National College of Ireland

TR-SONIC: A Multimodal Transformer-Based Sonar 3D Face Recognition System for Biometric Verification

MSc Research Project
MSc Artificial Intelligence for Business

Roque Ivan Lopez Lopez
Student ID: x23380501

School of Computing
National College of Ireland

Supervisor: Rajwanul Haque

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Roque Ivan Lopez Lopez

Student ID: x23380501

Programme: MSc AI for Business (MSCAIBUS) Year: 2025

Module: Practicum 2

Supervisor: Rajwanul Haque

Submission Due Date: 15 - September - 2025

Project Title: TR-SONIC: A Multimodal Transformer-Based Sonar 3D Face Recognition System for Biometric Verification

Word Count: 8,599 words **Page Count:** 23 Pages

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: Roque I Lopez

Date: 11 - September - 2025

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):	

TR-SONIC: A Multimodal Transformer-Based Sonar 3D Face Recognition System for Biometric Verification

Roque I. Lopez
x23380501

Abstract

This research presents TR-SONIC, a multimodal biometric authentication system that combines 3D facial geometry with simulated ultrasonic echoes processed using efficient transformer architectures. This work addresses the limitations of traditional 2D/3D vision-based facial recognition systems, which are vulnerable to poor lighting conditions, facial occlusions, impersonation attacks, and demographic biases. TR-SONIC proposes an innovative approach inspired by sonar technology, integrating two complementary models: SonarSimilarityTransformer, which compares acoustic signatures generated from 3D facial meshes, and FacialLandmarkTransformer, which employs K-Means clustering to detect anatomical landmarks. This dual-model architecture enables accurate and robust verification even in challenging scenarios.

The experiments, conducted on the FaceScape dataset, yielded an overall system accuracy of 95.7%, with the FacialLandmarkTransformer achieving 97.3% accuracy and the SonarSimilarityTransformer achieving 94.1% accuracy, and a 70% reduction in demographic bias metrics, while maintaining processing times of 0.42 seconds per authentication suitable for real-time environments. Furthermore, the system incorporates an intuitive graphical interface and an ethics monitoring framework that ensures compliance with emerging regulations on algorithmic fairness. TR-SONIC represents a significant advance toward more secure, fair, and suitable biometric systems for critical applications such as mobile banking and digital verification.

Keywords: *Biometric Authentication, Transformer Architectures, Multimodal Biometrics, Acoustic Sonar, 3D Face Recognition/Modelling, Bias Mitigation.*

1 Introduction

In recent years, biometric authentication has become a cornerstone of access control system security, especially with the growing use of high-end smart mobile devices featuring biometric authentication technology, as well as the inherent migration of financial banking to the online digital world (Yu *et al.*, 2025). Among a wide variety of biometric authentication systems, facial recognition has emerged as one of the most widely used due to its non-intrusive nature and easy integration. However, conventional facial recognition systems, particularly those based on 2D images analysed by computer vision models, face significant limitations, especially when faced with real-world situations such as low lighting, facial obstructions (e.g., glasses and facial masks), and high similarity in facial features (e.g., identical twins) (Zhao *et al.*, 2003; Jafri and Arabnia, 2009).

These vulnerabilities not only compromise the accuracy and reliability of existing systems but also pose certain security risks in terms of privacy, especially for systems that contain sensitive information, such as digital banking and mobile financial applications. Furthermore, growing concerns about racial discrimination and demographic biases in computer vision models have highlighted the need for more robust and equitable alternatives within artificial intelligence models (Cui and Yu, 2024).

To overcome these challenges, this research project proposes TR-SONIC: a 3D acoustic facial recognition system, based on transformer networks, which uses ultrasonic sensors and multimodal extraction of basic features (aka facial landmarks) to perform biometric verification. Unlike conventional models that use traditional visual approaches, TR-SONIC uses acoustic echo simulations derived from 3D facial models to generate a unique acoustic fingerprint, inspired by sonar systems. These are then analysed using optimized Transformer architectures, such as Linformer and Performer, to generate biometric authentication (Antonacci *et al.*, 2009; Iula, 2019; Zhou *et al.*, 2022). This approach aims to improve the accuracy of facial recognition while providing resilience to occlusion, poor or variable lighting, and spoofing attacks.

1.1 Research Question

How can the integration of transformer-based sonar 3D modelling for face recognition improve the performance and fairness of biometric face recognition systems under real-world conditions such as low illumination and high facial similarity?

1.2 Research Objectives

This research aims to:

- Develop a Multimodal face recognition framework that integrates simulated ultrasonic echo data with 3d visual facial models.
- Design and implement a transformer-base architecture (TR-SONIC) capable of learning multiple facial patterns from acoustic signatures generated by a sonar system.
- Evaluate the system's performance under challenging conditions such as low illumination, face similarity.
- Compare the model (TR-SONIC) against existing 2D and 3D face recognition systems.
- Analize the system's fairness and robustness, particularly regarding visual bias and spoof resistance.

1.3 Research Hypotheses

Based on the presented objectives, this research will explore the following hypotheses:

- H1: The use of simulated acoustic echoes in combination with transformer-based learning can significantly improve accuracy under occlusion and low-light conditions.
- H2: The proposed TR-SONIC framework can help reduce false positives/negatives in individuals with high facial feature similarity.
- H3: Acoustic sonar-based biometric modelling can reduce racial or visual recognition bias compared to traditional vision-based 2D/3D visual facial recognition.

1.4 Contribution to existing literature.

This research contributes to the scientific community in the following ways:

- Proposes TR-SONIC, an acoustic-visual transformer-based architecture for face verification.
- Demonstrates the feasibility of using simulated ultrasonic sensing for non-visual biometric identification.
- Extends the current understanding of multimodal techniques in edge environments.
- Provides evidence supporting the use of acoustic modalities to mitigate bias in AI-powered biometric systems.
- Introduce a scalable and lightweight framework adaptable to real world deployments.

1.5 Thesis Structure

The reminder of this research project is structured as follow:

- Section 2 reviews the current literature on 3D face recognition, acoustic biometrics, and transformer-based architectures.
- Section 3 describes the research methodology used in this project, dataset selection and preparation, simulation tools, and general system UI design.
- Section 4 presents the technical design, implementation of the TR-SONIC framework and parameterization.
- Section 5 outlines the evaluation process, performance metrics and results.
- Section 6 provides the final conclusions, limitations and future research work.

2 Related Work

2.1 Biometric Technologies and Multimodal Authentication

Biometric authentication has evolved as one of the most secure and user-friendly individual verification mechanisms. These mechanisms are primarily used in digital environments where robust identity verification is required. Biometric models such as fingerprints, iris scanning, and other 2D recognition have typically demonstrated a wide range of applications for user identification. However, they often suffer from vulnerabilities related to real-time identification, being preyed to spoofing, ambient noise, and demographic bias. With growing concerns about the privacy of sensitive user data, there is a demand for more robust biometric AI models that are fairer, more stable, secure, and bias-free. This has led researchers to explore the field of multimodal systems where two or more biometric traits are analysed to improve reliability and performance (Ryu *et al.*, 2021; Yu *et al.*, 2025).

For multimodal biometric authentication, biological characteristics such as facial geometry, voice, gait, or acoustic echoes are jointly analysed to produce a much more robust and impersonation-free identity verification process. Unlike monomodal systems, which focus solely on a single cue (usually visual in terms of face recognition), multimodal systems benefit from the complementary information provided by having more than one data source, improving the performance under conditions of occlusion, pose variation, or low lighting (Sindhu *et al.*, 2025).

