National College of Ireland

Deepfake Resistant Framework

Real Time Biometrics Identity Verification in Remote Banking

MSc Research Project

MSCCYBE_JAN24

Shanthini Chandrasagaran

Student ID: x22197460

School of Computing

National College of Ireland

Supervisor: Ross Spellman

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Shanthini Chandrasagaran
Student ID:	X22197460
Programme	MSCCYBE_JAN024 **Year:** 2024
Module:	THESIS
Supervisor:	Ross Spellman
Submission Due Date:	11th December 2025
Project Title:	Deepfake Resistant Framework in Real Time Biometrics Identity Verification in Remote Banking
Word Count:	6331 **Page Count** 26 (including Cover sheet, content page, appendix)

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Shanthini Chandrasagaran

Date: 09122025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

DEEPFAKE-RESISTANT FRAMEWORK FOR REAL-TIME BIOMETRIC IDENTITY VERIFICATION IN REMOTE BANKING

Shanthini Chandrasagaran

Student ID: x22197460

Abstract:

The progressive sophistication of deep-fake technologies, powered by breakthroughs in Generative Adversarial Networks (GANs) and deep learning, presents an opportunity to elevate the resilience of biometric identity verification systems across the banking and FinTech landscape. These advanced, synthetic assets—including facial, audio, and behavioral signatures—demand a higher standard of security than conventional methods currently offer.

My research is dedicated to designing and successfully implementing a leading-edge, multi-modal verification framework. This system is specifically designed for real-time, robust biometric identification within critical financial Know Your Customer (KYC) processes. The proposed defense leverages the power of four integrated layers—facial recognition, voice authentication, behavioral biometrics, and active liveness detection—to achieve optimal resilience and adaptability.

To validate its strength, I will generate sophisticated synthetic challenges using advanced deepfake tools. Detection will be handled by fine-tuned Convolutional Neural Networks (CNNs). A key contribution is the development of a vulnerability scoring dashboard, providing financial institutions with the crucial ability to quantitatively measure and proactively visualize their exposure status. Implemented in Python on Google Colab for real-time performance, this study delivers a powerful and essential defense system that secures the digital identity lifecycle against evolving AI-driven methodologies.

Table of Contents

Chapter 1: Introduction

1.1 Introduction and Problem Statement

The rapid digitalization of the financial industry has led to faster adoption of biometric identity verification systems for customers to perform identification, verification, and access services. These technologies, mostly using facial recognition and voice authentication, have also become an everyday part of convenience and security in remote banking and fintech settings (Monisha *et al.* 2024). The advent of generative adversarial networks (GANs) and advanced neural architecture, however, has raised a potentially disastrous security issue amid the fast-evolving field of deepfakes.

Deepfakes can be highly realistic, and even minor behavioral clues can be reproduced, allowing an attacker to impersonate a genuine customer with great precision. Such ability jeopardizes the security of remote onboarding operations, erodes confidence in KYC mechanisms, predisposes financial actors to identity fraud and account cloning, and subjects them to regulatory risk (Abdul-Al *et al.* 23024). Most systems do not address newer anti-spoofing techniques but focus on classic presentation attacks, such as videos or printed pictures. Not many have also succeeded in being tested with high-fidelity synthetic attacks generated by modern AI models.

Furthermore, there is a significant gap in the vulnerability assessment of biometric systems in the financial sector. Current alternatives typically lack an adaptive scheme to gauge resilience, visualize weak points, and deliver actionable risk scores. In the absence of such a mechanism, banks will have no systematic way to quantify their exposure to threats posed by deepfakes. They thus will be unable to prioritize defenses against them (Vatchala *et al.* 2025).

In this research, it is shown that the gaps in this domain are bridged by developing a deep-fake-resistant, real-time (made possible by a multi-modal biometric verification tool), scalable model equipped with vulnerability assessment capabilities (Sumalatha *et al.* 2024). This proposed system will maximize system security and robustness and provide financial institutions with a practical way to evaluate risk by combining multiple biometric modalities, integrating deepfake detection, and simulating realistic synthetic attacks.

1.2 Objectives

- To test an approximation to the real world (deepfake identity attacks) on biometric KYC systems using experimental generation tools (e.g., Face Swap, Descript) and other GAN-based synthesizers
- To create a live identity verification framework that combines face recognition, voice authentication, behavioral biometrics, and liveness
- To deploy a vulnerability assessment dashboard to rate biometric verification systems to certain quantifiable levels of risk under different attacking conditions
- To test the accuracy of detection and the robustness of systems against older types of spoofing and newer formats of synthetic identity attacks using AI
- To provide recommendations on security and a deployable prototype that could be deployed in a fintech and remote banking context.

1.3 Research Questions

- What would be the success rate of deepfake-based spoofing attacks on current biometric identity verification systems in the financial sector?
- Is a multi-modal biometric system any better able to detect individuals and resist synthetic identity fraud than a single-modality one?
- How can one measure, render, and score the susceptibility of such systems to AI forged forgeries using a solid methodology?

1.4 Scope and Limitation

Scope:

- The study will be carried out on remote onboarding and login processes that are widespread in banks and fintech
- The system will be made by using publicly available biometric data and attack data, synthesized with an attack dataset (WILLIAM *et al.* 2022)
- Many publicly available tools will be used to simulate deepfake attacks (FaceSwap, DeepFaceLab, and Descript, as well as custom GAN-based models)

1.5 Limitations:

- No actual customer information of banks will be used; testing will be based on synthetic/publicly accessible information
- The prototype could be technically oriented only and will not cover regulatory and compliance requirements like GDPR or PSD2 in their entirety (Kolluri *et al.* 2024)
- Depending on the hardware resources, e.g., when dealing with high-scale deployments, performance results may be different.

Chapter 2: Literature Review

2.1 Introduction to the Literature Review:

The rapid digitalization of the financial services industry has led to a hardening of biometric authentication requirements to ensure secure remote banking. Although biometrics is more secure than conventional types of passwords, the advent of deepfake technologies has opened the door to identity fraud (Wang *et al.* 2023). Recent studies have shown that neural networks can alter visual and audio images with a high degree of realism, casting doubt on the robustness of biometric systems.

2.2 Biometric Authentication in Remote Banking:

Biometric authentication is one such aspect of digital banking that has become a cornerstone of security, as it is convenient and prevents credential theft. Face recognition, voice authentication, and behavioral biometrics are generally used methods to verify a user in remote banking applications (Wang and Chow 2023). Such modalities offer non-authentic, un-Kathy-authenticate services, as well as making passwords and one-time codes less critical.

