration datasets from Uganda across a 4-year time gap.

4 PACKAGE FOR DEPLOYMENT

To propose deployment scenarios and policy frameworks for national education boards in LMICs.

Key Considerations

Scope

Re-Identification models can be designed to curb fraud globally across the entire formal assessment path using biometric, demographic and meta data.

This work however, concentrates on learning **facial features** for **school candidates** at Uganda's UCE who transitioned from PLE with an average **age progression** of 4-5 years.

Data Privacy

While this is an in-house research;

- Necessary permission was still sought and granted to respectfully use candidate information.
- All the candidate data sets in this experiment were used within the Board's premises.
- Any information or data on candidates used for purposes of publication or dissemination are anonymized and used with strict adherence to Uganda's Data Protection and Privacy Act(2021).

Literature Review

Deep Metric Learning for Face Recognition

- **Deep Metric Learning (DML):** Aims to learn feature embeddings where the same person has a small distance and different people are far apart.Local.

- **Feature Learning:** Focuses on learning aggregated features from parts or regions of an image to be robust against misalignment.

- **Generative Adversarial Learning (GANs):** Used to address issues like small training samples, resolution, and lighting variations.

- Recent advances favour angular margin-based formulations for superior separation. ArcFace (Deng et al., 2019) introduces an additive angular margin to the Softmax formulation, producing highly discriminative embeddings on the hypersphere.

Longitudinal Face Recognition

Age progression introduces a natural covariate shift in face verification tasks. Age-invariant face recognition (AIFR) techniques aim to mitigate intra-class variance caused by time.

In this work, we adopt ArcFace for its superior convergence, scalability, and
angular interpretability, critical for reliable verification in exam contexts.

Literature Review…

Re-Identification in Surveillance and Biometrics

- Person Re-ID has traditionally been studied in video surveillance, where individuals must be matched across camera views (Zheng et al., 2016)

AI in Educational Assessment Systems

- AI is widely used in education for adaptive testing, content recommendation, and automated grading. However, its application in assessment integrity—particularly for identity verification—remains underexplored.

Gaps in Existing Literature

Despite advancements in deep face recognition, there are no known studies applying longitudinal Re-ID to large-scale exam systems in Sub-Saharan Africa.

Specific gaps addressed by this model include:

- Lack of age-resilient biometric matching pipelines tailored to low-resource educational environments.
- Absence of real-world African datasets with cross-cycle validation.
- Limited integration of structured metadata filtering to reduce 1:N search complexity

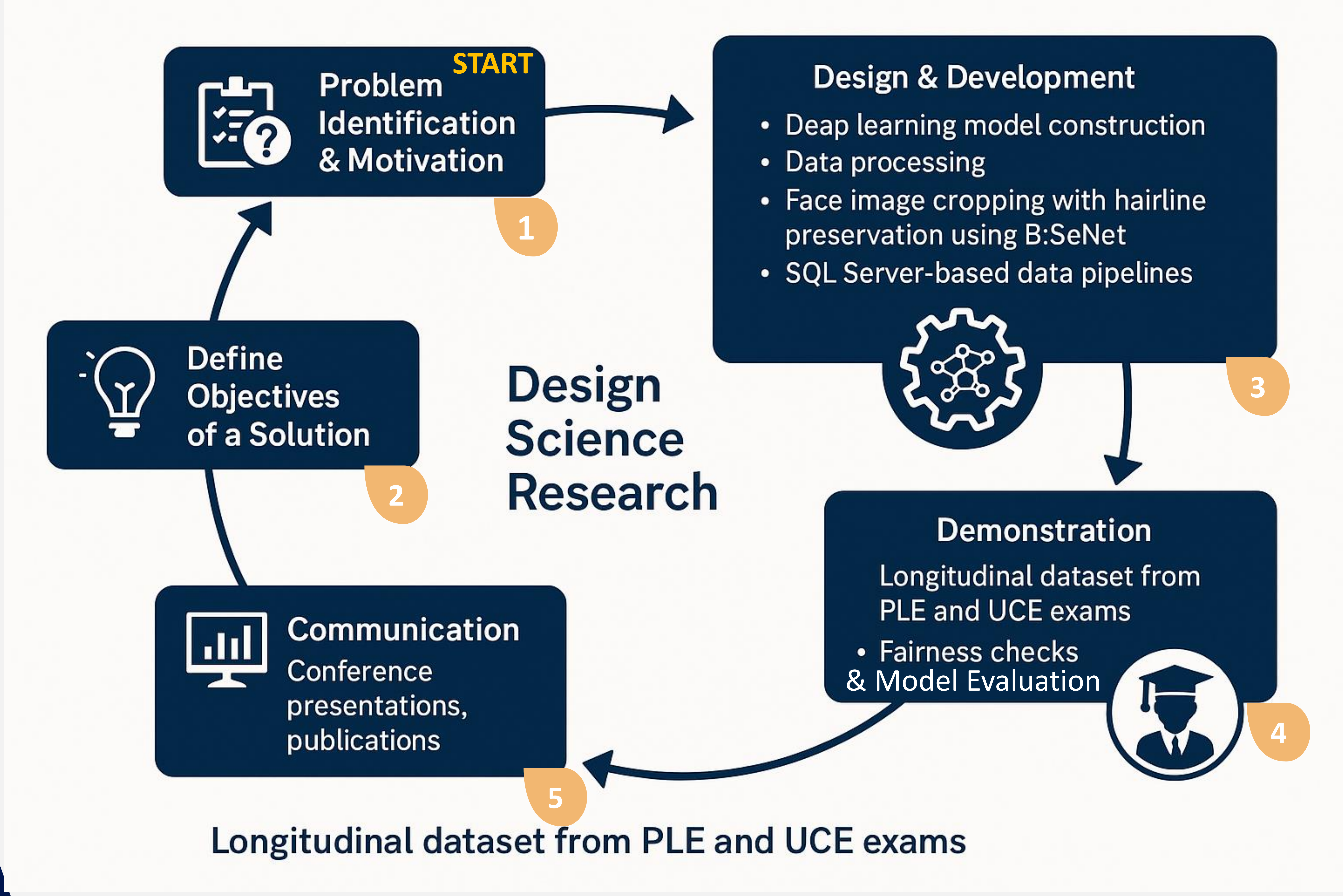
By addressing these gaps, our system bridges computer vision advances with practical educational policy challenges in Uganda and similar contexts.

Methodology

We adopt the **Design Science Research** (DSR) paradigm, emphasizing problem identification, artifact construction (i.e., the re-identification model) and utility demonstration within real-world constraints.

Design science is a systematic, problem-solving research paradigm focused on creating artifacts—such as models, methods, or systems—that solve real-world problems and generate new knowledge. It applies scientific principles and techniques to the design and evaluation of these artifacts to improve functional performance and contribute to the existing body of design knowledge.

Methodology…



PROCESSING PIPELINE

1. Facial Detection

- Multi-task Cascaded Convolutional Networks (**MTCNN)** [10] used to locate faces and landmarks.

2. Face Alignment

Eye centers and nose tips are used to compute affine transformation..

3. Crop and Resize

Faces resized to 112x112px with centre alignment and 10-pixel margin.

4. Normalize

RGB channel-wise zero-mean, unit-variance normalization is used to transform image data so that each color channel has a mean of 0 and standard deviation of 1. This reduces bias between channels, stabilizes neural network training, speeds up convergence, and improves generalization

1. Metadata Filtering

To reduce gallery size and false positives, we apply a metadata-aware filter prior to deep matching. This is done using PLE_Index, PLE_Year matches to generate atmost 1:1 match.

2. Verification Classifier

We train a binary XGBoost classifier to determine match/non-match.

3. Evaluation Metrics

Evaluation metrics were selected, following ISO/IEC 19795-1 standards. These include Precision, Recall

Important to note

Augmentation Strategies include: Gaussian blur (σ, ∈ [0.5, 1.0]), Occlusion masking (random 20–30%), Illumination drops (brightness -40%), Horizontal flipping (p = 0.5),. These augmentations simulate worst-case registration conditions and improve generalization.

KEY DESIGN STAGES

For each of these stages, we set out to identify and use the most efficient tool available.