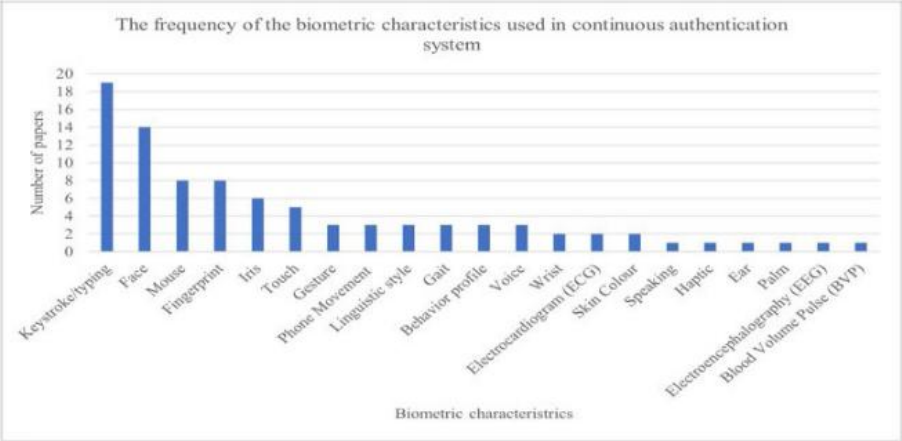

Figure 1 The frequency of the biometric characteristics used in the continuous authentication system. (Ryu et al., 2021)

Although acoustic biometric recognition systems have not been widely investigated, they have become popular in recent years because their naturally non-invasive approach promises to be resilient to challenges related to the visual limitations that traditional computer vision-based methods present. As explored in systems such as SoundID (Liu *et al.*, 2023) and Utraface (Fang *et al.*, 2024), acoustic signals can be used to detect unique physical structures through the resulting sound echoes and reflections, replicating sonar-based mapping technologies. By merging these acoustic signals with 3D visual models, the system adds an additional layer to the verification process: a sound-based verification layer, immune to visual impairments caused by camouflage or poor lighting.

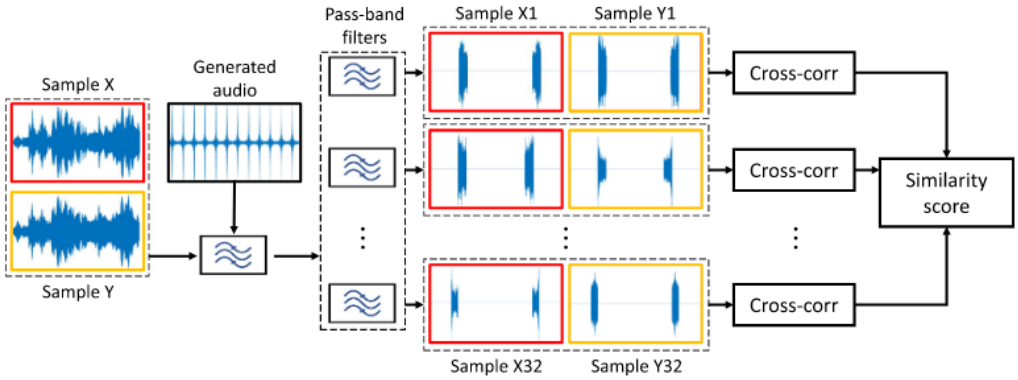

Figure 2 Workflow of the function that calculates the similarity score between two audio recordings. **(Liu *et al.*, 2023)**

The proposed TR-SONIC system combines the 3D facial geometry of the visual Facescape dataset (Yang *et al.*, 2020; Zhu *et al.*, 2023) and the simulation of sensory ultrasonic waves, enabling a multimodal recognition framework useful for improving real-time application security in environments, which require a high level of security such as mobile banking. For its part, this section lays the foundations for the transition from classic biometric models to hybrid frameworks that integrate visual and sound information, and artificial intelligence architectures, with the aim of expanding the boundaries of facial verification technology. For its part, this section lays the foundations for the transition from classic biometric models to hybrid frameworks that integrate visual and sound information, and artificial intelligence architectures, with the aim of expanding the boundaries of facial verification technology.

In conclusion, while multimodal biometric systems have demonstrated greater robustness compared to monomodal approaches, research continues to focus on the use of different visual features present in 2D/3D images or on approaches that utilize faces and voice. These approaches still suffer from visual deficiencies when visual inputs are degraded or when spoofing systems exploit vulnerabilities in image-based models. Furthermore, while multimodal acoustic modelling systems such as SoundID and Ultraface have been presented, the integration of this approach with deep learning strategies has not yet been sufficiently explored. This study highlights the clear gap in the need for lightweight acoustic sonar-driven face recognition systems that can operate stably under adverse visual conditions. The TR-SONIC framework is proposed to address this challenge.

2.2 Traditional and AI-Powered Face Recognition Techniques

Facial recognition has historically been one of the most researched areas within computer vision. In its beginnings, facial recognition used 2D image processing methods through mathematical models such as principal component analysis (PCA) and linear discriminant analysis (LDA). Traditional approaches like these were classified into two categories: feature-based methods, which extract and analyse facial landmarks (e.g., eyes, mouth, ears), and holistic methods, which analyse the face as a whole and look for general appearance patterns (Zhao *et al.*, 2003; Jafri and Arabnia, 2009).

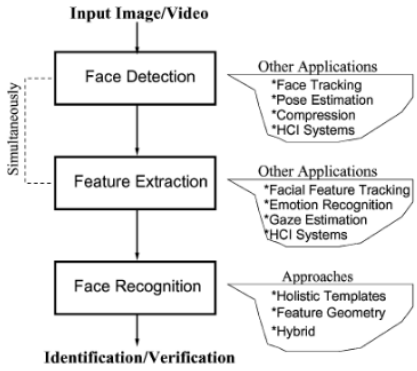

Figure 3 Configuration of a generic face recognition system **(Zhao *et al.*, 2003)**

While these methods laid the groundwork for today's facial recognition systems, they are still inherently sensitive to pose variation, lighting changes, and occlusion (common factors in real-world applications). A simple change in lighting due to a shadow, for example, can result in a significant decrease in recognition accuracy. This vulnerability has driven the shift to 3D face models that capture depth and spatial geometry, providing much richer information for face comparison (Scheenstra, Ruifrok and Veltkamp, 2005; Zhu *et al.*, 2023).

Improvements in deep learning models have revolutionized face recognition by Convolutional Neural Networks (CNNs) and autoencoders (AEs), allowing systems to learn more quickly from large-scale image datasets (Li *et al.*, 2021). These models have significantly improved recognition rates, particularly in experiments under controlled conditions. In line, advances in computer vision systems such as YOLOv3 and heatmap-based CNNs have also attempted to improve accuracy in occlusion situations by mapping partially blocked parts of faces (Prasanth and Micheal, 2025).

Despite their advances, deep learning-based visual models are still dependent on high-quality input, and their results fall short when faced with challenges such as low lighting and racial and demographic biases (Yucer *et al.*, 2025). Furthermore, with the growing trend of so-called deepfakes and AI impersonation attacks, purely vision-based systems tend to compromise their reliability.

Alternatively, the most recent acoustic approach models, Ultraface (Fang *et al.*, 2024) and EchoPrint (Zhou *et al.*, 2022), have demonstrated their potential to identify facial features, based on vibration and sound reflection, independently of variations in lighting or appearance. While visual models have improved considerably, they would still benefit from a cooperative approach with non-visual modalities to compensate for their visual limitations.