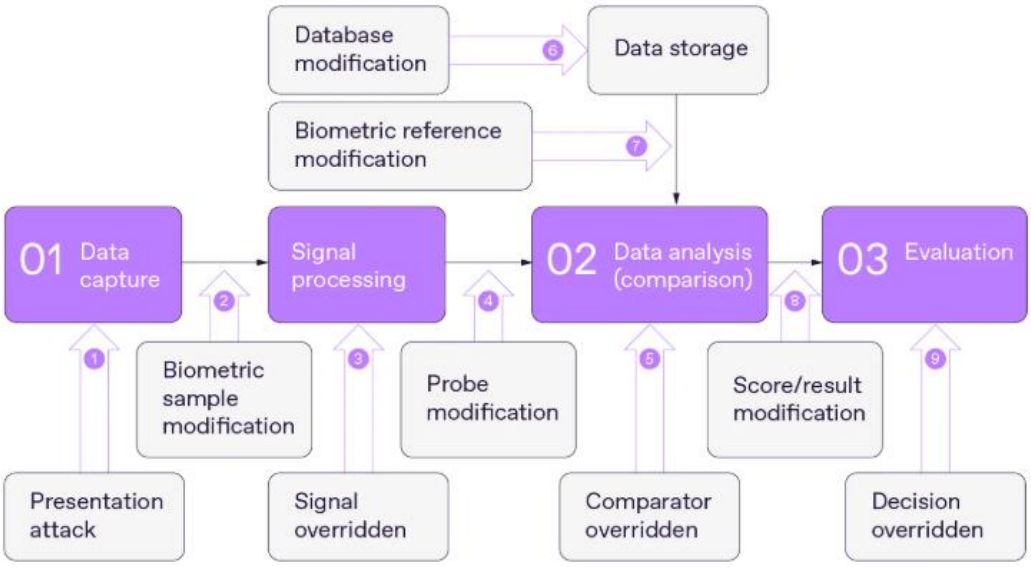

Figure 2.1: Deep-fake Detection with ID Verification Tool

(Source: Wang and Chow 2023)

2.3 Deepfake Technologies and Their Impact on Biometric Systems:

In remote biometric applications, the threat is critical, especially when visual and auditory information are essential for user verification (Zhang *et al.* 2024). Technologies for deep freezing have been significantly enhanced by new convolutional attention mechanisms and feature pooling, enabling greater feature-level spoofing of identities (Wang *et al.,* 2023). A similar trend is the creation of manipulations based on noise that interfere with liveness checks, making synthetic content even more difficult to detect (Wang and Chow, 2023).

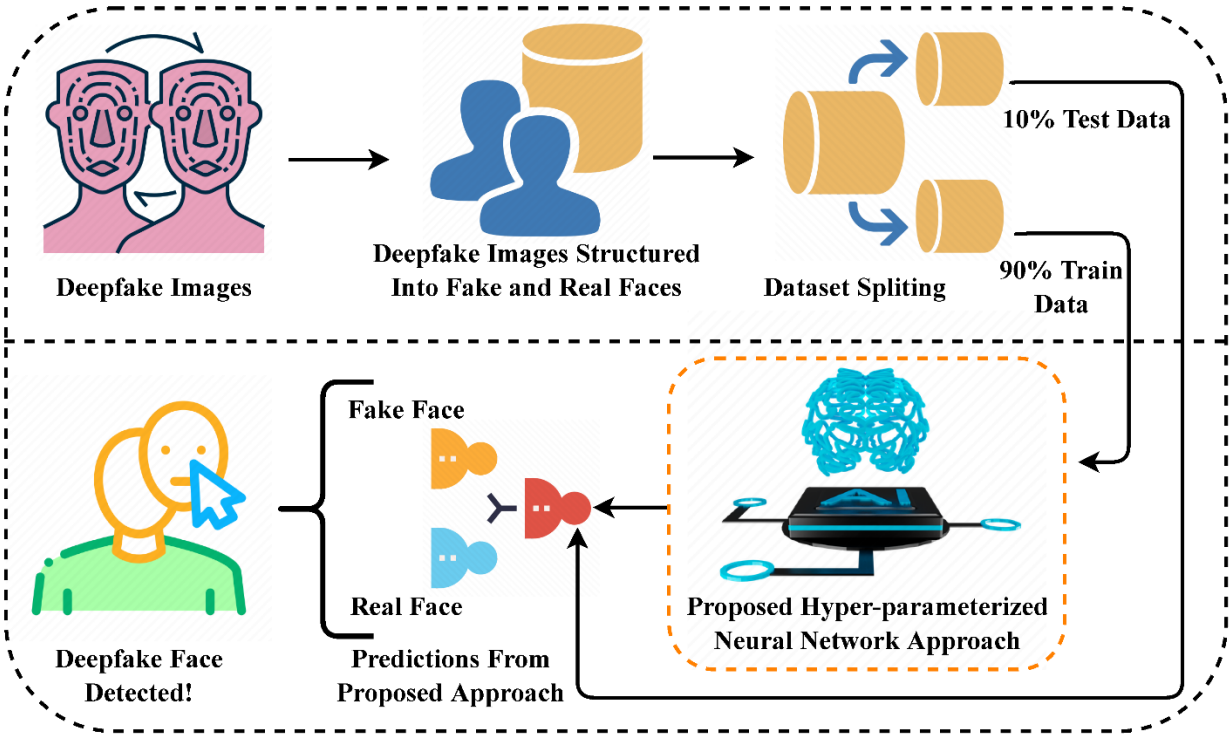

Figure 2.2: Novel Deep-Learning Approach

(Source: Wang *et al.* 2023)

2.4 Deepfake Detection Methods:

A single direction of change is based on the idea that deep convolutional transformer models can learn fine temporal and spatial divergences in manipulated content (Wang *et al.,* 2023). Other works focus on noise-based analysis, in which sensor-level noise patterns that deepfakes fail to match are analyzed by the detection models (Wang and Chow, 2023).

The most crucial issue is the model's generalization in identifying unseen forms of deepfakes. Studies note that implicit identity leakage poses a significant challenge, as detection models can accidentally overfit to specific identities during training (Dong *et al.,* 2023). At that, self-supervised learning has been considered an approach to training models on adversarial examples in the absence of large volumes of labelled data to enhance resistance to new deepfake capabilities (Chen *et al.,* 2022).

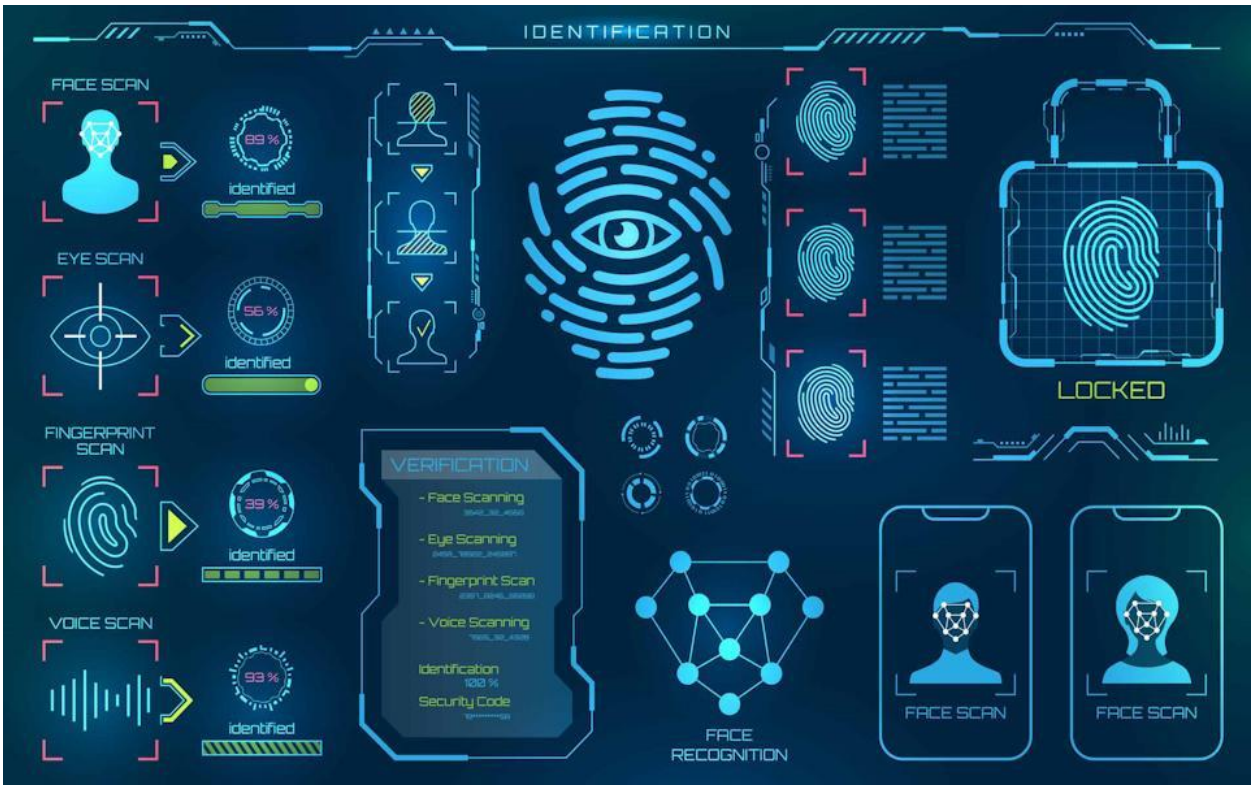

Figure 2.3: Trifecta of Biometric Security in Cloud and Software

(Source: Chen *et al.,* 2022)