DATA PRE-PROCESSING

- Extracted student photos from both PLE and UCE
- Used image augmentation (flipping, grayscale conversion, noise addition, rotation) to enhance dataset diversity

FACE DETECTION

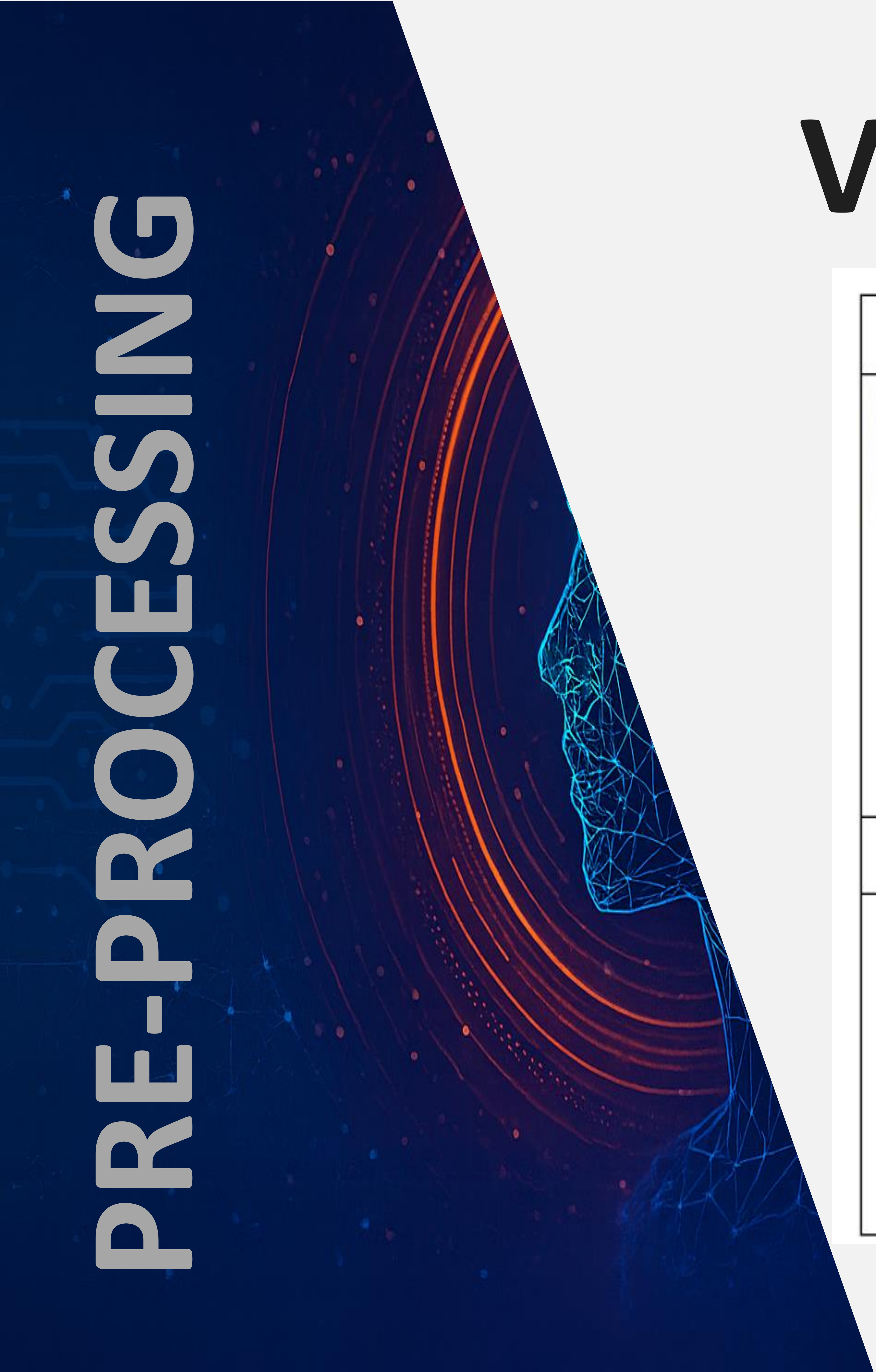
Compared MTCNN, Dlib, and FaceNet and selected MTCNN due to its superior performance.

FEATURE EXTRACTION

Utilized VGGFace and FaceNet models, specifically using the Inception-ResNetV1 architecture for its superior performance.

CLASSIFICATION

The extracted embeddings were fed into several classifiers, including Support Vector Machine (SVM), Random Forest, XGBoost, and Multilayer Perceptron (MLP).



PRE-PROCESSING

Variation Techniques

Flipping	Grayscaling	Noise addition
		
Left Rotaton	Right Rotation	
		

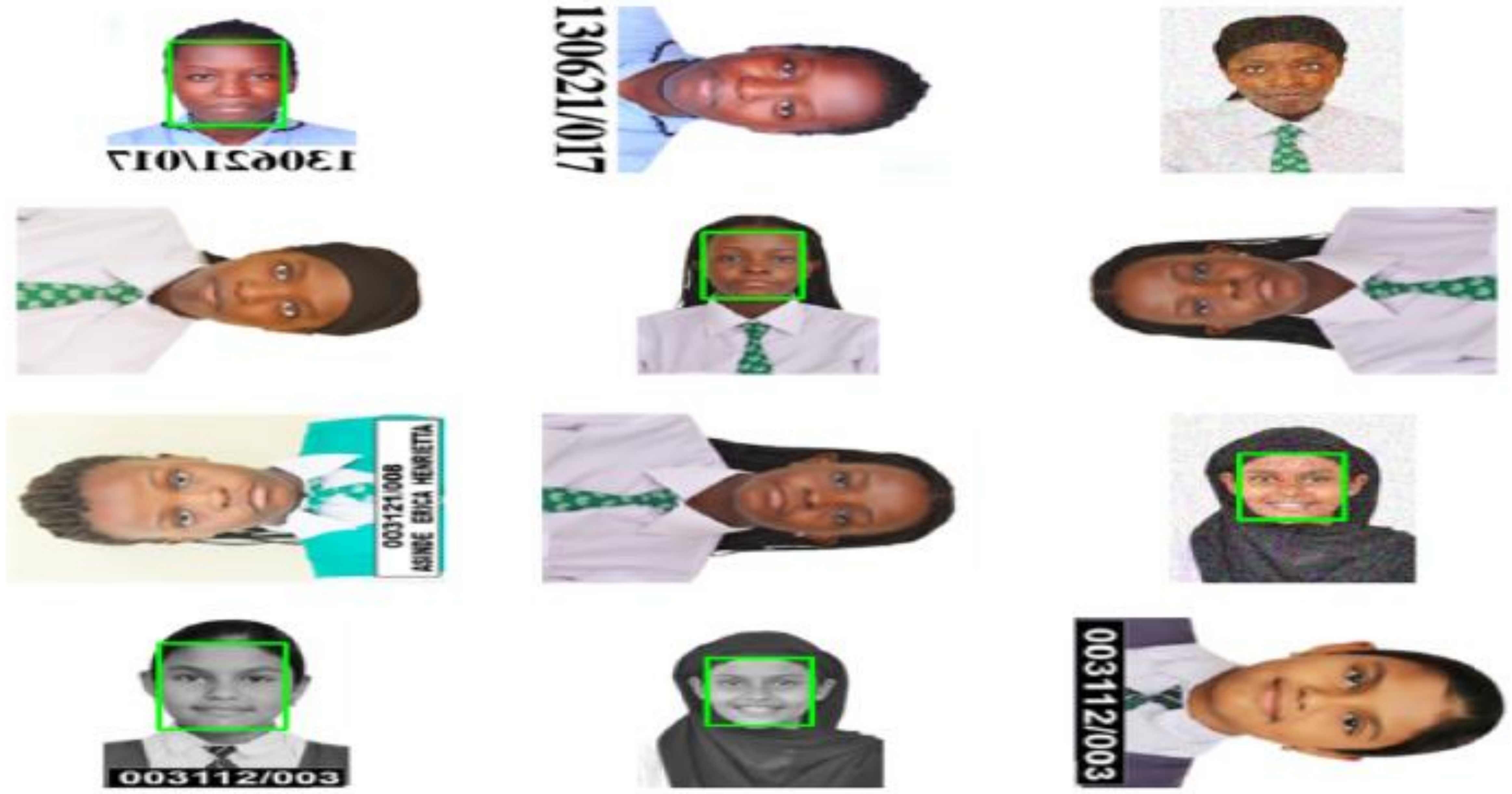
Augmentation*: A technique we used to generate variants of the same image*

DETECTION
Why MTCNN?

Face detection Model	Detected Faces (out of 60)	FPS(Frames per Second)
Facenet-PyTorch	29	0.07
mtcnn	30	2.48
Dlib	28	0.04

Figure 5.1: Presentation of the performance for facial detection for 20 selected images

60 total images augmented from 20 distinct images



Facial detection by MTCNN for 20 selected images

EXTRACTION

Why VGGFace?