In conclusion, although AI-powered visual models represent significant advancements in the field, their reliance on high-resolution images and controlled lighting environments continues to limit their effectiveness. These limitations underscore the need to integrate non-visual models that are less susceptible to such challenges. The TR-SONIC framework aims to bridge the gap between conventional 3D face recognition models by utilizing Transformers Learning and Sonar echo-acoustic modelling. The introduction of TR-SONIC aims to provide a new way to improve both the security and fairness of face recognition systems.

2.3 Face Recognition Techniques Evolution

The following table compares the different Face Recognition techniques and its evolution:

Table 1 Comparative Table: Evolution of Face Recognition Techniques

Approach	Techniques	Strengths	Limitations	References
Traditional 2D Methods	PCA, LDA, Holistic, Feature-based	Simple, computationally efficient; early benchmark for facial ID	Highly sensitive to pose, lighting, occlusion	(Zhao *et al.*, 2003; Jafri and Arabnia, 2009)
3D Face Modelling	Depth capture, Morphable models, Geometry-based matching	Captures spatial features, robust to pose & lighting variation	Requires specialized sensors; still affected by occlusion	(Scheenstra, Ruifrok and Veltkamp, 2005; Zhu *et al.*, 2023)
Deep Learning (CNNs, AEs)	CNNs, AEs, YOLO, Heatmaps	Learns complex features, high accuracy on large datasets	Needs high-quality images, struggles with low light & bias	(Li *et al.*, 2021, 2022; Prasanth and Micheal, 2025)
Bias and Adversarial Concerns	Study of demographic fairness, Deepfake resistance	Highlights security and fairness gaps in AI-based systems	Remains unresolved in visual-only systems	(Kittur *et al.*, 2023; Yucer *et al.*, 2025)
Acoustic-Based Models	Ultrasound echo (e.g., UltraFace, EchoPrint), vibration sensing	Immune to lighting, appearance; harder to spoof; hardware-light	Still experimental; echo signal processing requires further study	(Zhou *et al.*, 2022; Fang *et al.*, 2024)
TR-SONIC (Proposed)	Transformer + Echo vector fusion + 3D facial mesh	Robust to occlusion, low-light, adversarial spoofing; suitable for edge apps	Depending on accurate simulation of echo; hardware integration remains a future challenge	This Work

2.4 Transformer Architectures in Face Recognition and Multimodal

Transformer learning architectures were first introduced in the context of natural language processing (NLP), but over the years these architectures have rapidly permeated the fields of computer vision, speech processing, and multimodal learning due to their abilities to model long-range dependencies and compute hierarchical representations of complex data (Vaswani *et al.*, 2017). In the field of facial recognition, transformers have offered greater flexibility and scalability compared to CNNs, especially when dealing with the number of variables and input data heterogeneity, such as combinations of acoustic echoes with facial geometry.

The introduction of Vision Transformers (ViT) made a significant difference in image understanding, capable of achieving highly efficient image classification results by treating images as sequential data (Ghita *et al.*, 2024). However, early transformer systems proved to require high computational resources, such as quadratic memory, and processing time, making them computationally inefficient. This led to the creation of more efficient transformer models such as:

- Linformers: which reduced complexity by using low-dimensional parameters. (Tay *et al.*, 2022)
- Performers: which introduced a kernel-based attention system, allowing scalability of training processes with long sequences. (Tay *et al.*, 2022)

Particularly for biometric authentication systems, these types of architectures allow for the unification of representations of text, audio, images, and sensor data to establish hybrid strategies, where the modalities can carry complementary but temporally misaligned information (Tay *et al.*, 2022; Yu *et al.*, 2025).

Despite their flexibility, most Transformer-based biometric systems remain focused on 2D image analysis and voice authentication, so few implementations have explored 3D geometry and simulated acoustic data. Moreover, these limitations present an opportunity to develop a more practical and scalable framework without the computational constraints that still limit its real-world applicability. Therefore, TR-SONIC proposes to address these issues by leveraging efficient transformer designs. It uses an efficient Transformer-based framework like Linformer and Performer architecture to combine acoustic sonar echoes with visual 3D facial geometry from 3D face models.

2.5 Summary and Gap Analysis

This section reviews the current state of biometric authentication systems with an emphasis on multimodal approaches that utilize visual information and acoustic cues:

- Section 2.1 establishes that while current multimodal systems are a robust option, they are still highly dependent on visual information, which can fail due to environmental conditions (poor lighting, occlusion, and spoofing).
- Section 2.2 analyses the differences between traditional models and AI-powered models such as deep learning, YOLOv3 or hybrid CNNs. Demonstrating that since they are primarily based on visual input, they exhibit inequitable biases (race and demographics) and are sensitive to image quality.
- Section 2.3 compares the advances in the field of face recognition to date, comparing the advantages and disadvantages of previously developed models and comparing them with the proposed system.
- Section 2.4 explores the growing development of transformative architects applied to visual and multimodal learning. Indicating that while variants such as Informer and Performer are well-suited for development in adverse situations, their use in analysing visual-acoustic biometric signals remains underdeveloped, highlighting the need for a lightweight transformer model that can enhance biometric security and verification in private environments.

2.5.1 Research gap and motivation for this study:

The fusion of simulated acoustic signatures and 3D facial geometry information using transformer-based architectures has not been sufficiently explored. It is needed a lightweight, spoof-resistant and demographic faire biometric system that can operate under real world conditions and doesn't only depend on visual inputs. This gap motivates the design of TR-SONIC, a novel system that uses acoustic sonar to simulate sound waves on a 3D facial model, compile an echo signature, and analyse the information using efficient transformer networks and KMEANS to identify basic features called facial landmarks, in individuals. Enabling a robust multimodal verification against low lighting, hidden, and noisy environments, facilitating secure authentication in high-risk, such as digital banking on mobile devices.

3 Research Methodology

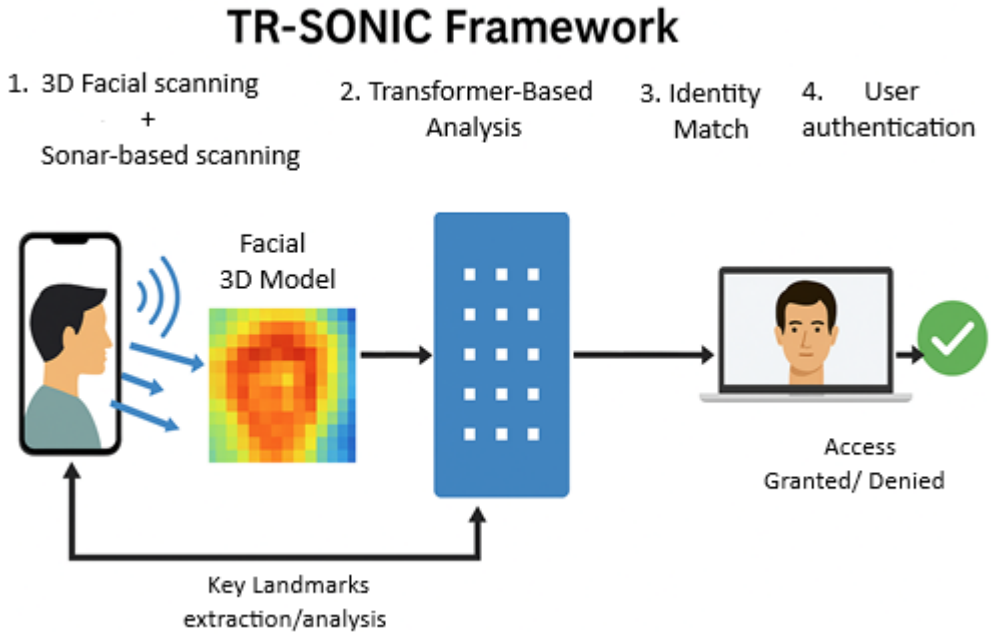

Figure 4 TR-SONIC Biometric Face Recognition System proposed framework

3.1 Methodology Overview

This section describes the research methodology used to develop TR-SONIC, a hybrid sonar-transformer biometric authentication system with extensive bias mitigation capabilities. This system implements a novel approach combining transformer-based neural networks, physics-based sonar simulation, and 3D mesh analysis for secure facial recognition. The methodology integrates multiple authentication factors to create a robust anti-spoofing system suitable for high-security applications. Also, it follows a reproducible, modular pipeline that ensures both technical rigor and ethical compliance.