Further studies suggest that frequency-domain learning identifies spectral-based abnormalities in frequency components as commonly found in deepfakes (Tan et al., 2024). This method increases generalizability.

In the domain of audio, privacy-preserving deepfake detection models, such as SafeEar, have been proposed; these methods analyze voice manipulations without processing sensitive speech content (Li *et al.,* 2024).

2.5 Multi-Modal Biometric Systems:

The latest research on multimodal fusion has shown greater resistance to environmental noise and intentional attacks. For example, behavioral biometrics, such as keystroke dynamics, mouse movement patterns, and so on, are remarkably immune to deep-fake manipulation because they involve human motor behaviors that are more difficult to synthesis. Such a combination of behavioral data, therefore, adds a security feature to a remote banking application (Dong *et al.* 2023).

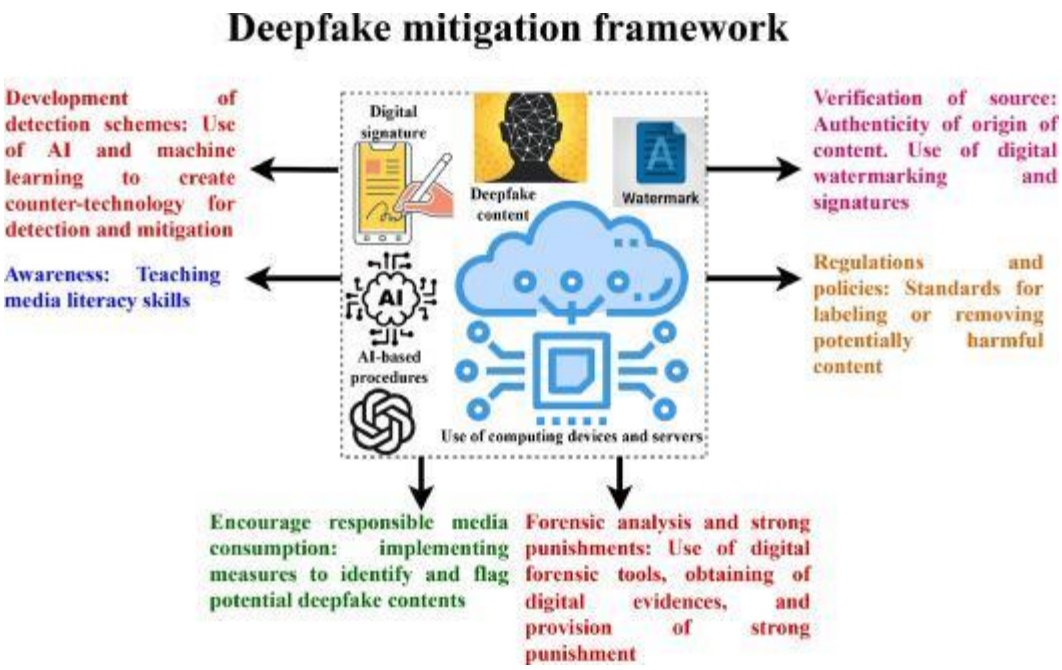

Figure 2.4: Deepfake Mitigation Framework

(Source: Dong *et al.,* 2022)

Nevertheless, the available literature indicates that very few studies have examined multimodal systems explicitly designed for deepfake resistance, highlighting a gap that could be addressed by improving biometric fusion with deepfake detectors through computer vision and audio processing (Chen *et al.* 2022).

2.6 Vulnerability Assessment in Biometric Systems:

Vulnerability assessment is a critical step in understanding the nature of biometric systems' responses to adversarial threats. The traditional evaluation process uses metrics such as the False Acceptance Rate (FAR), True Acceptance Rate (TAR), and Equal Error Rate (EER) to provide numerical measures of system accuracy (Kumar and Sharma 2023). Nevertheless, the emergence of deepfakes has led to new assessment metrics, such as Attack Success Rate (ASR), which refers to the degree of successful penetration of deepfake inputs into authentication systems.

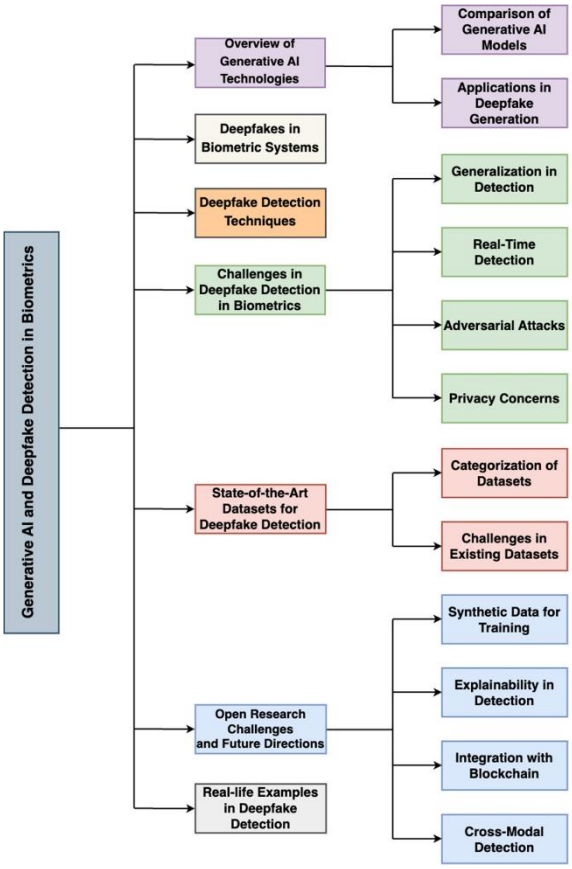

Figure 2.5: Deepfake Detection in Biometric System

(Source: Kumar and Sharma 2023)

2.7 Deepfake-Resistant Systems in Other Fields:

Other fields outside the banking industry have investigated deep-fake-resistant validation systems. GAN-based detection models can be applied on social media to analyze facial data and detect irregularities in audio, thereby helping detect suspicious content (Kumar and Sharma, 2023). Frequency-aware and noise-based methods are increasingly used in cybersecurity and content moderation because they are resistant to all forms of manipulation (Tan *et al.,* 2024).

