COMSATS UNIVERSITY ISLAMABAD

Lahore Campus

Department of Computer Engineering

Real-Time Two-Layer Contactless Authentication on an Edge Computing Platform

Proposal for Final Year Project - I

Submitted by:

Abdullah Laeeq - CIIT/FA22-BCE-026/LHR

Ali Hamza - CIIT/FA22-BCE-071/LHR

Muhammad Faizan Shurjeel - CIIT/FA22-BCE-086/LHR

Supervised by:

Dr. Zaid Ahmad

Co-Supervisor: Engr. Talha Naveed

Department of Computer Engineering

COMSATS University Islamabad, Lahore Campus

Group Members Signatures		
Abdullah Laeeq (Member's Signature)	**Ali Hamza** (Member's Signature)	**Faizan Shurjeel** (Member's Signature)

Dr. Zaid Ahmad
(Checked and Signed by the Supervisor)

Contents

1 ABSTRACT

Traditional authentication methods, such as passwords, PINs, and physical tokens, are increasingly insufficient in the face of modern security threats. The global emphasis on hygiene has further highlighted the risks of contact-based biometrics. This project addresses these security and hygiene gaps by proposing a Two-Layer Contactless Physical Authentication System, leveraging facial and voice biometrics for secure, touch-free identity verification. A fundamental challenge, however, is the deployment of computationally expensive deep learning models on low-cost, resource-constrained hardware. This proposal directly confronts this challenge by integrating state-of-the-art model optimization techniques into its core methodology.

The project's foundation is the implementation of advanced deep learning models for biometric analysis. For facial recognition, we will use an architecture based on ArcFace to generate discriminative facial embeddings. This will be complemented by a speaker verification layer powered by a modern model such as ECAPA-TDNN [2]. To ensure real-time performance on an embedded platform like the Raspberry Pi 4, we will employ a robust optimization strategy centered on low-bit quantization. Informed by cutting-edge frameworks like QuantFace [5], which demonstrate the ability to reduce model size by over 4x with minimal accuracy loss, we will optimize our models for efficient CPU-based inference. This approach will enable the creation of a Minimum Viable Product (MVP) that is portable, affordable, and suitable for widespread practical deployment.

2 INTRODUCTION

2.1 Overview

In an increasingly interconnected world, the need for robust identity verification has never been more critical. While single-factor biometric systems can achieve high security, systems that offer multiple authentication modalities provide significant advantages in user experience and resilience. This project explores this domain by combining two of the most natural human biometrics: the face and the voice. This dual-option approach allows users to choose their preferred method based on environmental conditions—if poor lighting affects facial recognition, voice verification remains a viable alternative, and vice versa—creating a system that is both user-friendly and robust. The ultimate goal is to move beyond theoretical models and create a system that is deployable in the real world on affordable, accessible hardware.

2.2 Problem Statement

The project aims to solve several key problems with existing authentication systems:

- **Limited Authentication Options:** Most systems rely on a single authentication method, leaving users without alternatives when conditions are not ideal.
- **Environmental Dependencies:** Facial recognition struggles in poor lighting, while voice recognition fails in noisy environments. A dual-option system provides resilience.
- **Computational Cost and Deployability:** State-of-the-art AI models for face and voice recognition are computationally expensive and too large to run efficiently on low-cost embedded hardware like a Raspberry Pi without significant optimization.
- **Hygiene Concerns:** Contact-based systems are a vector for germ transmission. A contactless solution eliminates this risk entirely.
- **Lack of Accessibility:** High-security biometric systems are often expensive and proprietary. There is a need for an affordable and adaptable open-source solution.

2.3 Objectives

To address these issues, we have defined the following project objectives:

1. To conduct a thorough investigation of state-of-the-art deep learning models for both facial recognition and speaker verification, with a specific focus on architectures suitable for edge deployment.
2. To design and develop a modular software pipeline for facial recognition capable of accurate face detection, feature extraction, and unique embedding generation.
3. To design and develop a parallel software pipeline for speaker verification that can process an audio clip and generate a unique speaker embedding.
4. To integrate these two pipelines with a decision fusion engine that grants access when either facial OR voice authentication succeeds.
5. To implement comprehensive anti-spoofing (liveness detection) mechanisms for both modalities to prevent attacks using photos, videos, or recordings.
6. To optimize the selected deep learning models for edge deployment using low-bit quantization, informed by frameworks like QuantFace [5] and Ef-QuantFace [6], to ensure real-time performance on a Raspberry Pi 4.

7. To professionally evaluate the system's performance using standard biometric metrics, including False Acceptance Rate (FAR) and False Rejection Rate (FRR), as well as deployment metrics like latency.