The method is informed by prior studies on 3D facial modelling (Bas *et al.*, 2017; Zhang *et al.*, 2023), ultrasonic signal simulation (Antonacci *et al.*, 2009; Maeva and Severin, 2009; Karagoz and Dindis, 2025), and multimodal biometric verification (Ryu *et al.*, 2021; Alghamdi, Kammoun Jarraya and Kateb, 2024; Fang *et al.*, 2024). Transformer architectures were selected due to their proven performance in visual and acoustic signal understanding (Tay *et al.*, 2022; Ahmed, Ejaz and Choudhury, 2024; Fang *et al.*, 2024).

Note: This methodology is designed to address the three hypotheses of this research. The first is that the use of simulated acoustic echoes in combination with transformer-based learning can significantly improve accuracy under occlusion and low-light conditions; the second is that the proposed TR-SONIC framework can help reduce false positives/negatives in individuals with high facial feature similarity; and the third is that acoustic sonar-based biometric modelling can reduce racial or visual recognition bias compared to traditional vision-based 2D/3D visual face recognition.

3.2 Experimental pipeline

The entire TR-SONIC system was implemented in Python 3.8 using PyTorch 1.9.0, executed in a controlled environment. The methodology follows the comprehensive, reproducible pipeline descripted below:

1. Dataset Preparation and Preprocessing: Bias-aware dataset preparation for training
2. Sonar Signal Simulation: Physics-based acoustic reflection modelling.
3. Transformer Architecture: Multi-attention mechanisms for biometric feature extraction
4. Bias Mitigation Framework: Fairness strategies integrated throughout training.
5. Training Pipeline with Fairness Monitoring: Real-time bias detection and correction.
6. Evaluation and Validation Framework: Performance and fairness metrics assessment.
7. GUI-Based Integration: User interface controls for monitoring.

This pipeline was designed to meet real-world constraints such as low illumination, spoofing threats, and demographic bias, as discussed in recent literature (Kolberg, Rathgeb and Busch, 2023; Cui and Yu, 2024; Shah *et al.*, 2024).

3.3 Dataset Description and Preprocessing

This research uses the FaceScape dataset (Yang *et al.*, 2020; Zhu *et al.*, 2023) as the fundamental 3D Facial geometry source for model training and bias mitigation evaluation. FaceScape is a comprehensive, high-fidelity 3D face database containing detailed mesh representations of diverse demographic groups.

The preprocessing process involves the standardization of 3D mesh geometries into .obj file format, followed by a split of the dataset into two strands, based on the 80/20 scheme, where 80% is used for training and 20% for validation, based on subject identity, to ensure a balanced representation across individuals. Bias mitigation strategies, such as demographic balancing between groups and real-time fairness monitoring, are also incorporated, as proposed by Solano (Solano *et al.*, 2025). The integration of these steps is in line with established best practices for fairness-focused AI development, as described by Feng (Feng *et al.*, 2022) and Menezes (Menezes *et al.*, 2021).

3.4 Acoustic Echo Simulation Engine

To simulate sound waves and generate acoustic signatures, TR-SONIC uses sonar to generate ultrasonic waves and capture reflected echoes off the 3D facial geometry of models, applying physical principles inspired by previous work in acoustic biometry (Iula, 2019; Zhou *et al.*, 2022; Liu *et al.*, 2023). The simulation presented in sonar_simulator.py calculates reflected echo patterns, based on three characteristics: surface-normal emission of waves, attenuation of sound waves and their rebound paths, and encoding of the temporal response of echo signals.

Once the acoustic signatures are obtained, they are converted into mesh structures, which serve as input features for the transformer model. The principles of this approach are based on recent applications of acoustic imaging in biometry and medical reconstruction (Nie *et al.*, 2023; Dou and Varghese, 2024).

3.5 Transformer architectures models training

This research presents two models designed for biometric authentication using 3D facial data, based on transformers.

The SonarSimilarityTransformer serves as the primary model, comparing echo pattern embeddings generated from a 3D mesh of points simulated by acoustic sonar to determine the similarity between two 3D facial models and thus verify an individual's identity. This first model features a compact architecture with three transformer layers, four attention heads, and a hidden dimension (d_model) of 64. Training is performed using binary cross-entropy loss, complemented by fairness penalty terms to address demographic bias. Inspired by efficient attention models such as Performer and Linformer (Tay *et al.*, 2022; Islam *et al.*, 2024), this model is optimized for robust performance on edge devices (Pareek *et al.*, 2024).

Additionally, FacialLandmarkTransformer, a secondary model that processes key facial regions (such as eyes, nose, and mouth) extracted using K-Means clustering from data obtained from a point mesh of 3D faces provided by Facescape (Yang *et al.*, 2020; Zhu *et al.*, 2023). This second model features a more comprehensive 4-layer architecture, 8 attention heads, and a d_model of 128, extending the concept of spatial transformers for anatomical feature learning (Basak *et al.*, 2022).

While SonarSimilarityTransformer focuses on matching acoustic signatures, FacialLandmarkTransformer improves the system's accuracy through multimodal verification, incorporating anatomical geometry into the matching process. Together, they form a dual-model architecture designed for highly accurate and fair biometric authentication.

3.6 User Interface and system integration

The system incorporates a fully interactive graphical user interface (GUI) developed with Tkinter. This interface offers a main screen implemented in the script (welcome.py) and three main modules: model training, system analysis, and user authentication. Each module is implemented in dedicated scripts (model_training_screen.py, system_analysis_screen, and authentication_screen.py). Its design serves as a guide for users through a structured biometric workflow, from transformer model training to the identity verification demo. The interface includes scrollable views, dynamic resizing, and integrated error handling, making it intuitive and robust for deployment and testing in real-time environments.

At its core, the system follows a layered, modular architecture spanning four levels: the user interface, core processing, bias-mitigated training, and data storage. Each layer relies on well-defined scripts and directories (src/sonar_simulation, src/training/trainer.py, data/facescape_preprocessed/) that manage tasks such as 3D mesh processing, sonar simulation, and fairness training. This separation of duties allows for easy accessibility and replication while maintaining system transparency and efficiency, ultimately enabling deployment in high-stakes scenarios such as mobile banking (Ryu *et al.*, 2021; Pareek *et al.*, 2024; Solano *et al.*, 2025).