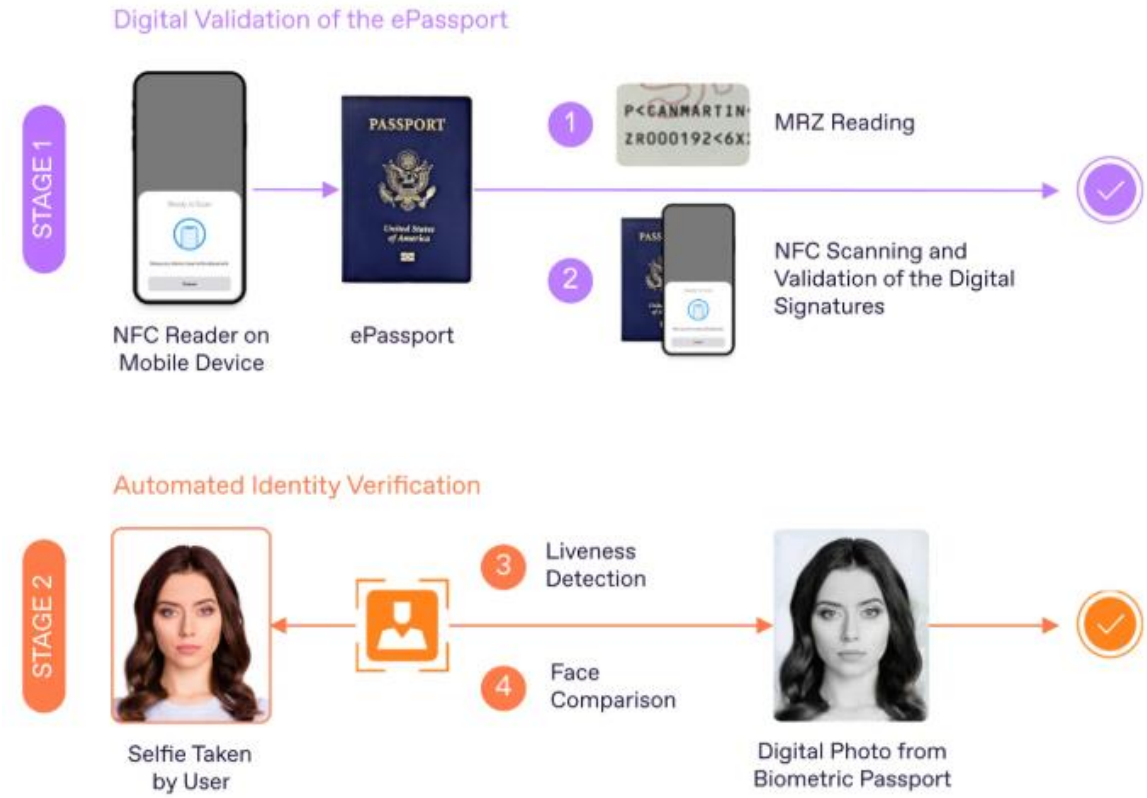

Figure 2.6: Biometrics Application in Banking

(Source: Tan *et al.,* 2024)

Besides, it has offered common-sense reasoning frameworks to detect deepfakes and to focus on semantic inconsistencies that AI-generated content tends to overlook (Zhang *et al.,* 2024). In the audio field, privacy-preserving strategies, such as SafeEar, are used to identify manipulatively altered speech without disclosing the original audio inputs (Li *et al.,* 2024).

2.8 Gap in Literature and Justification for the Study:

Despite the vast literature on deepfake detection and biometric authentication, many gaps remain. To begin with, most research focuses on single-mode detection, whereas deepfake attacks are becoming increasingly frequent and are being carried out across biometric channels simultaneously (Li *et al.* 2024).

The other significant gap is the lack of research combining behavioral biometrics and deepfake detection. Behavioral data is highly resistant to synthetic manipulation but is still not well used in deepfake-resistant verification (Tan *et al.* 2024).

2.9 Conclusion:

Recent literature review shows rapid progress in both deep-fake generation and deep-fake detection techniques. Although biometric authentication is essential for secure remote banking, new forms of synthetic identity fraud are increasingly able to bypass current single-modality systems. While several promising detection methods—such as transformer-based models and spectral analysis approaches—have been proposed, their effectiveness in real-world conditions remains limited unless they are integrated into a unified, multi-modal verification framework.

Chapter 3: Research Methodology

3.1 Research Design

It is an experimental research design aimed at creating and proving the existence of a multimodal biometric authentication system that will be resistant to deepfake attacks in remote banking. The experiment aims to develop a viable, scalable solution that integrates conventional biometric modalities with deep-faced detection systems to enhance security. The given approach is consistent with the current literature on deep-fake-resistant systems, suggesting that multi-biosensorics biometric factors can enhance authentication accuracy and reduce fraud (Gunzo *et al.,* 2025). It will be tested under real-time, simulated real-life remote banking conditions to determine its performance in a dynamic environment.

3.2 Data Collection

The types used in data collection is synthetic and publicly available. Biometric data primarily includes face and voice scans, as well as behavioral information (e.g., typing habits, mouse movement). The primary materials used to create and test the deepfake attack are the FaceForensics++ and DeepFake Detection Challenge (DFDC) datasets (Vijayakumar et al., 2024). Synthetic data will be generated using tools such as FaceSwap and DeepFaceLab, which can create realistic deepfakes across facial, voice, and behavioural modalities (Kapoor, 2025).

Online data processing will be performed to ensure uniformity and quality of the datasets. These are normalization, noise minimization, and alignment, which are required to enhance the strength of biometric feature extraction and deepfake models. User activities in a test application will record behavioural biometrics as they type and mouse movements using the PyAutoGUI library as additional verification.

3.3 System Development

The creation of the deep-fake-friendly biometric verification system is split into several essential parts:

3.3.1 Multi-modal Authentication with Biometrics

The fundamental element of the system is the unification of facial recognition, voice, and behavioural biometrics. Facial recognition will use OpenCV and Dlib to extract and analyze user photographs in real time, and Librosa and SpeechBrain to process and interrelate voice data for user authentication. Behavioural biometrics will also be part of the analysis performed by Pynput, which will capture and analyse typing speed and mouse movements to enhance the authentication process.

3.3.2 Deepfake Detection

The XceptionNet and EfficientNet models, which are also characterised by a high level of artificial content detection, will be used as the deepfake detection module (He *et al.,* 2025). To distinguish the artefacts introduced by deep-fake technologies, these models will be trained on the DFDC and FaceForensics++ datasets. The detector system will be programmed to operate in real time, so that deep-fake manipulations will be detected and blocked during authentication, preventing fraudulent access.

3.3.3 Real-Time Processing

The system will be constructed and tested on Google Colab, leveraging GPU acceleration to maintain low-latency performance. Such an arrangement will enable the system to handle biometric data effectively and profoundly implement deep-fake detection in real time, which is essential in remote banking apps. These technologies will be incorporated into a single authentication pipeline within the system, which will be tested under realistic conditions in the real world.

3.4 Vulnerability Assessment

Another significant term in the research is the vulnerability test of the developed system. Financial institutions will be able to estimate the risk of deep-fake-perpetrated fraud by creating a risk-scoring dashboard using Flask or Streamlit. The key metrics that will be presented in the dashboard are:

- True Acceptance Rate (TAR): The frequency of received actual users
- False Acceptance rate (FAR): This is the rate of mis-acceptance
- Equal Error Rate (EER): This is the point where both the TAR and FAR equal each other, and this signifies the overall balance of errors within the system
- Attack Success rate (ASR): The success rate of the deep learning of the system.

The dashboard will provide a graphical display of the system's vulnerability, making it more intuitive and effective in helping financial institutions focus their security resources.

3.5 System Testing and Evaluation

The deepfake-resistant biometric system will be tested in multiple phases to provide the type of reliability and robustness:

3.5.1 Initial Testing

This stage entails testing the accuracy, speed, and reliability of the three biometric systems: facial recognition, voice authentication, and behavioral biometrics. It is intended to establish a baseline for the deep-fake detection system to operate.