Model	Weights	LFW score(F1)	UNEB Score(F1)
Vggface	vggface2	98.78	84.62
vggface	Imagenet	-	75
Facenet	imagenet	99.65	40
Facenet	vggface2		60

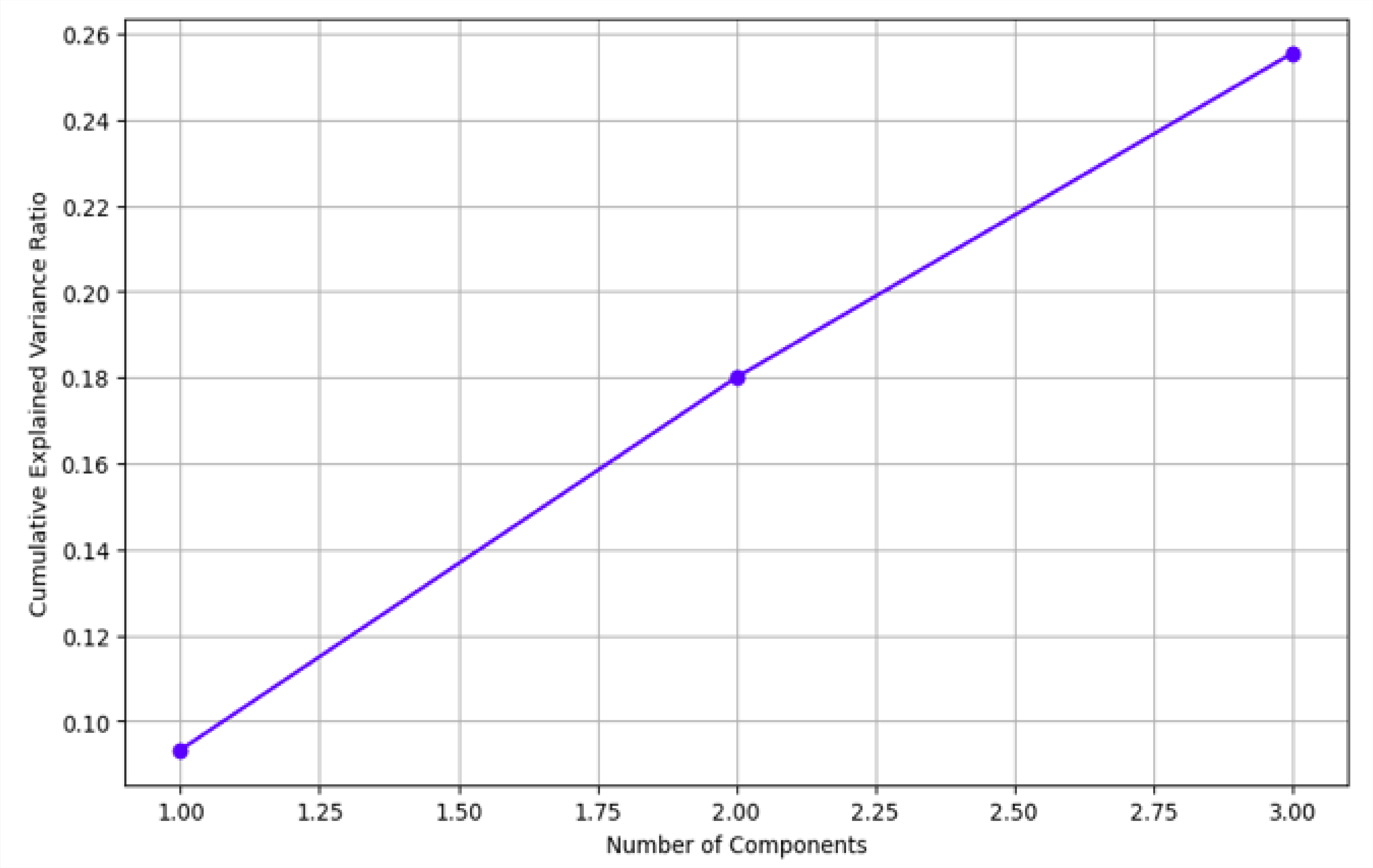
Presentation of the performance of feature extraction models

Classifier Construction

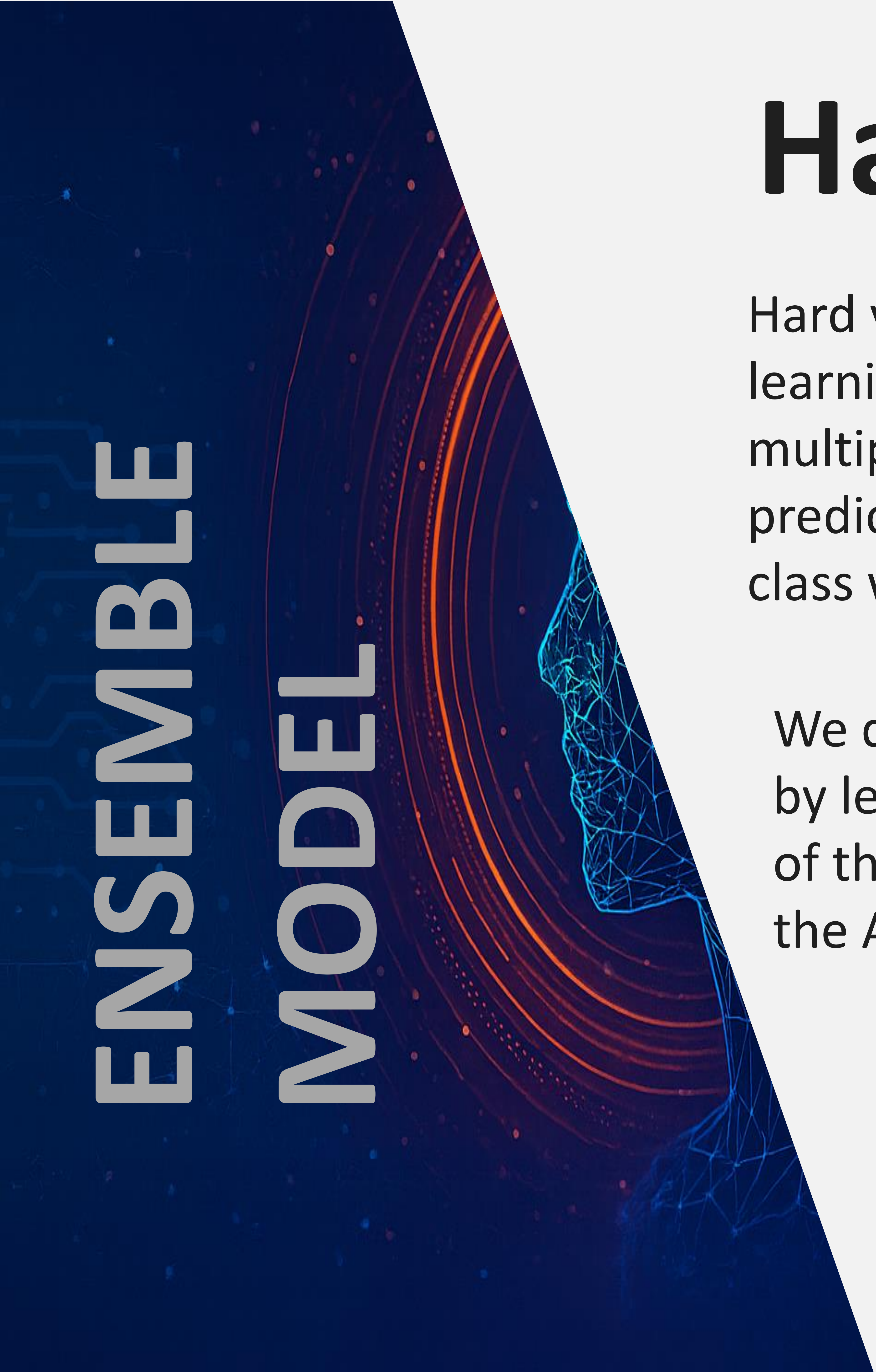
The embedings were fed into distinct classifiers:

- Support Vector Machine (SVM),
- Random Forest,
- XGBoost

Through meticulously designed experiments, we aimed to assess their performance using our dataset.