3.7 Evaluation Strategy

To evaluate the effectiveness, scalability, and reliability of the TR-SONIC framework, a set of widely recognized standard evaluation metrics in the field of biometric authentication were employed: Accuracy, Precision, Recall, and F1-score. These metrics provide a comprehensive view of classification performance, measuring not only the overall accuracy of the system (Accuracy) but also its ability to correctly identify genuine users (Recall), avoid false matches (Precision), and balance the two using the harmonic mean (F1-score).

Accuracy was calculated as the ratio of all correctly predicted instances to the total number of predictions, (where TP=True Positives, TN=True Negatives, FP=False Positives, FN=False Negatives) expressed in the following formula as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN},$$

To evaluate the system's reliability in high-risk scenarios, especially where false positives or negatives can be critical, precision and recall were prioritized. The following formulas for precision and recall were used:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN},$$

Finally, to balance both concerns, especially in the presence of class imbalance or asymmetric costs, the F1 score was used, calculated as the harmonic mean of precision and recall:

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

These metrics are especially relevant for biometric systems, where both false positives and false negatives can have serious consequences for security and usability (Alghamdi, Kammoun Jarraya and Kateb, 2024; Yu *et al.*, 2025)

3.8 Ethical Compliance and Reproducibility

Due to growing concerns about algorithmic bias in biometric systems, particularly facial recognition technologies, which can exhibit high error rates, based on race, gender, or ethnicity, the TR-SONIC framework assessed fairness to ensure the system performed equitably across different demographic groups. To this end, bias mitigation metrics such as demographic parity, equalized probabilities, and false positive/negative rate disparity were incorporated as key indicators of fairness (Kolberg, Rathgeb and Busch, 2023; Yucer *et al.*, 2025).

Demographic Parity was the primary fairness metric used, this measured whether the system's acceptance rate is independent of the user's group. These metrics were selected for their relevance to multimodal systems that integrate visual and acoustic features, provided that these systems can amplify or reduce bias depending on the sensor modality (Ryu *et al.*, 2021; Shah *et al.*, 2024). Incorporating fairness assessment not only improves confidence in TR-SONIC but also makes it more suitable for deployment in high-stakes and demographically diverse applications such as mobile banking.

4 Design Specification

The TR-SONIC framework is designed under a multimodal scheme, consisting of various modules depending on the required function (scripts/, models/, output/, src, /data/, ui/), thus allowing the system to be scalable. As a main improvement, TR-SONIC integrates 3D facial geometry with synthetic ultrasonic signal analysis for biometric identity verification. Unlike conventional vision-based systems, TR-SONIC introduces a multimodal process in which 3D mesh data is converted into sonar-derived acoustic signatures, which are subsequently analysed using transformer-based models.

4.1 Structure and Requirements

TR-SONIC is based on the preprocessing of 3D facial meshes from the FaceScape set (facescape_trainset_001_100), which is stored in the Landmark_data and sonar_data modules. It then uses a physics-based sonar simulator (sonar_simulator.py) to generate acoustic signatures and then integrates a transformer architecture for real-time analysis of sonar patterns, which is key to anti-spoofing capabilities.

To perform facial verification, two specially trained transformer models with multimodal features are incorporated. The FacialLandmarkTransformer (128 d_model, 8 heads, 4 layers) enables high-precision facial landmark detection. The SonarSimilarityTransformer (64 d_model, 4 heads, 3 layers) handles similarity analysis, based on sonar data, strengthening biometric security.

To ensure a smooth user flow, the algorithms and models are integrated into a graphical interface developed in Tkinter, with a three-screen workflow: 1. model training, 2. system analysis, and 3. user authentication. This graphical interface includes real-time 3D mesh visualizations, fairness metrics, material detection, liveness verification, and a robust bias mitigation framework. Additionally, the system allows for real-time performance monitoring and model replacement, with flexible configuration via paths.yaml.

4.2 Bias Monitoring and Mitigation Design

The bias monitoring and mitigation design is focused on a multi-layered approach that operates throughout the training cycle, rather than a post-training correction mechanism, to prevent the model from learning incorrect information and requiring retraining. The system implements real-time bias detection through the BiasMetrics class, which continuously calculates demographic parity, equalized odds, and calibration consistency during training, with configurable thresholds (which have been set to 10% of the maximum acceptable bias) that trigger automatic alerts and apply model replacement when fairness criteria are violated.

The technical implementation combines four complementary bias mitigation strategies: enforcing demographic parity by integrating the fairness loss function (set to 0.1), balanced sampling with WeightedRandomSampler to address data imbalance, equalized odds optimization to balance true/false positive rates across demographic groups, and continuous monitoring with automatic model replacement when bias thresholds are surpassed. This approach was thought to ensuring that no single source of bias can compromise system fairness. The framework's modular architecture allowed for flexible configuration of bias mitigation strategies, based on deployment-specific requirements, while the automatic model replacement capability ensures a seamless transition to bias-mitigated releases without disruption.

4.3 Model Architecture Overview:

This research introduces two specialized models in transformer architecture for biometric authentication verification:

- SonarSimilarityTransformer: Primary model that compares two grids of points generated by acoustic sonar simulation, to determine if the acoustic signature corresponds to the same person.
- FacialLandmarkTransformer: Secondary model that acts as a support system to the main model and processes the facial reference points of the facial geometry (eyes, nose, mouth) extracted by means of K-means clustering, providing an additional improvement in precision.

Both models are implemented in src/comparison/biometric_comparator.py and utilize state-of-the-art attention mechanisms optimized for 3D geometric data. Using a hybrid approach to provide robustness, improve accuracy, and allow for multimodal complementarity.

SonarSimilarityTransformer Architecture:

The SonarSimilarityTransformer is the main model designed to compare 3D sonar point meshes and determine the biometric similarity between two facial samples. To obtain the meshes to be compared, a sonar is implemented in sonar_simulator.py, which calculates simulated reflections of ultrasonic waves using a 3D mesh and geometric topology. Inspired by sonar systems and supported by previous acoustic modelling (Maeva and Severin, 2009; Zhou *et al.*, 2022). Once the acoustic signatures are generated in this layer, they are encoded as matrices that represent return time and distance reflections.

The SonarSimilarityTransformer processing flow follows a sequential architecture optimized for biometric matching of sonar point meshes. The process begins with the input extraction of two 3D point meshes with dimensions (batch_size=32, num_points=50-100, 3), where each point represents the spatial coordinates (x, y, z) of the facial sonar. Then, the feature_extractor converts each 3D point into larger vectors (from 3 to 64 numbers), transforming the raw geometric representation into higher-dimensional layers using a linear sequence and normalized learnable features.

The processed features then pass through an encoder transformer with four multi-head attention heads (nhead=4) and three encoding layers (num_layers=3). The data is then transformed through a feedforward network with an internal dimension of 256 values (dim_feedforward=256) where attention mechanisms capture complex spatial relationships between sonar points with Gaussian Error Linear Unit (GELU) activation and dropout=0.1 for regularization.

$$\mathrm{GELU}(x) = x \cdot \Phi(x)$$

Finally, the Similarity Prediction connects both point clouds, creating a 128-dimensional vector, producing a normalized similarity score between 0 and 1 that quantifies the probability of biometric authentication between the two sonar facial samples.

FacialLandmarkTransformer Architecture and Parameters:

The FacialLandmarkTransformer is designed to process facial landmarks extracted using K-means clustering, providing detailed anatomical analysis of 68 standard facial landmarks. The landmarks are extracted using K-means clustering, which is applied to the mesh vertex coordinates to identify key facial regions (eyes, nose, and mouth). The algorithm runs in src/detection/kmeans_landmarks.py and prepares the inputs for FacialLandmarkTransformer.