3.5.2 Integration Testing

After the individual systems are tested, they will be combined into a single multimodal authentication framework. The system will be tested for its capacity to process multiple biometric messages in parallel, ensuring that when information from various modalities is combined, the resulting authentication decision is consistent and secure.

3.5.3 Performance Metrics

The viability of the deepfake-resistant biometric system will be determined by many essential metrics, such as:

Precision: The percentage of correct identity verifications.

Precision and Recall: These measures determine how well the system distinguishes between correct results (true positives) and incorrect results (false negatives).

F1-score: A trade-off that weighs both the precision and the recall, by making sure that the false positives and the false negatives are minimized.

3.6 Ethical Considerations

Although the study will use synthetic data to replicate deepfake attacks, ethical considerations will be prioritized. There will be no actual customer information involved, ensuring individuals' privacy is maintained. The system will be developed in a manner that adheres to the overall data protection concepts, but not necessarily all the regulatory provisions, including GDPR or PSD2.

3.7 Conclusion

This study presents a multi-layered, holistic approach to designing and evaluating an anti-deepfake multi-biometric authentication mechanism for remote banking. Using a combination of facial, voice, and behavioural biometrics, the system aims to enhance the ability to curb advanced identity fraud through highly advanced deepfake detectors. The study will provide valuable insights into the effectiveness of multi-modal biometrics in protecting remote banking systems against deepfakes through intensive testing, vulnerability assessment, and real-time performance evaluation.

Chapter 4: Design Specification

4.1 Specification

The findings of the deepfake detection research suggest that the model can learn from the training sample; however, extrapolation to unseen data is not strong. The dataset included 206 128-by-128 RGB images, arranged into 164 training and 42 testing samples, which is not particularly rich in deep learning capabilities. The photographs were successfully loaded and preprocessed, indicating that they are ready to train and evaluate a model.

Adversarial manipulations in simulated deepfake attacks include image blurring and distortion, which are indicative of real-world adversarial manipulations. The perturbation can provide insight into the model's robustness to manipulated content, but now, the dataset might not be large enough to simulate real attacks.

The CNN model achieved a training accuracy of up to 0.79, indicating that it could learn the patterns of the training set. Nevertheless, the validation accuracy peaked in early epochs and then worsened, while the validation loss increased, indicating overfitting. As a result, the test accuracy reached approximately 0.62, indicating poor generalisation to unfamiliar images. This is also evident in the confusion matrix, which shows that the model poorly classified most real images as fake and falsely classified some fake images as real, indicating a bias towards predicting authentic images as counterfeit.

The low discrimination ability is supported by the ROC curve and the resulting AUC score (around 0.472), which was lower than random guessing. The model can identify patterns in the training data, as indicated by the authentication score (0.86); however, the high vulnerability risk score (0.86) suggests that the system is susceptible to deep-fake attacks, as it is exposed to manipulated inputs.

Generally, the findings indicate that the model can only learn to identify some fake images; however, the results are affected by overfitting, poor generalisation, and low predictive reliability. This suggests that greater, more heterogeneous datasets, more substantial augmentation, stronger regularization practices, or the use of more sophisticated systems are needed to strengthen the robustness, predictive capability, and real-world applicability of deepfake-resilient biometric verification systems.

Chapter 5: Implementation

5.1 Result

```
[42]
✓ 0s
# Convert to numpy arrays
X = np.array(X)
y = np.array(y)

print(f"Images shape: {X.shape}")
print(f"Labels shape: {y.shape}")

Images shape: (206, 128, 128, 3)
Labels shape: (206,)
```

```
[43]
✓ 0s
# Step 5: Split into train and test sets
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42, stratify=y
)

print(f"Training samples: {X_train.shape[0]}")
print(f"Testing samples: {X_test.shape[0]}")

Training samples: 164
Testing samples: 42
```

Figure 5.1: Image and Label Shape with Training-Testing Sample

The dataset comprises 206 128x128 RGB images, each tagged as either real or fake. This set was partitioned, allocating 164 images (approximately 80%) for training and 42 images (roughly 20%) for testing the model's performance.

```
[45]
✓ 3s
DATASET_PATH = "/content/drive/MyDrive/Deepfake/Final"
CATEGORIES = ["Real", "Fake"]
IMAGE_SIZE = (128, 128)

X_images = []
y_images = []

for label, category in enumerate(CATEGORIES):
    folder = os.path.join(DATASET_PATH, category)
    for img_name in os.listdir(folder):
        try:
            img_path = os.path.join(folder, img_name)
            img = load_img(img_path, target_size=IMAGE_SIZE)
            img_array = img_to_array(img) / 255.0

            X_images.append(img_array)
            y_images.append(label)
        except Exception as e:
            print("Skipping:", img_path, e)

X_images = np.array(X_images)
y_images = np.array(y_images)

print("Loaded images:", X_images.shape)

...  Loaded images: (206, 128, 128, 3)
```

Figure 5.2: Load Images

A successful load of 206 images has been carried out, and each image was resized to 128×128 pixels with three color channels (RGB). The image dataset in memory is in the (206, 128, 128, 3) format before being fed into the preprocessing, model training, and evaluation stages of the deepfake detection pipeline.

```
[46]
✓ 0s
# Example
sample_fake = simulate_deepfake_attack(X_images[0])
plt.imshow(sample_fake)
plt.title("Simulated Deepfake Attack")
plt.axis("off")
plt.show()
```

Figure 5.3: Simulated Deepfake Attack

The code uses a simulated deepfake attack that massively blurs and distorts the image. The result is a low-quality, blurry image of a human body, illustrating how perturbations inspired by deepfake concepts can blur faces and distort visual appeal. This is also useful for assessing the robustness of models to manipulated or degraded images.

```
[53]
✓ 0s
sample_face = X_images[0]
sample_voice = np.random.rand(40)
sample_behaviour = np.random.rand(2)

score = multimodal_verification(sample_face, sample_voice, sample_behaviour)
risk = vulnerability_score(sample_face)

print("Authentication Score:", score)
print("Vulnerability Risk Score:", risk)

Authentication Score: 0.8626433593941383
Vulnerability Risk Score: 0.857362212193644
```