2.4 Deliverables

Upon successful completion of the Final Year Project, we will deliver the following:

1. **A Working Proof-of-Concept Application:** A functional software system that demonstrates the complete two-layer authentication process, from data capture to the final access decision, running on the chosen MVP hardware.
2. **A Comprehensive FYP Final Report:** Detailed documentation covering the project's background, literature review, methodology, system design, implementation details, testing procedures, results, and conclusion.
3. **Source Code Repository:** A well-documented Git repository containing all the code for the AI models, optimization scripts, integration logic, and any associated applications.
4. **Final Presentation & Live Demonstration:** A final presentation summarizing the project and a live demonstration of the working prototype.

3 LITERATURE REVIEW

3.1 Facial Recognition Systems

Modern facial recognition leverages Convolutional Neural Networks (CNNs) to learn discriminative features. Architectures like Google's **FaceNet** pioneered the use of a triplet loss to learn a 128-dimensional embedding space. The current state-of-the-art is largely defined by loss functions like **ArcFace**, which introduces an additive angular margin loss to improve the discriminative power and separability of the learned 512-dimensional embeddings. For the initial pre-processing step of finding the face, Google's **BlazeFace** [4] presents a hyper-efficient face detector designed for mobile GPUs, capable of running at over 200 FPS by using a novel architecture with 5x5 kernels.

3.2 Speaker Verification Systems

Speaker verification confirms a speaker's identity from their voice. **x-vector** [3] systems became a robust baseline, using a Deep Neural Network to extract a fixed-dimensional speaker embedding from variable-length speech. The current state-of-the-art is represented

by the **ECAPA-TDNN** [2] architecture, available in toolkits like **SpeechBrain** [1]. It uses channel attention and propagation mechanisms to create more powerful and noise-robust speaker embeddings.

3.3 Model Optimization and Edge Deployment

Deploying the aforementioned models on resource-constrained hardware like a Raspberry Pi presents a significant challenge. Recent academic literature provides a clear pathway to solving this problem through model optimization, primarily via low-bit quantization.

Two independent studies, by Froiz-Míguez et al. [8] and Le et al. [9], have confirmed the feasibility of running complex, single-modal voice recognition on a Raspberry Pi 4. Both studies, using entirely different AI architectures (a modern transformer and a classic HMM+ANN, respectively), established a consistent performance benchmark of 1.5 to 2.0 seconds for CPU-only inference, validating the Pi 4 as a capable platform.

The core technique for achieving such performance is quantization. A key paper by Bunda et al. [7] focuses specifically on sub-byte quantization of MobileFaceNet. It empirically proves that this architecture can be quantized down to 4-bits, resulting in an 8x model size reduction (from over 4MB to 0.5MB) with an almost negligible drop in accuracy on the LFW benchmark (98.85% vs 98.63%). This validates that face recognition models are highly amenable to aggressive quantization.

However, quantization requires a fine-tuning step to regain lost accuracy, which traditionally needs the original, often private, training dataset. A groundbreaking framework, QuantFace [5], solves this problem. It presents a methodology to fine-tune quantized models using unlabeled, synthetically generated face data and knowledge distillation. An evolution of this work, Ef-QuantFace [6], further streamlines this process, proving that a massive dataset is not required; fine-tuning can be achieved in just 15 minutes using a small, public dataset of 14,000 images and a more advanced technique called Evaluation-oriented Knowledge Distillation (EKD). These frameworks provide a state-of-the-art, privacy-preserving, and highly practical methodology for our project's optimization phase.

4 METHODOLOGY

4.1 System Architecture

Our proposed methodology is structured into five distinct phases, creating a systematic development process. The final system architecture is visualized in Figure 1.

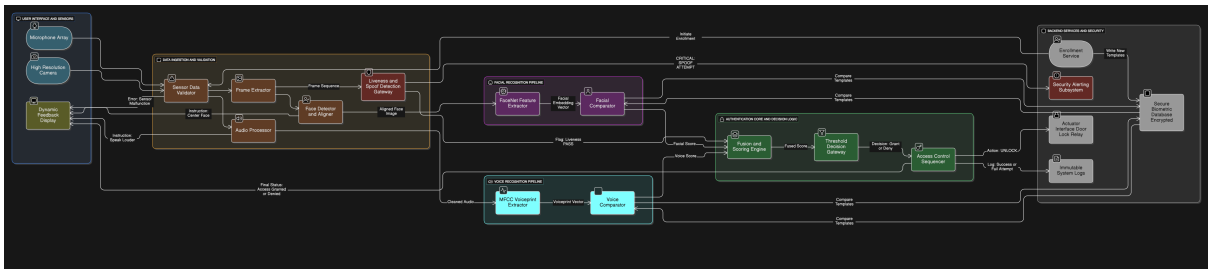

Figure 1: Proposed System Architecture Block Diagram.

Phase 1: Research and Model Selection. Based on our extensive literature review, we have selected an **ArcFace-based model** for facial recognition and the **ECAPA-TDNN** architecture from the **SpeechBrain toolkit** for speaker verification.

Phase 2: Modular Pipeline Development. The facial and voice pipelines will be developed as independent modules on a PC. The facial pipeline will utilize a hyper-efficient detector like **BlazeFace** [4] for the initial detection, followed by the ArcFace model for embedding extraction. The voice pipeline will process a 16kHz audio clip with the ECAPA-TDNN model.