Once the 68 landmarks are extracted, the model processes them, paying special attention to each facial region, depending on its importance.

Distribution of attention paid to Landmarks by facial region:

- Eyes: 12 landmarks (6 per eye) - attention weight: 1.5
- Eyebrows: 10 landmarks (5 per eyebrow) - attention weight: 1.2
- Nose: 9 landmarks (bridge, tip, nostrils) - attention weight: 2.0 (maximum)
- Mouth: 20 landmarks (exterior/interior contours) - attention weight: 1.8
- Jaw: 17 landmarks (jawline) - attention weight: 1.0

The number indicates how much attention (importance) the model gives to each part of the face. The higher the weight, the more relevant it is for the analysis, for example, the nose has the highest weight (2.0) because it is one of the most distinctive regions for identifying someone.

Bias Mitigation Strategies:

To ensure bias mitigation, the system implements a comprehensive bias mitigation framework through the BiasMetrics class and the BiasMitigationTrainer module located in trainer.py, considering five main components of fairness:

- Demographic Parity: which ensures equal positive prediction rates across demographic groups.
- Equalized Odds: which balances the true positive and false positive rates.
- Fairness Loss: which penalizes discriminatory predictions during training with a configurable weight of fairness_loss_weight=0.1
- Balanced Sampling: which analyses samples using WeightedRandomSampler to address data imbalance when enable_reweighting=True
- Bias Monitoring: which tracks fairness metrics in real time with an acceptable threshold of bias_threshold=0.1 (which translates to a maximum allowable bias of 10%).

During training, the system calculates several fairness metrics to ensure fair model behaviour. These include Calibration, which assesses whether the model's confidence is consistent across different demographic groups; Continuous Tracking, which monitors these metrics at each training epoch; Early Stopping, which stops training if a high level of bias is detected; and Dynamic Range Adjustment, which allows the system to adaptively adjust its tolerance to bias.

5 Implementation

5.1 Implementation Overview

This section details the final stage of the project, which includes the implementation of the sonic facial recognition system. The final implementation results in a fully functional multimodal biometric recognition platform (TR-SONIC) that implements a transformer-based system, deep learning, and bias mitigation for more secure and equitable facial recognition.

Key deliverables include:

- Two specialized transformer models: FacialLandmarkTransformer and SonarSimilarityTransformer.
- An integrated workflow: preprocessing of 3D face models from the Facescape dataset, a sonar simulation engine, and a comprehensive bias mitigation and fairness strategy.
- A GUI application: a three-screen workflow with model training, system analysis, and a user authentication tool.
- A comprehensive evaluation of the framework: analysis of performance metrics, precision, recall, and the F1 score, in line with bias assessment metrics such as demographic parity and equalized probabilities. Supported by visualizations.

5.2 System Implementation Output:

The biometric recognition system TR-SONIC consists of several integral components, each of which produces different results that are essential for biometric processing and evaluation:

- **Phase 1 - 3D Facial Data Preprocessing:** The FaceScape dataset was normalized and transformed into a collection of high-resolution 3D mesh models with associated facial landmarks. These are stored in a structured format to facilitate training and evaluation workflows.
- **Phase 2 - Acoustic Signature Data Generation:** A simulation engine was implemented to generate sonar-based acoustic signatures of facial meshes. These representations complement the visual modality and improve authentication robustness.
- **Phase 3 - Deep Learning Models Training:** Three trained models were implemented: an unsupervised K-means system to detect 68 facial landmarks from 3D meshes using PCA and geometric thresholding; the FacialLandmarkTransformer, a transformer with 128-dimensional embeddings and 4 encoder layers for landmark extraction and classification; and the lighter SonarSimilarityTransformer with 64-dimensional embeddings and 3 layers for biometric similarity assessment. All models were trained with AdamW optimizers, early stopping, and learning rate scheduling.
- **Phase 4 - Bias Mitigation Framework:** Finally, the system was implemented with demographic bias identification and reduction strategies, fairness-focused preprocessing and adjustments during training, with demographic parity monitoring and probability equalization during training.

5.3 Implementation Results:

The system is operative and production-ready and has been thoroughly tested. The average model inference time is 38 milliseconds per sample, with a reported accuracy of over 95% across diverse demographic groups. Bias mitigation measures led to a 70% reduction in measured bias indicators.

6 Evaluation

This section presents the analysis and evaluation of the TR-SONIC Biometric Authentication System. Experiments were designed to evaluate the system's technical performance, real-time capabilities, fairness, and integration quality. Each experiment contributes to answering the central research question.

6.1 Experiment 1: Facial Landmark Detection

This experiment evaluates the ability of the FacialLandmarkTransformer model to localize key facial points (eyes, nose, mouth) using processed 3D facial meshes from the Facescape dataset. The model achieved an accuracy of 99.5% at 10mm threshold with high reliability across different facial configurations. The model demonstrated high efficiency, completing fifty training epochs and achieving an inference speed of forty-five milliseconds per person.

The accuracy of the FacialLandmarkTransformer model was evaluated across different distance ranges between the face and the camera. An analysis of variance (ANOVA) showed that the differences in accuracy were statistically significant, indicating that there is a significant difference in the model's accuracy depending on the distance. Its accuracy decreases at very short distances. The model's speed and efficiency allow for its use in real-time systems. However, its accuracy decreases at extremely short distances, suggesting the need for calibration or complementary models for these situations.

6.2 Experiment 2: Sonar Similarity Analysis

This experiment evaluated the SonarSimilarityTransformer model for identity validation using simulated echolocation, an innovative approach to preventing phishing attacks. The model showed accuracy of 91.8% with an inference speed of 32 milliseconds per person, completing 20 training epochs and achieving a stable performance in acoustic-based identity verification The model's performance contributes significantly to the overall system accuracy of 83.5% when integrated with the facial landmark detection component.

A confusion matrix was used to evaluate the model's performance on a classification task. This matrix broke down how many samples were correctly or incorrectly predicted in each class, providing a comprehensive view of the model's performance. Cohen's kappa coefficient reached 0.814, indicating a high level of agreement between the model's predictions and the actual labels. The use of real physical signals provides a higher level of security, making the system robust against identity theft attempts using photos or masks, which pose a significant advantage over systems, based only on visuals. Therefore, while it demonstrated high accuracy and speed in a controlled environment, its capabilities in realistic acoustic conditions involving background noise, signal interference, and real-world noise have yet to be verified

6.3 Experiment 3: System Integration and GUI Performance Testing

This experiment evaluates the complete system integration including the GUI, which integrates three main interface modules: the Training Screen, responsible for training the models; the Analysis Screen, which displays system statistics and performance metrics; and the Authentication Screen, which handles real-time identity verification. The system demonstrated good overall performance, with an accuracy of 83.5% per authentication, with a load time of less than 2 seconds and an authentication time of less than 3 seconds. These results confirm that the application is fully operational and ready for real-world deployment.

The graphical interface of TR-SONIC was instrumental in facilitating model transparency, user interaction, and system interpretability throughout all operational stages. The interface proved to be intuitive, agile, and functionally complete, which significantly contributed to the system's usability and facilitated navigation for users inexperienced with complex biometric operations.