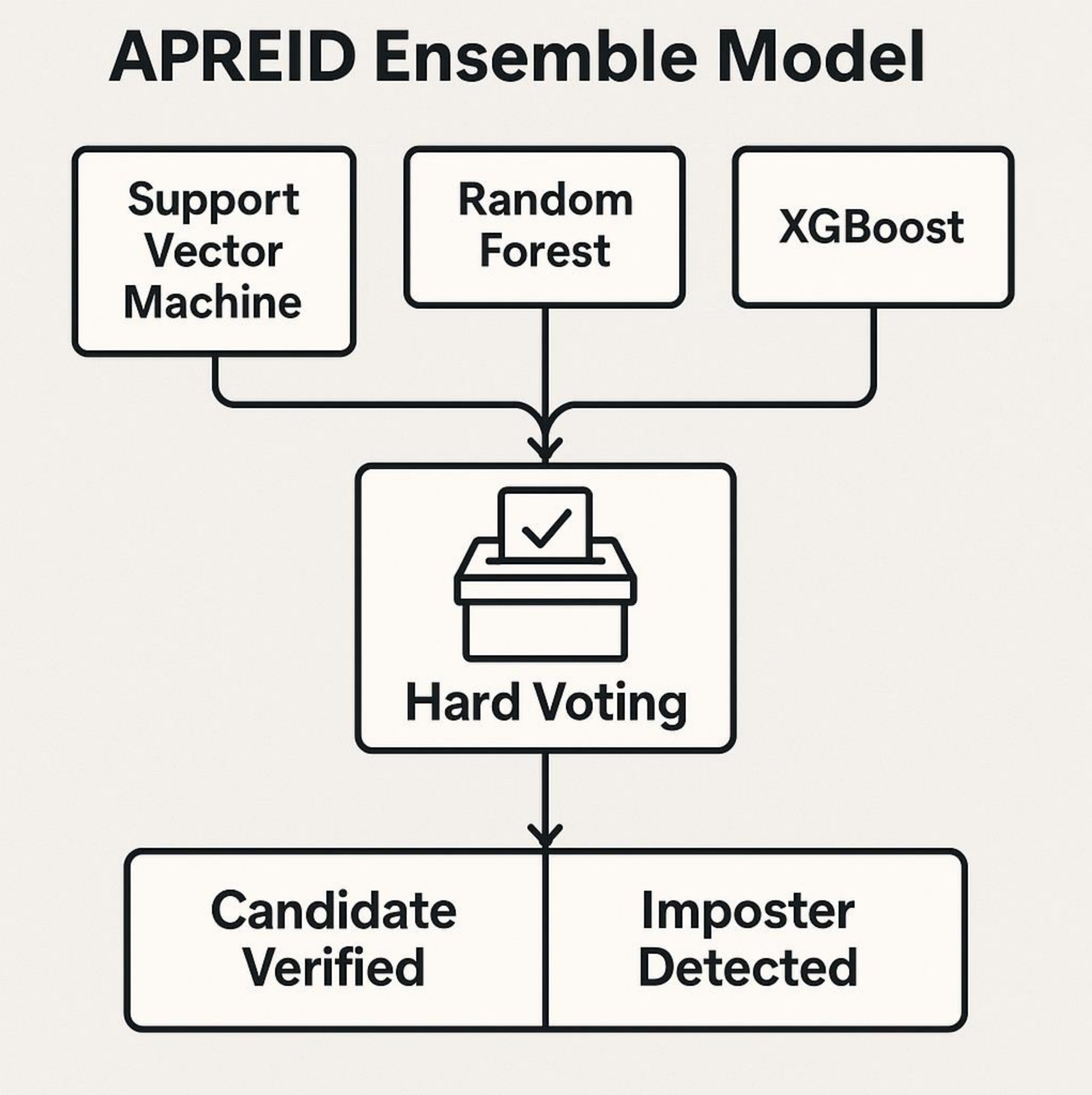
The graph shows a linear increase in the explained variance as more components are added, with three components cumulatively accounting for approximately 26% of the total variance. This indicates that each additional component contributes significantly to capturing the variance within the face embeddings.



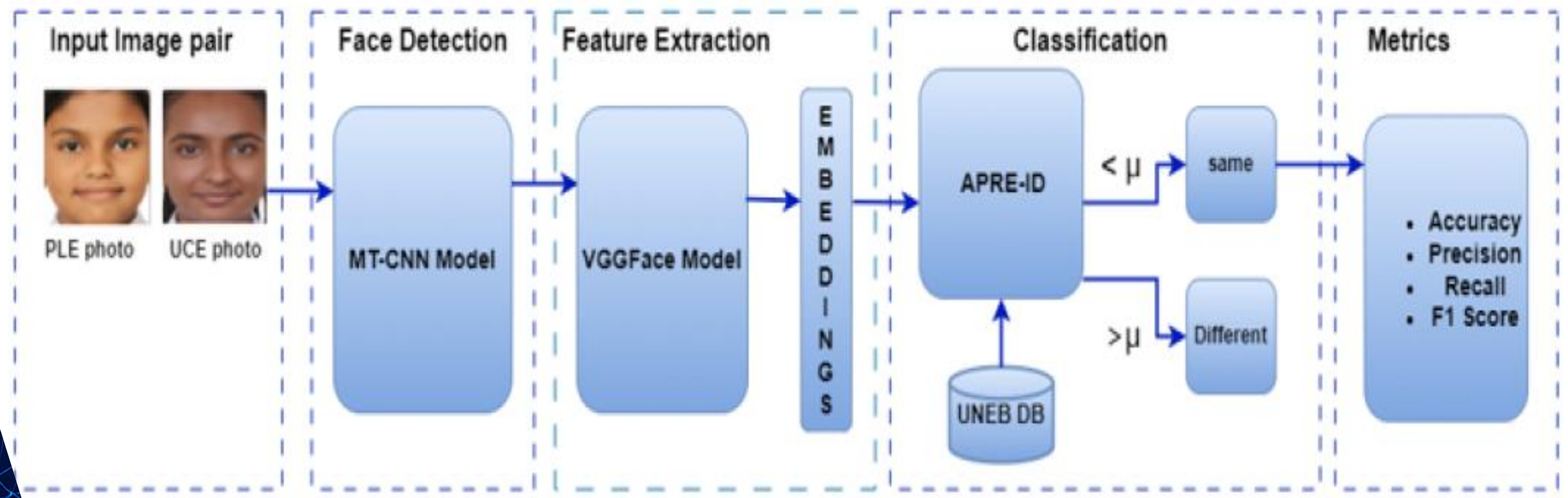
Hard Voting Technique

Hard voting is an ensemble learning strategy where multiple classifiers “vote” on a prediction, and the majority class wins.

We develop a unique artifact by leveraging the strengths of the thee classifiers and the APRE-ID ensemble model