Phase 3: Integration and Decision Fusion. The independent pipelines will be integrated into a unified system. A Decision Fusion Engine will be developed to receive the verification scores from both modules. The system will use an **OR gate logic**: access is granted if either the face OR the voice passes its respective similarity threshold, offering maximum user flexibility.

Phase 4: Security Hardening (Liveness Detection). To prevent basic spoofing attacks, we will implement anti-spoofing measures. The proposed method is eye-blink detection for the facial modality, using OpenCV to track facial landmarks and require a blink before authorizing the feature extraction step.

Phase 5: Hardware Prototyping and Optimization. The core optimization strategy will involve Quantization-Aware Training (QAT). Following the highly efficient methodologies presented in Ef-QuantFace [6] and the specific validation for models like MobileFaceNet by Bunda et al. [7], we will fine-tune the quantized models to regain accuracy. We will target 8-bit or potentially 4-bit quantization and convert the final models to the ONNX format for high-speed inference on the Raspberry Pi 4, aiming for the sub-2-second latency benchmark established in literature.

4.2 Hardware Implementation

The system will be implemented using the following hardware components:

- **Processing Unit:** Raspberry Pi 4 (4GB)
- **Camera Module:** High-resolution camera

- **Microphone:** Digital USB microphone
- **Storage:** High-endurance A2-rated MicroSD card

5 EXPECTED OUTCOMES

- A functional dual-option biometric authentication prototype on a Raspberry Pi 4.
- High accuracy for each modality (target more than 95% on test sets).
- Real-time performance with a target response time under 2 seconds.
- Implementation of liveness detection to prevent basic spoofing.
- A comprehensive final report detailing the system and its performance against academic benchmarks.

6 PROJECT TIMELINE

Phase	**Activities**	**Start Week**	**Duration**
Phase 1	Literature review, model selection, requirement analysis	Week 1	2 weeks
Phase 2	Facial recognition pipeline development	Week 3	3 weeks
Phase 3	Voice recognition pipeline development	Week 4	3 weeks
Phase 4	Pipeline integration and decision fusion	Week 7	2 weeks
Phase 5	Anti-spoofing implementation (Liveness Detection)	Week 9	2 weeks
Phase 6	Model Quantization and Optimization (QAT)	Week 11	3 weeks
Phase 7	Hardware deployment, testing, evaluation, and documentation	Week 14	2 weeks

7 BUDGET ESTIMATION

Component	Quantity	Cost (PKR)
Raspberry Pi 4 (4GB)	1	18,000
Camera Module v3	1	8,000
USB Microphone	1	3,000
7" LCD Display (Optional)	1	12,000
MicroSD Card (64GB A2-rated)	1	2,000
Official Power Supply (5V/3A)	1	2,500
Enclosure/Case	1	3,500
Miscellaneous Components	-	3,000
Total		**52,000**

8 REFERENCES

References

[1] Ravanelli, M., et al. (2021). SpeechBrain: A General-Purpose Speech Toolkit. *arXiv preprint arXiv:2106.04624*.

[2] Desplanques, B., Thienpondt, J., & Demuynck, K. (2020). ECAPA-TDNN: Emphasized Channel Attention, Propagation and Aggregation in TDNN Based Speaker Verification. *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

[3] Snyder, D., Garcia-Romero, D., Sell, G., Povey, D., & Khudanpur, S. (2018). X-vectors: Robust DNN embeddings for speaker recognition. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

[4] Bazarevsky, V., Kartynnik, Y., et al. (2019). BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs. *arXiv preprint arXiv:1907.05047*.

[5] Boutros, F., Damer, N., & Kuijper, A. (2022). QuantFace: Towards Lightweight Face Recognition by Synthetic Data Low-bit Quantization. *arXiv preprint arXiv:2206.10526*.

[6] Gazali, W., Kho, J. M., Santoso, J., & Williem. (2024). Ef-QuantFace: Streamlined Face Recognition with Small Data and Low-Bit Precision. *arXiv preprint arXiv:2402.18163*.

[7] Bunda, S., Spreeuwers, L., & Zeinstra, C. (2022). Sub-byte quantization of Mobile Face Recognition Convolutional Neural Networks. *2022 International Conference of the Biometrics Special Interest Group (BIOSIG)*.

[8] Froiz-Míguez, I., Fraga-Lamas, P., & Fernández-Caramés, T. M. (2023). Design, Implementation, and Practical Evaluation of a Voice Recognition Based IoT Home Automation System for Low-Resource Languages and Resource-Constrained Edge IoT Devices. *IEEE Access, 11*, 63623-63647.

[9] Le, V.-H., Luc, N.-Q., and Quach, D.-H. (2024). Developing a secure voice recognition service on Raspberry Pi. *Bulletin of Electrical Engineering and Informatics, 13*(5), 3544-3551.