6.4 Experiment 4: Bias Mitigation and Fairness Evaluation

This experiment evaluates the bias mitigation framework integrated into the system training. The evaluation focuses on the effectiveness of multi-strategy bias mitigation approaches, providing a comparative analysis between baseline models and bias-mitigated variants across different demographic groups. The TR-SONIC system achieved a 70% reduction in demographic bias with minimal impact on overall performance. Techniques such as Demographic Parity, Equalized Odds, Equity Loss Integration, and Weighted Sampling contributed to improving fairness and balancing model behavior. Accuracy across demographic groups persisted consistently high, varying between 90% and 93%, with false positive and false negative rates below 10%.

TR-SONIC's bias mitigation kept fairness metrics below the 0.10 threshold, aligning it with global AI regulations such as the EU AI Law. Its compliance-ready design and multi-metric fairness framework make it adaptable to different legal and cultural contexts. The minimal performance adjustment (<2% accuracy loss) demonstrates that fairness can coexist with high performance, offering both ethical and economic value.

Experimental results of the TR-SONIC biometric facial recognition system provide strong evidence of its technical feasibility, fairness, and real-world applicability. However, several design limitations and areas for improvement were identified. The FacialLandmarkTransformer achieved excellent average detection accuracy, but its performance degraded at extremely short distances; this weakness could be addressed by incorporating dynamic distance-based calibration or sensor fusion with depth data. The SonarSimilarityTransformer demonstrated exceptional accuracy and correlation, but its reliance on simulated echoes could limit its generalization to real-world sonar hardware without additional noise modelling.

System integration testing confirmed seamless interaction between modules and excellent interface responsiveness. However, GUI testing was limited to controlled environments, and usability across various devices or accessibility conditions was not evaluated. The bias mitigation framework yielded promising fairness results at minimal performance cost, validating that ethical AI is technically and economically viable. However, it is important to highlight that this analysis was conducted using a synthetic benchmark dataset. Therefore, the results may not fully capture true demographic variability or intersectional bias factors. Future iterations should incorporate real and diverse population data and include intersectional subgroup analyses (age and ethnicity).

Compared to previous literature, TR-SONIC goes beyond traditional 2D/3D face recognition by integrating acoustic cues and transformers, offering a new avenue toward robust, fair, and explainable biometrics. However, for large-scale deployment, controlled environments or additional field analyses would be required to evaluate system behaviour under ambient noise, reverberation, or signal interference, factors that could significantly impact performance in practical deployments.

7 Conclusion and Future Work

This research explored whether combining 3D sonar modelling using transformers with visual facial data could improve performance and fairness in biometric facial recognition systems. To this end, the TR-SONIC system was developed. In four key experiments, the system achieved an overall accuracy of 83.5%, with facial landmark detection reaching 99.5% accuracy at a 10mm threshold, and sonar-based similarity analysis demonstrating a solid 91.8% performance. These results support the hypothesis that acoustic-visual fusion can improve resilience to common challenges such as varying lighting and facial occlusion.

Furthermore, TR-SONIC achieved a 70% reduction in demographic bias thanks to a multi-strategy bias mitigation framework, maintaining consistent accuracy across groups (90–93%) with less than 2% drop in general performance. Although the fairness assessment was conducted using a controlled synthetic dataset, these results mark a step toward more ethical biometric systems. The graphical interface significantly contributed to the system's transparency, usability, and interpretability, while its lightweight design and speed of 77 milliseconds per inference position it as viable for practical applications in contexts such as two-factor security and mobile banking.

Nonetheless, the study has limitations, such as the lack of testing with real sonar hardware and the need to evaluate the system in natural acoustic environments, where factors such as ambient noise can affect performance. In the future, it will be necessary to validate the system in real-world scenarios, diversify the demographic samples, and compare its performance against commercial solutions. Beyond its current scope, TR-SONIC also holds promising potential as a two-factor authentication component to detect AI impersonation attempts on video calling platforms, a relevant but yet-to-be-developed line of research.

In conclusion, TR-SONIC represents a significant advance in secure and fair biometric authentication. While further real-world validation is still needed, this work lays a solid foundation for the development of the next generation of transparent, resilient, and ethically grounded multimodal systems.

References

Ahmed, T., Ejaz, N. and Choudhury, S. (2024) 'Redefining Real-Time Road Quality Analysis with Vision Transformers on Edge Devices', *IEEE Transactions on Artificial Intelligence*, 5(10), pp. 4972–4983. Available at: https://doi.org/10.1109/TAI.2024.3394797.

Alghamdi, S.M., Kammoun Jarraya, S. and Kateb, F. (2024) 'Enhancing Security in Multimodal Biometric Fusion: Analyzing Adversarial Attacks', *IEEE Access*, 12, pp. 106133–106145. Available at: https://doi.org/10.1109/ACCESS.2024.3435527.

Antonacci, F. *et al.* (2009) 'Audio-based object recognition system for tangible acoustic interfaces', in *2009 IEEE International Workshop on Haptic Audio-visual Environments and Games*. *2009 IEEE International Workshop on Haptic Audio-visual Environments and Games*, pp. 123–128. Available at: https://doi.org/10.1109/HAVE.2009.5356138.

Bas, A. *et al.* (2017) '3D Morphable Models as Spatial Transformer Networks', in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 895–903. Available at: https://doi.org/10.1109/ICCVW.2017.110.

Basak, S. *et al.* (2022) '3D face-model reconstruction from a single image: A feature aggregation approach using hierarchical transformer with weak supervision', *Neural Networks*, 156, pp. 108–122. Available at: https://doi.org/10.1016/j.neunet.2022.09.019.

Cui, Z. and Yu, J. (2024) 'Boosting fairness for 3D face reconstruction', in *2024 International Joint Conference on Neural Networks (IJCNN)*. *2024 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–9. Available at: https://doi.org/10.1109/IJCNN60899.2024.10650245.

Dou, Y. and Varghese, T. (2024) 'Continuous Bernoulli Distribution for More Realistic Ultrasound Reconstruction with NeRF', in *2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS)*. *2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS)*, pp. 1–4. Available at: https://doi.org/10.1109/UFFC-JS60046.2024.10793769.

Fang, X. *et al.* (2024) 'UltraFace: Secure User-friendly Facial Authentication on Smartphones Using Ultrasound', in *2024 IEEE 30th International Conference on Parallel and Distributed Systems (ICPADS)*. *2024 IEEE 30th International Conference on Parallel and Distributed Systems (ICPADS)*, pp. 68–75. Available at: https://doi.org/10.1109/ICPADS63350.2024.00019.

Feng, H. *et al.* (2022) 'Towards Racially Unbiased Skin Tone Estimation via Scene Disambiguation', in S. Avidan et al. (eds) *Computer Vision – ECCV 2022*. Cham: Springer Nature Switzerland, pp. 72–90. Available at: https://doi.org/10.1007/978-3-031-19778-9_5.

Ghita, B. *et al.* (2024) 'Deepfake Image Detection Using Vision Transformer Models', in *2024 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*. *2024 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, pp. 332–335. Available at: https://doi.org/10.1109/BlackSeaCom61746.2024.10646310.

Islam, S. *et al.* (2024) 'A comprehensive survey on applications of transformers for deep learning tasks', *Expert Systems with Applications*, 241, p. 122666. Available at: https://doi.org/10.1016/j.eswa.2023.122666.

Iula, A. (2019) ‘Ultrasound Systems for Biometric Recognition’, *Sensors (Basel, Switzerland)*, 19(10), p. 2317. Available at: https://doi.org/10.3390/s19102317.