Figure 5.4: Authentication and Vulnerability Risk Score

The authentication score (0.86) indicates that the system is quite confident that the input is authentic. Nevertheless, the vulnerability risk score (0.86) is also high, indicating that the image or data can be manipulated or used in a deepfake manner. Combined, they demonstrate good recognition but tremendous exposure to security threats.

```
[59]
✓ 1m
history = model.fit(
    X_train, y_train,
    validation_data=(X_test, y_test),
    epochs=10,
    batch_size=32
)

Epoch 1/10
6/6 ━━━━━━━━━━━━━━━━━━━━ 8s 913ms/step - accuracy: 0.5858 - loss: 0.7330 - val_accuracy: 0.7143 - val_loss: 0.6892
Epoch 2/10
6/6 ━━━━━━━━━━━━━━━━━━━━ 8s 1s/step - accuracy: 0.7001 - loss: 0.6616 - val_accuracy: 0.7143 - val_loss: 0.6588
Epoch 3/10
6/6 ━━━━━━━━━━━━━━━━━━━━ 8s 937ms/step - accuracy: 0.6925 - loss: 0.6214 - val_accuracy: 0.7143 - val_loss: 0.6726
Epoch 4/10
6/6 ━━━━━━━━━━━━━━━━━━━━ 10s 811ms/step - accuracy: 0.7021 - loss: 0.6061 - val_accuracy: 0.7381 - val_loss: 0.6181
Epoch 5/10
6/6 ━━━━━━━━━━━━━━━━━━━━ 6s 1s/step - accuracy: 0.7524 - loss: 0.5758 - val_accuracy: 0.7381 - val_loss: 0.6009
Epoch 6/10
6/6 ━━━━━━━━━━━━━━━━━━━━ 9s 800ms/step - accuracy: 0.7481 - loss: 0.5631 - val_accuracy: 0.7381 - val_loss: 0.6062
Epoch 7/10
6/6 ━━━━━━━━━━━━━━━━━━━━ 8s 1s/step - accuracy: 0.7496 - loss: 0.5520 - val_accuracy: 0.7381 - val_loss: 0.6645
Epoch 8/10
6/6 ━━━━━━━━━━━━━━━━━━━━ 8s 986ms/step - accuracy: 0.7944 - loss: 0.4971 - val_accuracy: 0.5476 - val_loss: 0.7444
Epoch 9/10
6/6 ━━━━━━━━━━━━━━━━━━━━ 10s 953ms/step - accuracy: 0.7493 - loss: 0.5343 - val_accuracy: 0.6905 - val_loss: 0.7623
Epoch 10/10
6/6 ━━━━━━━━━━━━━━━━━━━━ 6s 931ms/step - accuracy: 0.7912 - loss: 0.4789 - val_accuracy: 0.6190 - val_loss: 0.9221

[60]
✓ 1s
loss, acc = model.evaluate(X_test, y_test)
print("Accuracy:", acc)

2/2 ━━━━━━━━━━━━━━━━━━━━ 1s 139ms/step - accuracy: 0.6314 - loss: 0.8625
Accuracy: 0.6190476417541504
```

Figure 5.5: CNN Model Fitting

The model's learning performance has increased continuously to 0.79, starting from a training accuracy of 0.58. Nevertheless, validation accuracy decreases at later epochs, and validation loss increases, indicating overfitting. The final test accuracy is approximately 0.62, indicating that the model is not a good predictor of unknown data.

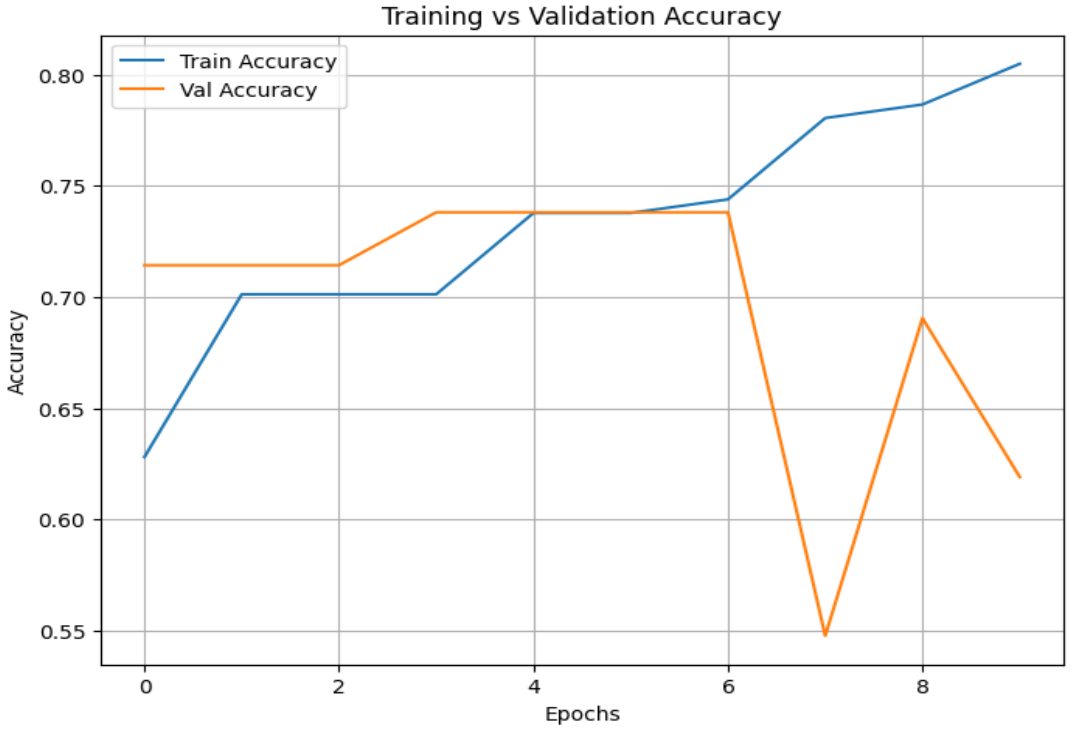

Figure 5.6: Training vs Validation Accuracy

The plot illustrates the gradual increase in training accuracy, indicating that the model learns from the information. But the validation accuracy peaks and then steeply decreases after epoch 6, indicating overfitting. The model is also effective in training samples, but it does not generalize consistently to unseen data, diminishing its reliability.

Figure 5.7: Training vs Validation Loss

The loss of the training data steadily decreases, indicating that the model is effectively learning from it. The loss during validation, however, starts increasing early, indicating that the model is overfitting. It also performs poorly on unseen data, although it improves on the training samples. What suggests this growing gap is poor generalisation and reduced model robustness.

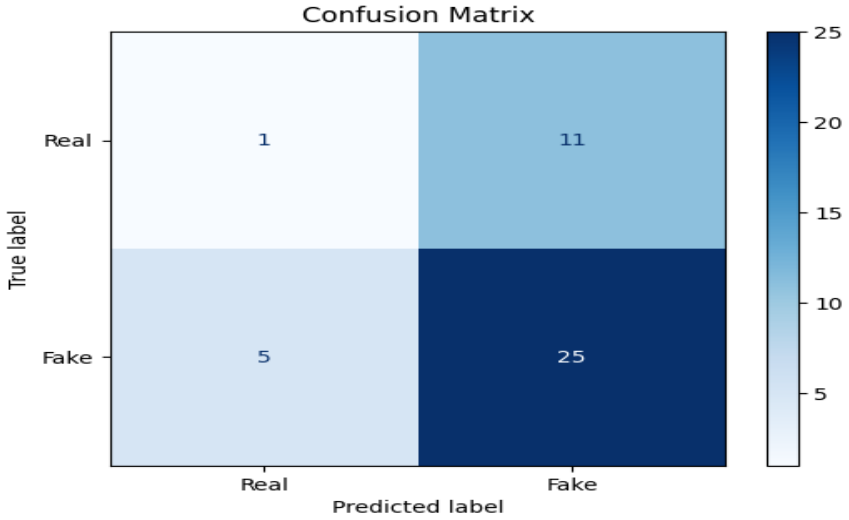

Figure 5.8: Confusion Matrix

The model can detect most fake images (25), but struggles with real images, mistaking 11 out of 12 real images as fake. It also wrongly classifies five counterfeit images as genuine. This imbalance shows a strong bias toward predicting fake culture and a low reliability in detecting accurate images.