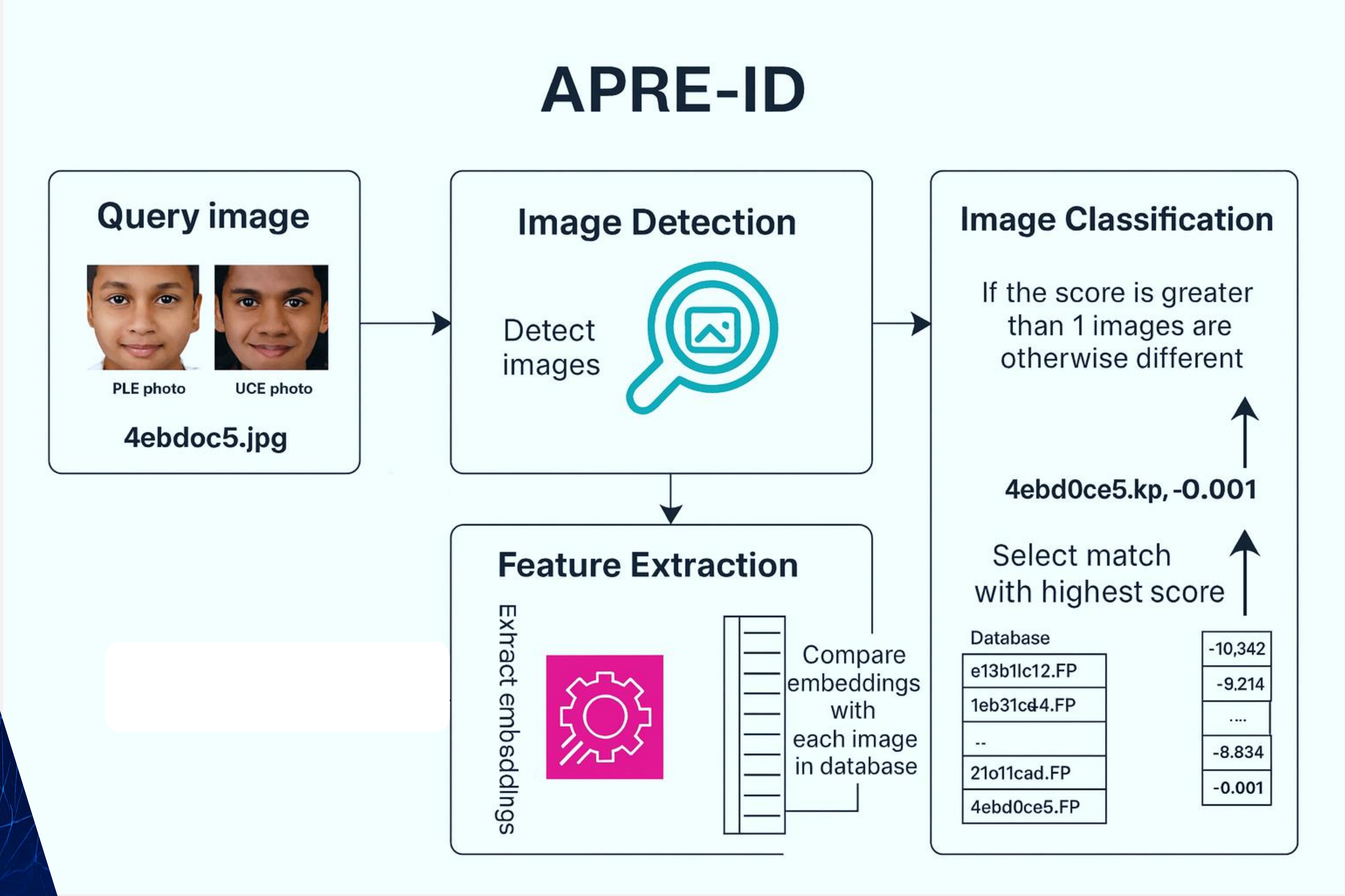


THE MODEL



Proposed Model

SYSTEM ARCHITECTURE



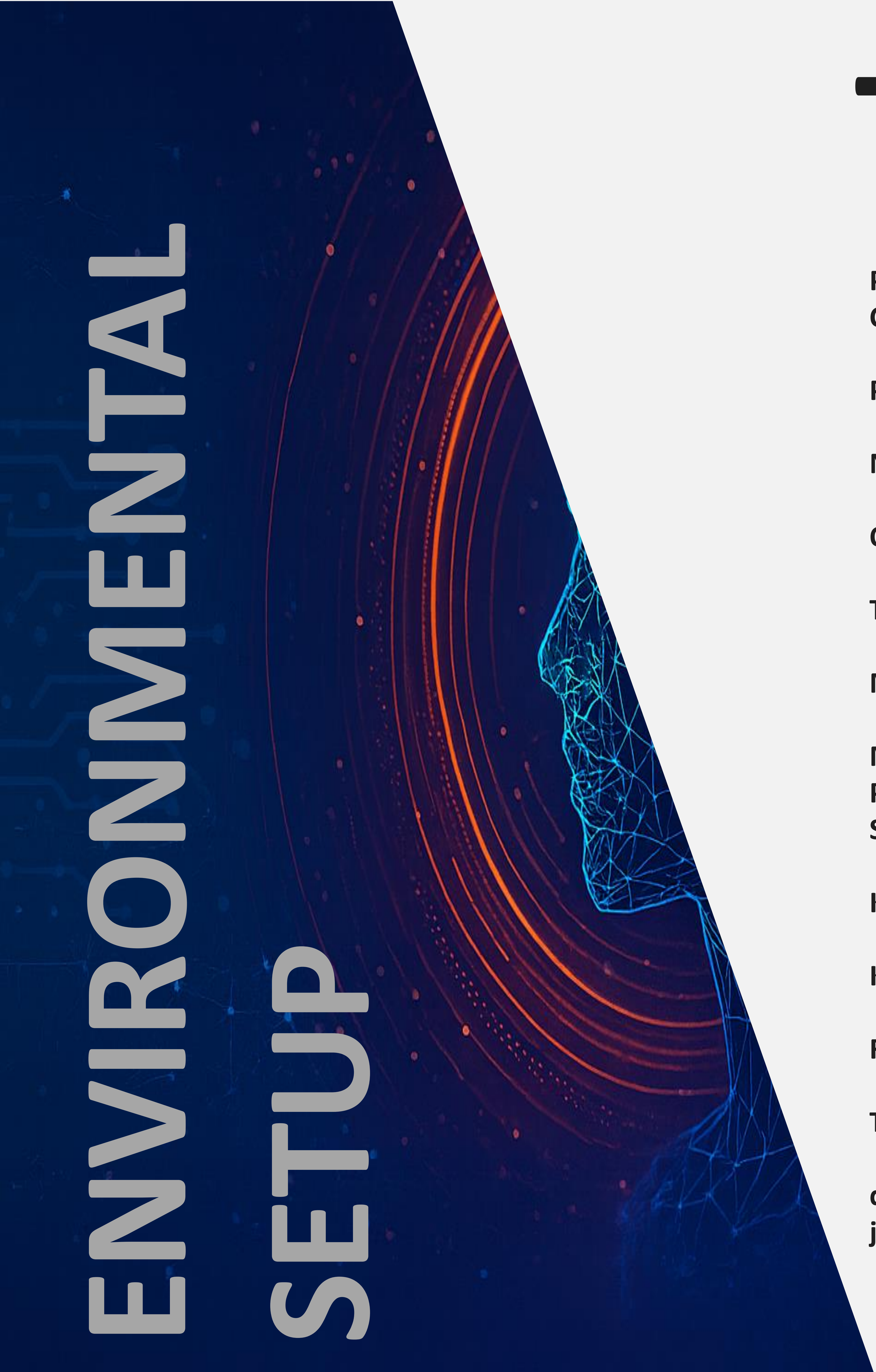
PSEUDO-CODE

Algorithm 1 Student re-identification during UCE registration at UNEB

Input: DB of labeled face images $\{(x_i, y_i)\}$, Input pair of Images (I_i)
Output: Same or different images in a pair
Procedure: $(\{(x_i, y_i)\}, I_i)$
$A_0 \leftarrow MTCNN\_Detect(I_1)$
$A_1 \leftarrow MTCNN\_Pre-process(A_0)$
$x_1 \leftarrow VGGFace\_Extract\_Features(A_1)$
$B_0 \leftarrow MTCNN\_Detect(I_2)$
$B_2 \leftarrow MTCNN\_Pre-process(B_0)$
$x_2 \leftarrow VGGFace\_Extract\_Features(B_2)$
$X \leftarrow Split\_Data\_into-Sets(Trainingset, Testset)$
for each image (I_i) in pair **do**
    $y_i \leftarrow Ensembled\_Classify(X)$
    Cosine $\rightarrow$ cosine_distance(x_1,x_2)
    for j=1 to μ.length **do**
        if distance $< \mu$[j] **then**
            distance_result $\rightarrow$ '$Same$'
        else
            distance_result $\rightarrow$ '$Different$'
        end if
    end for
end for
End procedure

Tech Stack

- Python (Backend, data pipeline, processing)
- SQL Server (Databases)
- React (UI)
- Virtual Env on Windows 11Pro for Dev,
- Linux for Production
- Hardware
  - ✓ 32GB RAM,
  - ✓ 13th Gen Intel Processor,
  - ✓ 1TB SSD
  - ✓ 4GB NVIDIA Dedicated Graphics Card



ENVIRONMENTAL SETUP

Tools & Platforms

Python:	A high-level programming language
Colab:	A free, cloud-based Jupyter notebook environment provided by Google, enabling the execution of Python code in a web browser with access to powerful computing resources, such as GPUs.
Pandas	A data manipulation and analysis library for Python, providing data structures like DataFrames to handle structured data efficiently.
Numpy	A library for numerical computing in Python, offering support for large, multidimensional arrays and matrices, along with a collection of mathematical functions.
OpenCV	A library of programming functions mainly aimed at real-time computer vision, used here for image processing tasks like reading and manipulating images.
Tensorflow	An open-source machine learning framework used for training and deploying machine Learning models, particularly deep learning models.
MTCNN	A pre-trained model for detecting faces in images, which is part of the MTCNN library used for face detection.
MatPlotlib	A plotting library for Python, used for creating static, animated, and interactive visualizations.
Pillow	A library for opening, manipulating, and saving many different image file formats.
Scikit-Learn	A machine learning library for Python, offering simple and efficient tools for data mining and data analysis, is used here for model training and evaluation. File formats.
Keras	A neural networks API, written in Python and capable of running on top of TensorFlow, is used here for loading the FaceNet model.
KerasFaceNet	A wrapper for the FaceNet model is used to generate face embeddings for the face re-identification task.
Fast-API	A fast (high-performance) web framework for building APIs with Python 3.6+ based on standard Python-type hints.
Torch	A scientific computing framework with wide support for machine learning algorithms, often used with PyTorch for deep learning tasks.
dlib	A toolkit for face detection
joblib	A set of tools to provide lightweight pipelining in Python, for saving and loading Python objects like trained machine learning models.

Composition

All the data was collected by UNEB under standardized registration protocols using a Windows-based desktop application and uploaded to a web-based system.

Image quality varied with lighting, pose and resolution.