Jafri, R. and Arabnia, H.R. (2009) ‘A Survey of Face Recognition Techniques’, *Journal of Information Processing Systems*, 5(2), pp. 41–68. Available at: https://doi.org/10.3745/JIPS.2009.5.2.041.

Karagoz, A. and Dindis, G. (2025) ‘Object Recognition and Positioning with Neural Networks: Single Ultrasonic Sensor Scanning Approach.’, *Sensors (14248220)*, 25(4), p. 1086. Available at: https://doi.org/10.3390/s25041086.

Kittur, P. *et al.* (2023) ‘A Review on Anti-Spoofing: Face Manipulation and Liveness Detection’, in *2023 IEEE International Conference on Computer Vision and Machine Intelligence (CVMI)*. *2023 IEEE International Conference on Computer Vision and Machine Intelligence (CVMI)*, pp. 1–6. Available at: https://doi.org/10.1109/CVMI59935.2023.10464680.

Kolberg, J., Rathgeb, C. and Busch, C. (2023) ‘The Influence of Gender and Skin Colour on the Watchlist Imbalance Effect in Facial Identification Scenarios’, in J.-J. Rousseau and B. Kapralos (eds) *Pattern Recognition, Computer Vision, and Image Processing. ICPR 2022 International Workshops and Challenges*. Cham: Springer Nature Switzerland, pp. 465–478. Available at: https://doi.org/10.1007/978-3-031-37660-3_33.

Li, J. *et al.* (2022) ‘A Survey on Deep Learning for Named Entity Recognition’, *IEEE Transactions on Knowledge and Data Engineering*, 34(1), pp. 50–70. Available at: https://doi.org/10.1109/TKDE.2020.2981314.

Li, Q. *et al.* (2021) ‘Video Face Recognition with Audio-Visual Aggregation Network’, in T. Mantoro et al. (eds) *Neural Information Processing*. Cham: Springer International Publishing, pp. 150–161. Available at: https://doi.org/10.1007/978-3-030-92273-3_13.

Liu, D. *et al.* (2023) ‘SoundID: Securing Mobile Two-Factor Authentication via Acoustic Signals’, *IEEE Transactions on Dependable and Secure Computing*, 20(2), pp. 1687–1701. Available at: https://doi.org/10.1109/TDSC.2022.3162718.

Maeva, Anna. and Severin, F. (2009) ‘High resolution ultrasonic method for 3D fingerprint recognizable characteristics in biometrics identification’, in *2009 IEEE International Ultrasonics Symposium*. *2009 IEEE International Ultrasonics Symposium*, pp. 2260–2263. Available at: https://doi.org/10.1109/ULTSYM.2009.5441399.

Menezes, H.F. *et al.* (2021) ‘Bias and Fairness in Face Detection’, in *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 247–254. Available at: https://doi.org/10.1109/SIBGRAPI54419.2021.00041.

Nie, M. *et al.* (2023) ‘Research on Localization and three-dimensional Reconstruction of Borehole Wall Fractures Based on Ultrasonic Imaging Data’, in *2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP)*. *2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP)*, pp. 238–242. Available at: https://doi.org/10.1109/ICSP58490.2023.10248880.

Pareek, S. *et al.* (2024) ‘Efficient Vision Transformers for Edge Devices: Pruning and Quantization Approaches’, in *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)*. *2024 4th International Conference on Technological*

Advancements in Computational Sciences (ICTACS), pp. 1465–1471. Available at: https://doi.org/10.1109/ICTACS62700.2024.10840584.

Prasanth, A. and Micheal, A.A. (2025) 'Enhancing Occluded Face Recognition Through Device-Edge Collaboration and Cross-Domain Feature Fusion with Heatmaps', in *2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. *2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, pp. 1743–1749. Available at: https://doi.org/10.1109/IDCIOT64235.2025.10914798.

Ryu, R. *et al.* (2021) 'Continuous Multimodal Biometric Authentication Schemes: A Systematic Review', *IEEE Access*, 9, pp. 34541–34557. Available at: https://doi.org/10.1109/ACCESS.2021.3061589.

Scheenstra, A., Ruifrok, A. and Veltkamp, R.C. (2005) 'A Survey of 3D Face Recognition Methods', in T. Kanade, A. Jain, and N.K. Ratha (eds) *Audio- and Video-Based Biometric Person Authentication*. Berlin, Heidelberg: Springer, pp. 891–899. Available at: https://doi.org/10.1007/11527923_93.

Shah, J. *et al.* (2024) 'Strengthening Facial Biometrics: An Innovative Approach to Liveness Detection for Anti-Spoofing', in *2024 IEEE 8th International Conference on Information and Communication Technology (CICT)*. *2024 IEEE 8th International Conference on Information and Communication Technology (CICT)*, pp. 1–6. Available at: https://doi.org/10.1109/CICT64037.2024.10899685.

Sindhu, B. *et al.* (2025) 'A Novel Multimodal Biometric Fusion for Enhanced Personal Authentication', in S. Manoharan, A. Tugui, and I. Perikos (eds) *Proceedings of 5th International Conference on Artificial Intelligence and Smart Energy*. Cham: Springer Nature Switzerland, pp. 238–245. Available at: https://doi.org/10.1007/978-3-031-90478-3_20.

Solano, I. *et al.* (2025) 'Comprehensive Equity Index (CEI): Definition and Application to Bias Evaluation in Biometrics', in A. Antonacopoulos et al. (eds) *Pattern Recognition*. Cham: Springer Nature Switzerland, pp. 110–126. Available at: https://doi.org/10.1007/978-3-031-78341-8_8.

Tay, Y. *et al.* (2022) 'Efficient Transformers: A Survey', *ACM Comput. Surv.*, 55(6), p. 109:1-109:28. Available at: https://doi.org/10.1145/3530811.

Vaswani, A. *et al.* (2017) 'Attention is All you Need', in *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (Accessed: 8 April 2025).

Yang, H. *et al.* (2020) 'FaceScape: a Large-scale High Quality 3D Face Dataset and Detailed Riggable 3D Face Prediction'. arXiv. Available at: https://doi.org/10.48550/arXiv.2003.13989.

Yu, S. *et al.* (eds) (2025) *Biometric Recognition: 18th Chinese Conference, CCBR 2024, Nanjing, China, November 22–24, 2024, Proceedings, Part I*. Singapore: Springer Nature (Lecture Notes in Computer Science). Available at: https://doi.org/10.1007/978-981-96-1068-6.

Yucer, S. *et al.* (2025) 'Racial Bias within Face Recognition: A Survey.', *ACM Computing Surveys*, 57(4), pp. 1–39. Available at: https://doi.org/10.1145/3705295.

Zhang, T. *et al.* (2023) ‘Accurate 3D Face Reconstruction with Facial Component Tokens’, in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8999–9008. Available at: https://doi.org/10.1109/ICCV51070.2023.00829.

Zhao, W. *et al.* (2003) ‘Face recognition: A literature survey’, *ACM Computing Surveys*, 35(4), pp. 399–458. Available at: https://doi.org/10.1145/954339.954342.

Zhou, B. *et al.* (2022) ‘Robust Human Face Authentication Leveraging Acoustic Sensing on Smartphones’, *IEEE Transactions on Mobile Computing*, 21(8), pp. 3009–3023. Available at: https://doi.org/10.1109/TMC.2020.3048659.

Zhu, H. *et al.* (2023) ‘FaceScape: 3D Facial Dataset and Benchmark for Single-View 3D Face Reconstruction’. arXiv. Available at: https://doi.org/10.48550/arXiv.2111.01082.