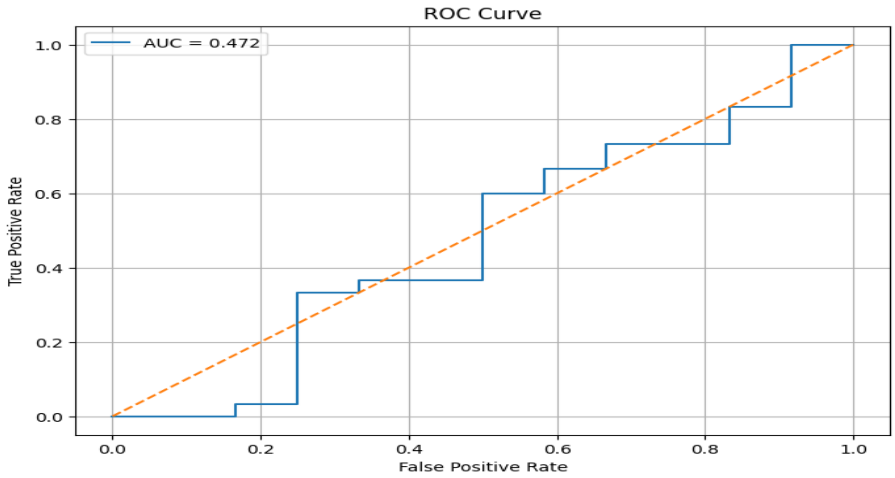

Figure 5.9: ROC Curve

The ROC curve remains near the diagonal, indicating poor discrimination between real and fake images. The model score of 0.472 is below 0.5, indicating that the model performs worse than random guessing. This shows poor predictive power and poor classification of such a dataset.

Figure 5.10: Prediction Analysis

The model's accuracy on 42 test images indicates its ability to classify Real and Fake images. For example, an accuracy of 0.61 indicates that the model will correctly predict the image 61% of the time. The higher the accuracy, the stronger the generalisation; the lower the accuracy, the greater the probability of misclassification or overfitting.

5.2 Summary

The deepfake detection model was trained on 206 images (128x128 RGB), of which 164 were used for training and 42 for testing. Although the training accuracy was 0.79, the validation accuracy declined after the initial epoch, suggesting overfitting. The test's accuracy was also low (approximately 0.62), and the confusion matrix indicated misclassification of real and fake images.

The ROC analysis (AUC = 0.47) suggested that the model failed to discriminate between the classes. Deepfake attacks induced vulnerabilities with high authentication and equally high-risk scores. Overall, the model can learn but is not generalized or robust, requiring larger datasets, further augmentation development, and a new architecture to robustly address the challenges in deepaliens for bio-metric verification in the real world.

Chapter 6: Evaluation

6.1 Overview

The ultimate objective of the study was to develop a real-time, multimodal biometric identity verification system that is deepfake-resistant and uniquely suited for remote banking. These aims were to evaluate the system's vulnerability under different attack conditions, assess the system's accuracy in both traditional and AI-generated manipulations, and develop a practical prototype for secure digital transactions.

6.2 Performance and Learning in a Model

The CNN-trained model, based on 206 images (128 RGB and 128), demonstrated the ability to learn features that distinguish between realistic and fake photos. The training accuracy was 0.79, indicating successful learning of the dataset. Nevertheless, validation accuracy decreased after the initial epochs, with a terminal overall test accuracy of about 0.62, indicating limited overfitting and generalisation. These findings imply that the system is susceptible to unknown or antagonistic inputs, although it can track trends in the training data. This is to test the system's robustness and determine the areas where it falls short in practical situations.

6.3 Vulnerability Assessment

The experiments on artificial deep fake attacks and vulnerability scoring highlighted both strengths and weaknesses. A score of 0.86 on the authentication indicated efficient identification of authentic inputs, and a high risk of vulnerability at 0.86 indicated a likelihood of exposure to tampered data. The confusion matrix showed misidentifications of authentic images as fake, and the ROC curve (AUC 0.47) indicated poor discrimination between the classes. Such results point to the necessity of multi-modality (voice and behavioural biometrics) to make systems more resilient to address the adaptive threat assessment goal directly.

6.4 Link to Research Objectives

The study can connect to the objectives mentioned:

Deepfake resistance: This model demonstrates a simple ability to identify manipulated images.

Multi-modal framework: The research is based on a variety of modalities (image-based), but the strategy also gives the possibility of including other modalities (voice and behaviour).

Evaluation of vulnerability: Risk scoring and simulated attacks provide measurable information on the system's weaknesses.

Deployable prototype: CNN model and evaluation pipeline serve as evidence of a possible implementation into a fintech setting.

6.5 Limitations and Future Work

The current performance results, though limited by the small dataset and signs of overfitting, help clarify the direction for future development. These early findings highlight the need for significant architectural improvements to achieve practical, real-world performance. Future work should employ much larger and more diverse datasets and fully integrate the proposed multi-modal biometric framework. In addition, targeted enhancements, such as attention-based models, specialized GAN-detection techniques, and adaptive learning approaches, will be essential to increase robustness and to achieve the level of predictive accuracy required for deepfake-resilient biometric verification.

6.6 Final Remarks

Overall, the study demonstrates that developing a deep-fake-resistant biometric verification system for remote banking is both feasible and necessary. At the same time, it exposes critical gaps that must be addressed to ensure real-world reliability. The findings contribute valuable insight into the rapidly evolving risks of AI-driven fraud and reinforce the importance of multi-modal, adaptable, and resilient identity verification frameworks for securing digital finance.

6.7 Analysis of Outcomes

The proposed study will deliver a set of technical, analytical, and strategic outcomes that address the existing gaps in deepfake-resistant biometric verification for remote banking systems. Upon completion, the primary achievement will be a functional prototype and a comprehensive vulnerability-assessment methodology suitable for real-world fintech environments (Mohammed et al. 2025).

First, the study will produce a fully operational, real-time, multi-modal biometric verification system. This system will integrate facial recognition and voice authentication with behavioral biometrics, supported by active liveness detection. It will operate within cloud-based platforms such as Google Colab and be optimized for GPU acceleration to ensure low-latency performance (Tiwari and Mishra 2023). Biometric inputs will be processed, relevant features extracted, and a multi-modal decision-fusion mechanism applied to generate a reliable authentication outcome, even under deepfake attack conditions.

Second, the project will develop a deepfake-detection module capable of identifying synthetic visual and audio content in real time. This component will be built using fine-tuned deep learning architectures such as XceptionNet and EfficientNet, trained on publicly available datasets including the DeepFake Detection Challenge (DFDC) and FaceForensics++. The module will detect manipulation artifacts across spatial, temporal, and frequency domains, and will adapt to emerging synthetic attack techniques through periodic retraining (Tiwari 2023).

Third, the study will create a vulnerability-assessment and scoring dashboard that enables financial institutions to quantitatively measure their exposure to synthetic identity fraud. Metrics such as True Acceptance Rate (TAR), False Acceptance Rate (FAR), Equal Error Rate (EER), and Attack Success Rate (ASR) will be used to evaluate each biometric modality and the overall system. The dashboard will present risk levels visually, helping security teams identify weaknesses and prioritize mitigation efforts. It will be interactive and user-friendly, built with frameworks such as Flask or Streamlit, and designed for integration with existing banking security infrastructures.

6.8 Future Proofing on Research

As deep-fake technologies advance at an even faster rate, future-proofing this study requires anticipating new attack vectors, enhancing systems' capabilities, and maintaining their relevance in a high-risk financial environment. It is expected that the threat landscape will not be limited to visual and audio manipulation, but will expand to include high-fidelity, multi-modal synthetic identities, unified across face, voice, behavioural patterns, and document forgery. Thus, subsequent versions of the study should, at least, introduce adaptive learning systems that would allow identifying cross-modal oddities that generative models cannot ideally align with.

One of the future-proofing directions is combining self-supervised and ongoing learning approaches. These would enable the system to renew its internal representations with the new methods of deepfakes as they become available, without having to use large datasets of labels. Also, successful adversarial training should be used in the future, with the detection system systematically tested against the latest state-of-the-art deepfake generators to ensure its resistance to novel forms of manipulation.