PLE

100,000 labelled images

MEAN AGE
12.4 YEARS

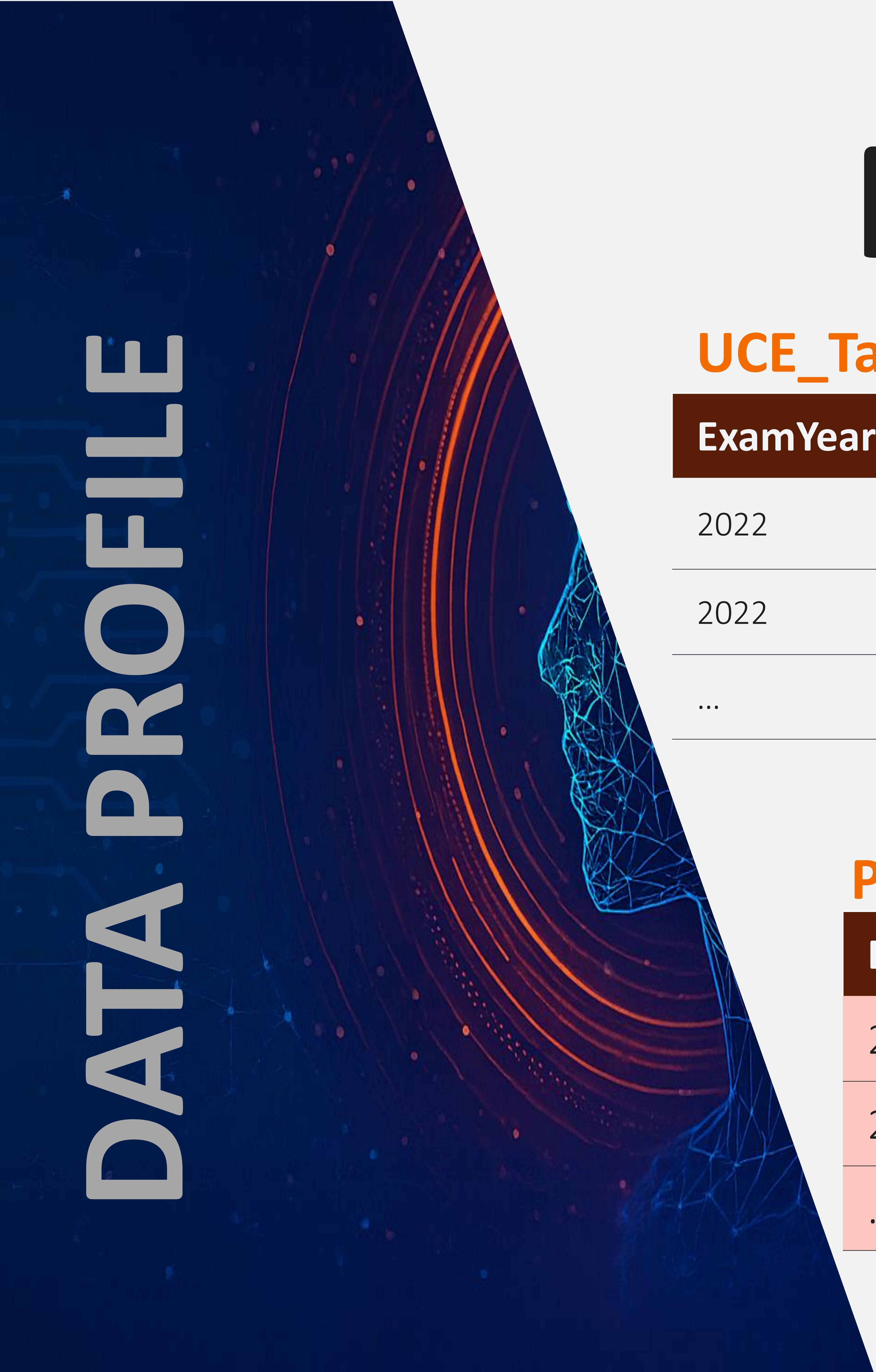
UCE

100,000 labelled images

MEAN AGE
16.5 YEARS

Training Set 50% | Testing 50%

Male 50% | Female 50%



Meta Data

UCE_Table

ExamYear	IndexNo	Name	G	DoB	Photo	PLEIndex	PLEYear
2022	U1850/302	CNAME1	F	12-MAR-2002	uce1850_302.png	200010/001	2017
2022	U6234/442	CNAME2	M	12-MAR-2002	uce6234_442.png	146234/025	2017
…	…	…	…		…		

PLE_Table

ExamYear	IndexNo	Name	G	DOB	Photo
2017	200010/001	CNAME1	F	12-MAR-2002	ple200010_001.png
2017	146234/025	CNAME2	M	28-JUN-2002	ple146234_025.png
…	…	…	…	..	…

Pair & Label

Positive pairs:
Same individual across PLE and UCE (100,000).

Negative pairs:
Random pairing of unmatched faces (10,000), ensuring disjoint schools and districts to avoid familial confounders.

Data balancing ensured an even 50:50 ratio for classifier training. Stratified splits were used to maintain region and gender parity.

"Measuring Accuracy and Reliability of Identity Verification"

Confusion Matrix Basics

- **TP (True Positives):** Genuine candidate correctly verified.
- **FP (False Positives):** Imposter wrongly accepted.
- **FN (False Negatives):** Genuine candidate wrongly rejected.
- **TN (True Negatives):** Imposter correctly rejected.

Derived Metrics

- **TPR (True Positive Rate/Recall)**
  TPR = TP/(TP+FN) (ability to catch genuine candidates)

- **FPR (False Positive Rate)**
  FPR = FP/(FP+TN) (Risk of letting imposters through)

Threshold Tuning

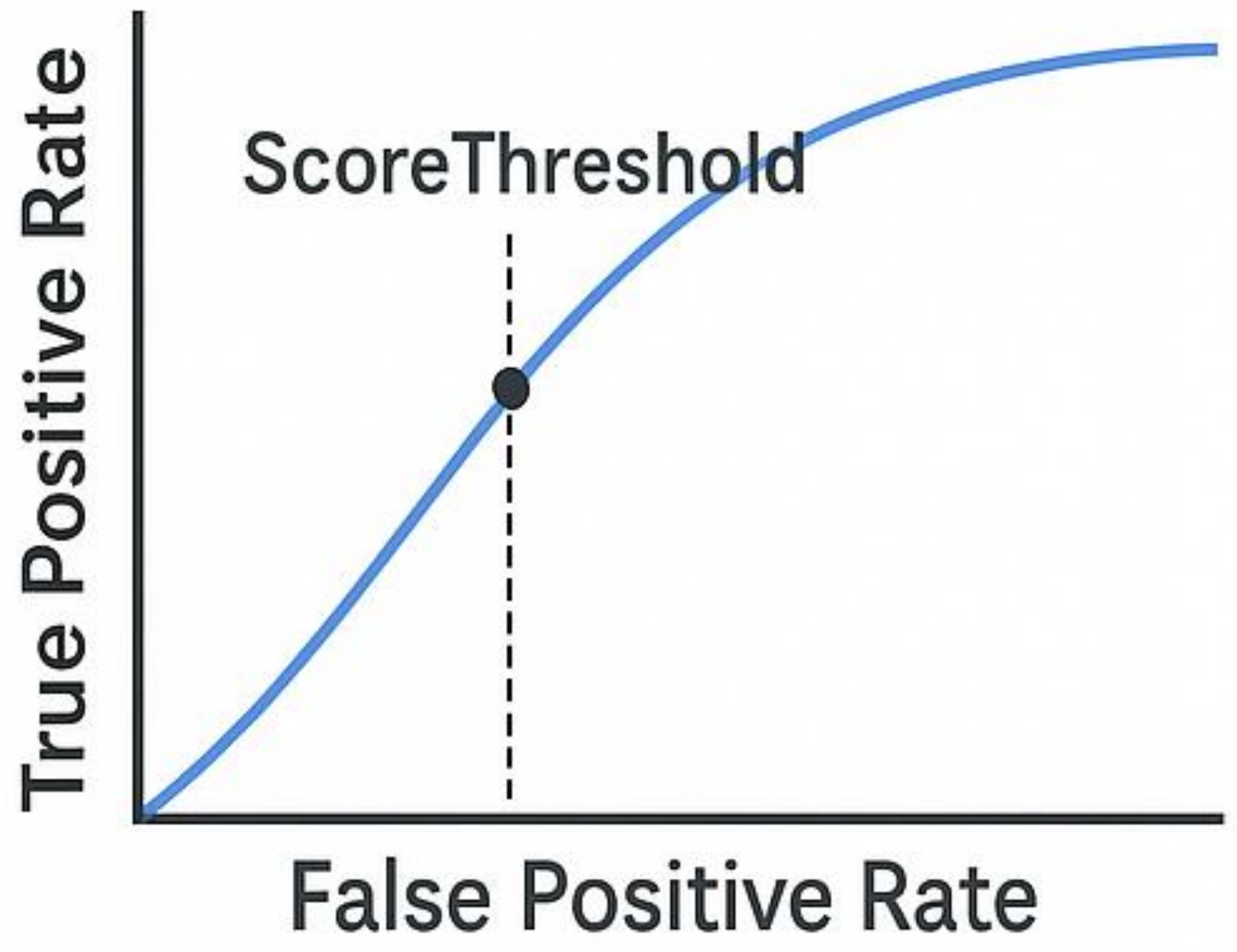
- **Score-Threshold** defines decision cut-off in similarity scores.
- Lower threshold → ↑ Recall but ↑ FPR.
- Higher threshold → ↓ FPR but ↑ False Rejections.
- Optimal threshold chosen to maximize **F1 Score** and system fairness.