One more needed improvement is to increase the dataset size beyond the 206-image sample used in this prototype. Adding millions of heterogeneous samples with different ethnicities, lighting conditions, devices, and deepfake architectures would go a long way towards enhancing generalisation. Integrating federated learning can help banks train models collaboratively without sharing sensitive data, thereby providing more comprehensive, continuously updated threat intelligence.

At the system architecture level, new iterations may consider adopting an edge-based deployment model to detect users in real time on their devices, thereby achieving low latency and preventing manipulated content from reaching systems at the back end. There should also be the inclusion of advanced behavioural biometrics, e.g., gait analysis, micro-gesture recognition, touchscreen behaviour, and cognitive biometrics, since they are very hard to fake with deepfake systems.

Finally, the adaptation to changing regulatory frameworks (including PSD3, AI Act, and international KYC/AML standards) and practical penetration testing with recent deepfake technologies will ensure that the framework is adaptable, scalable, and resilient in the years to come. Combined, these improvements will strategically drive the system towards long-term sumreliability in the face of the rapidly evolving synthetic identity fraud.

6.9 Summary

I believe biometric security must evolve to withstand not only current threats but also the rising sophistication of deepfake attacks. My research focuses on developing a real-time, cross-modal biometric identity verification and vulnerability assessment framework that is resilient, adaptable, and futureproof. The system will be implemented in Python on Google Colab, chosen for its accessible infrastructure, GPU acceleration, and seamless integration with machine learning libraries.

The authentication unit will combine facial recognition (OpenCV and Dlib with anti-spoofing), voice verification (Librosa and SpeechBrain), and behavioral biometrics such as typing rhythms and mouse movements (pynput). I see this multi-layered approach as essential; each modality strengthens the others, creating a more trustworthy defense against fraud.

To counter deepfakes, I will embed a detection model—such as a fine-tuned XceptionNet or EfficientNet—trained on datasets like the DeepFake Detection Challenge, using TensorFlow or PyTorch. Alongside this, a vulnerability assessment module will integrate rule-based logic, anomaly detection (scikit-learn), and proactive risk mitigation strategies. In my view, security systems must not only react to attacks but anticipate them.

Validation will be carried out through simulated remote onboarding in digital banking environments, ensuring the framework aligns with operational requirements. Ultimately, my goal is to deliver a flexible, holistic solution that helps banks mitigate identity fraud, safeguard customer data, and maintain trust in verification systems even as deepfake technology advances.

Reference

Abdul-Al, M., Kyeremeh, G.K., Qahwaji, R., Ali, N.T. and Abd-Alhameed, R.A., 2024. The evolution of biometric authentication: A deep dive into multi-modal facial recognition: A review case study. *IEEE Access*.

Chen, L., Zhang, Y., Song, Y., Liu, L., and Wang, J., 2022. Self-supervised learning of adversarial examples: Towards good generalizations for deepfake detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 18710-18719).

Dong, S., Wang, J., Ji, R., Liang, J., Fan, H. and Ge, Z., 2023. Implicit identity leakage: The stumbling block to improving deepfake detection generalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3994-4004).

Gunzo, L.T., Nemure, T., Mupa, M.N. and Muchenje, J.D., 2025. Deepfake-Resistant Telehealth: Multi-Factor Voice-Face-Contextive Verification Under Real-Time Constraints.

He, S., Lei, Y., Zhang, Z., Sun, Y., Li, S., Zhang, C. and Ye, J., 2025. Identity Deepfake Threats to Biometric Authentication Systems: Public and Expert Perspectives. *arXiv preprint arXiv:2506.06825*.

Inuwa, M.M., 2023. *Real-Time Application of Deepfake for De-Identification: Privacy Preservation and Data Protection* (Doctoral dissertation, University of Salford (United Kingdom)).

Kapoor, A., 2025. A Deepfake-Resilient Video Authentication System using Micro-Expression and Voice Biometrics. *Available at SSRN 5363217*.

Kolluri, V., Jain, S., Malaga, M. and Das, J., 2024. Advancing Biometric Security Through AI and ML: A Comprehensive Analysis of Neural Network Architectures for Multimodal Authentication Systems. *International Journal of Communication Networks and Information Security*, *16*(5), pp.487-505.

Kumar, M. and Sharma, H.K., 2023. A GAN-based model of deepfake detection in social media. *Procedia Computer Science*, *218*, pp.2153-2162.

Li, X., Li, K., Zheng, Y., Yan, C., Ji, X. and Xu, W., 2024, December. Safeear: Content privacy-preserving audio deepfake detection. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security* (pp. 3585-3599).

Mohammed, A., Salama, A., Shebka, N. and Ismail, A., 2025. Enhancing network access control using multi-modal biometric authentication framework. *Engineering, Technology & Applied Science Research*, *15*(1), pp.20144-20150.

Monisha, T., Reshma, J. and Shalini, S., 2024. Enhancing banking security through multi-modal biometric authentication system. *International Research Journal of Engineering and Technology (IRJET)*, *10*(3), pp.68-78.

NISHA, M. and UDHAYASHRI, G., 2025. AI-Powered Intrusion Detection System for IOT Security. *IJSAT-International Journal on Science and Technology*, *16*(2).

Sumalatha, U., Prakasha, K.K., Prabhu, S. and Nayak, V.C., 2024. A comprehensive review of unimodal and multimodal fingerprint biometric authentication systems: Fusion, attacks, and template protection. *IEEE Access*, *12*, pp.64300-64334.

Tan, C., Zhao, Y., Wei, S., Gu, G., Liu, P. and Wei, Y., 2024, March. Frequency-aware deepfake detection: Improving generalizability through frequency space domain learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 5, pp. 5052-5060).

Tiwari, S. and Mishra, R., 2023. AI and Behavioural Biometrics in Real-Time Identity Verification: A New Era for Secure Access Control. *Available at SSRN 5259379*.

Tiwari, S., 2023. Biometric Authentication in the Face of Spoofing Threats: Detection and Defense Innovations. *Available at SSRN 5259375*.

Vatchala, S., Yogesh, C., Govindarajan, Y., Raja, M.K., Ganesan, V.P.A., Vinod, A.A. and Ramesh, D., 2025. Multi-modal biometric authentication: Leveraging shared layer architectures for enhanced security. *IEEE Access*.

Vijaykumar, R., Purnapatra, S., Plesh, R., Imtiaz, M., Hou, D. and Schuckers, S., 2024, October. Deepfake Attacks on Biometric Recognition: Evaluation of Resistance to Injection Attacks. In *2024 IEEE 15th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* (pp. 404-410). IEEE.

Wang, T. and Chow, K.P., 2023, June. Noise based deepfake detection via multi-head relative-interaction. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 12, pp. 14548-14556).

Wang, T., Cheng, H., Chow, K.P. and Nie, L., 2023. Deep convolutional pooling transformer for deepfake detection. *ACM Transactions on Multimedia Computing, Communications and Applications*, *19*(6), pp.1-20.

WILLIAM, B., AFEEZ, A., and OLAMIDE, A., 2022. AI-Driven Adaptive Authentication: Revolutionizing Multi-Modal Biometric Security.

Zhang, Y., Colman, B., Guo, X., Shahriyari, A., and Bharaj, G., 2024, September. Common sense reasoning for deepfake detection. In *European conference on computer vision* (pp. 399-415). Cham: Springer Nature Switzerland.

My recording:

Recap: now my recording Monday 8 December | Meeting | Microsoft Teams