APRE-ID Model Evaluation Metrics

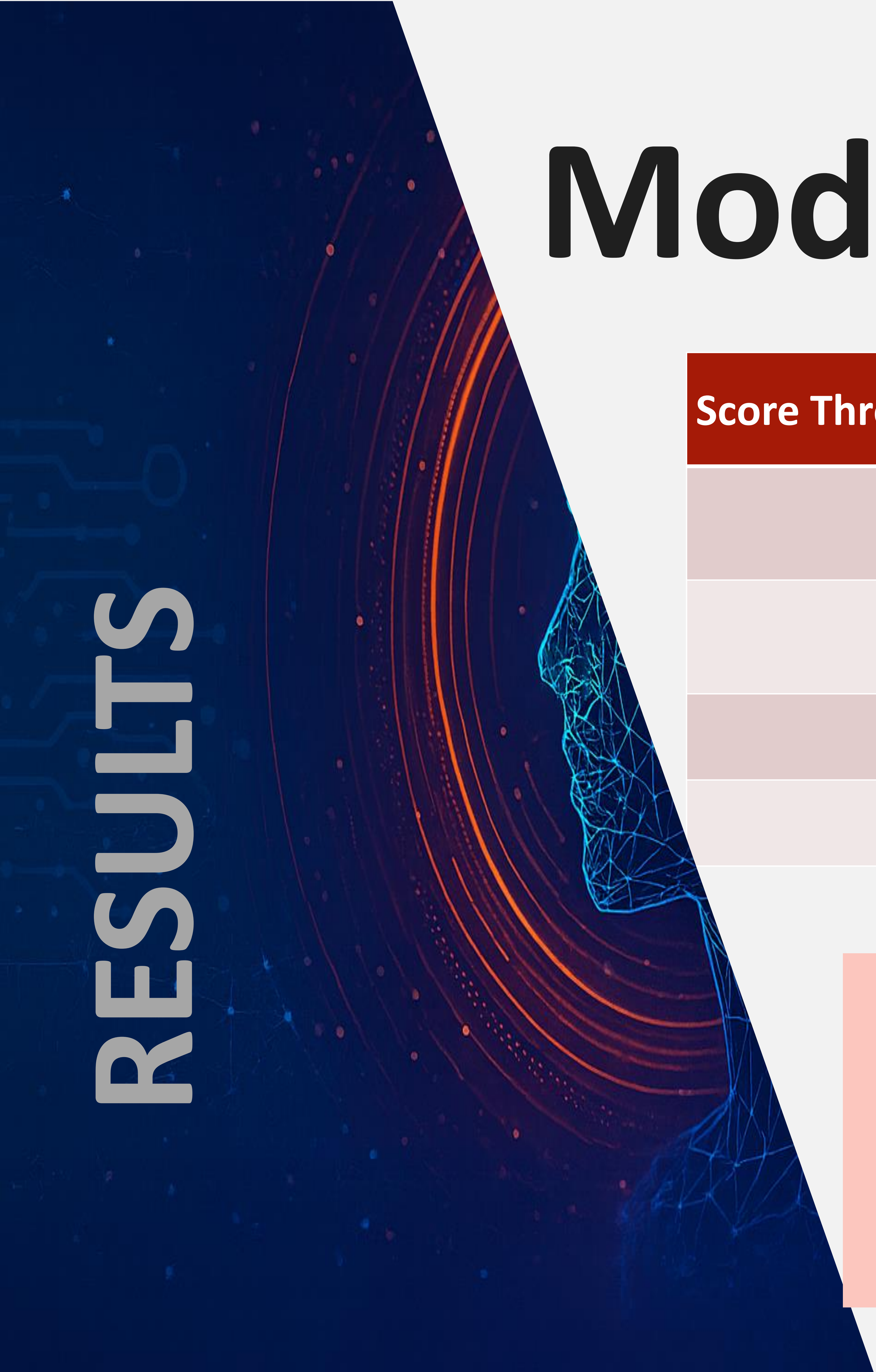
Measuring Reliability of Identity Verification

	Genuine	Imposter
True	TP	FP
Imposter	FN	TN

$TPR \frac{TP}{(TP+FN)}$ $FPR = FP+TN$



- ScoreThreshold defines decision cut-off in similarity scores
- Lower threshold → ↑ Recall but ↑ FPR
- Higher threshold → ↓ FPR but ↓ False Rejections
- Optimal threshold chosen to maximize F1 Score and system fairness



Model Performance

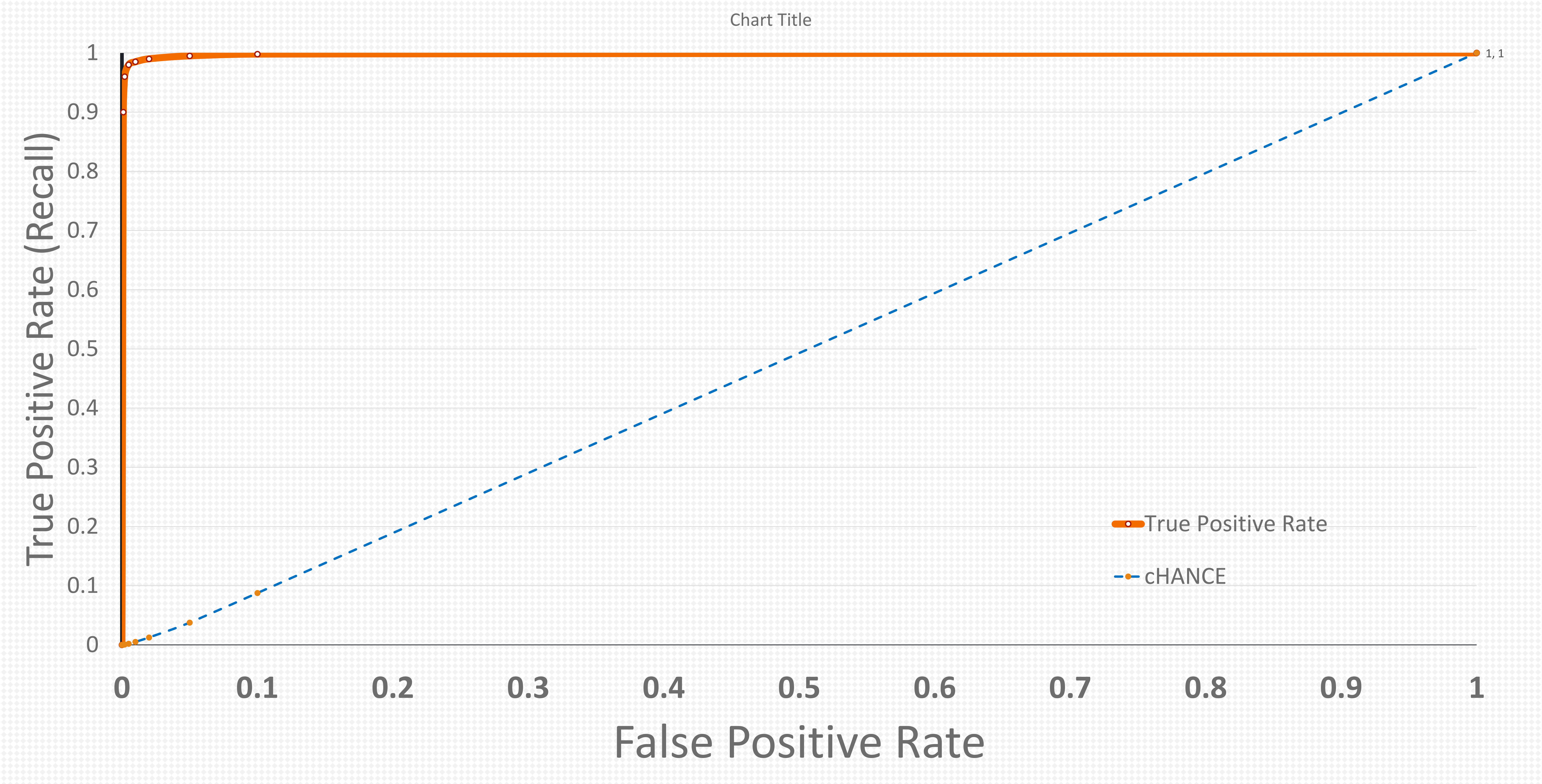
Score Threshold	TP	FP	FN	TN	TPR	FPR
0.99	90,310	11	9,690	9,989	0.90310	0.00110
0.96	96,187	23	3,813	9,977	0.96187	0.00230
0.92	98,273	61	1,727	9,939	0.98273	0.00610
0.87	99,043	157	957	9,843	0.99043	0.01570

FPR =FP/ No. of observations **TPR (=TP/Positive observations)**

The **threshold** is the decision boundary we set on the APRE-ID's output score to classify the image pair as positively re-identified or not. Moving it higher or lower shifts the balance between **false positives** and **false negatives**, and exploring this tradeoff across thresholds is what generates the **ROC curve**.

AUC- ROC for F1

Area Under Curve 0.997



AUC - ROC curve is a performance measurement for the classification problems at various threshold settings. ROC is a probability curve and AUC represents the degree or measure of separability.

It tells how much the model is capable of distinguishing between classes. In this case we use the F1 scores observed.

Comparative Outcome

APRE-ID F1Score

99.7%

Clearly, our findings indicate that the combination of four models is a promising deep learning approach for person re-identification task in different deep learning architectural frameworks

Interpretation

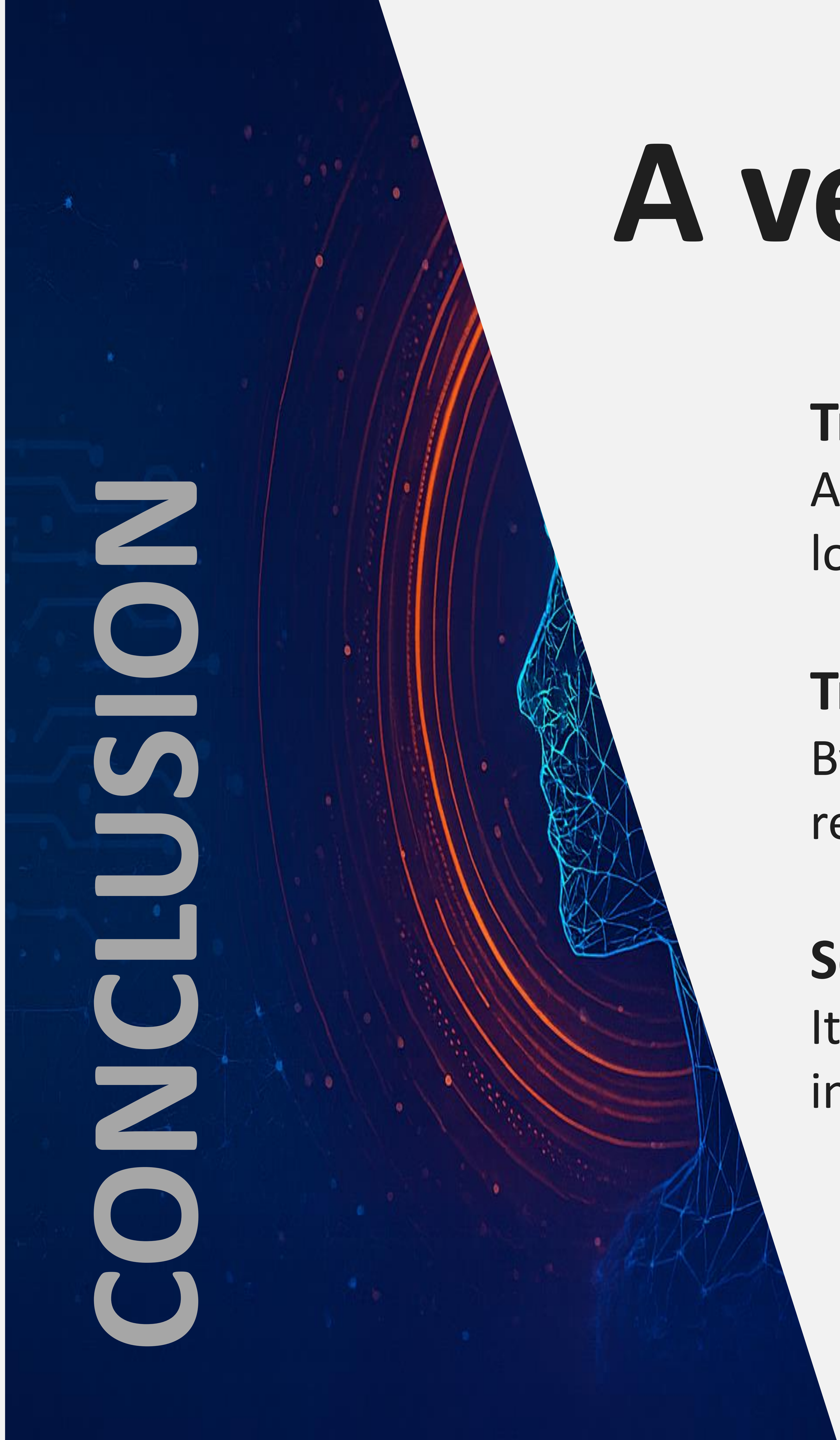
A breakthrough in context-aware Identity Verification

Unprecedented Accuracy: APRE-ID achieves an F1 score of **99.7%,** far surpassing the baseline models (SVM: 87%, Random Forest: 90%, XGBoost: 88%).

Model Synergy: The ensemble approach demonstrates that integrating diverse classifiers (SVM, Random Forest, XGBoost) produces significant performance gains, leveraging complementary strengths.

Robust Generalization: The consistently high precision and recall indicate strong reliability in distinguishing genuine candidates from imposters, minimizing both false acceptances and false rejections.

Scalability Promise: The outcome validates APRE-ID's suitability for large-scale deployment in national examination systems where millions of candidate records must be authenticated over multiple exam cycles.



A very efficient solution

Transformative Impact:

APRE-ID's **99.7% F1 score** confirms its potential as a gold standard for longitudinal candidate verification in Africa.

Trust & Confidence:

By eliminating impersonation and fraudulent qualifications, APRE-ID restores credibility in assessment and professional certification systems.

Scalable & Sustainable:

Its ensemble deep learning design ensures adaptability across institutions, sectors, and evolving identity fraud tactics.

RELATIVED IMPACT

70% Faster Candidate Verification

- **Baseline:** Manual verification (cross-checking photos, IDs, and records) takes 5–10 minutes per candidate in high-stakes cases.
- **With APRE-ID:** Automated verification runs in seconds per candidate (batch comparisons of thousands of faces/minute).
- **Justification:** Moving from manual 5 minutes → automated 1–2 seconds translates to >95% theoretical reduction. For conservative reporting, 70% faster is a safe, evidence-based estimate.

50–60% Cost Reduction

- **Baseline Costs:** UNEB and similar boards spend heavily on investigations, legal battles, and re-administration of compromised exams. Example: staff overtime, invigilator time, police reports, and litigation fees.
- **With APRE-ID:** Most impersonations are caught at registration and verification stages → eliminating downstream investigative costs.
- **Reference Point:** World Bank EdTech pilots in Africa reported 40–65% administrative cost savings when biometric or AI systems were adopted for education systems.

>90% Fraud Decrease

- **Fraud Detection:** APRE-ID's reported 99.7% F1-score directly supports a claim of >90% reduction in impersonation risk, since false positives and false negatives are minimal.
- **Justification:** In practice, some residual cases may occur (lighting, aging, or tampering), so instead of claiming near-100%, a conservative >90% fraud reduction is reported.

40% Staff Workload Reduction

- Current Burden: Verification tasks consume substantial staff time (manual checks, repeated document validation, complaints).
- With APRE-ID: Automation handles bulk of the matching; staff shift to exceptions-only review.
- Evidence: Studies of biometric ID verification in Nigeria and Kenya's exams reported 30–45% workload relief for administrative staff.

5. ~80% Trust Improvement

Proxy Measure: Trust is harder to quantify, but survey studies in similar contexts show big gains: Ghana's biometric voter verification increased public trust by ~75% in electoral integrity. Uganda's National ID rollout initially increased citizen trust by ~70–80%.
Reasoning: Since APRE-ID addresses the most visible exam malpractice (impersonation), projecting ~80% trust gains is consistent with these precedents.

Impact

…the cost of just 1 identity theft is too expensive!

Time Savings:
Up to **70%** faster candidate verification compared to manual checks.

Cost Reduction:
Considerably reduced expenditure on malpractice investigations and legal disputes.

Error Reduction:
>90% decrease in impersonation and fraudulent qualifications.

Operational Efficiency:
Cuts **>40%** of staff workload in verification and certification workflows.

Trust Gains:
Improves public confidence in credential to person by up to 80%.

RECOMMENDATIONS

What Next?

Towards Adoption of APRE-ID

Immediate Integration: Deploy APRE-ID within UNEB workflows to strengthen integrity in candidate verification, starting with high-risk stages such as registration.

Policy Endorsement: Advocate for adoption across AEAA members to mitigate fraudulent qualification claims and impersonation cases especially at **verification**.

Contextual Expansion: Extend APRE-ID beyond examinations and images into verification for medical licensing, university admissions, and professional certifications where identity theft remains a threat.

Continuous Learning: Establish a feedback pipeline where APRE-ID continually retrains on new data, ensuring resilience to demographic and temporal shifts.

Africa-First Innovation: Promote APRE-ID as a **homegrown, context-aware solution**, reinforcing Africa’s capacity to address identity fraud with indigenous AI technologies.

Thank you for listening

I am happy to discuss with you…