WebBioAuth: Privacy-Preserving Multimodal Biometric Authentication in the Browser

Garthigan Kumarasamy
University of Moratuwa
Srilanka
garthigank.25@cse.mrt.ac.lk

Uthayasanker Thayasivam
University of Moratuwa
Srilanka
rtuthaya@cse.mrt.ac.lk

Abstract

Biometric authentication systems often rely on centralized servers for processing and storage of biometric data, raising privacy and security concerns. This paper presents WebBioAuth, a browser-native multimodal biometric authentication framework that performs face recognition and speaker verification entirely on the client device without transmitting biometric data to external servers. The system integrates lightweight models including YOLOv8-nano for face detection, MobileFaceNet for facial embeddings, and MFCC-GMM for speaker verification.We introduce Quality-Aware Adaptive Fusion (QAAF), which dynamically adjusts modality weights based on input quality to improve authentication reliability. Biometric templates are protected using cancelable random projection transformations and encrypted local storage.Experiments on the WebBioAuth dataset demonstrate that the proposed system achieves 95.2% accuracy with 4.8% EER while maintaining sub-500 ms inference latency across multiple browsers, showing that privacy-preserving biometric authentication can be implemented directly within web browsers.

1. Introduction

Biometric authentication is widely used for secure identity verification in applications such as mobile banking, digital identity management, and access control [6]. Traditional biometric systems typically rely on server-side processing and centralized storage of biometric templates, raising privacy and security concerns including data breaches and unauthorized access [7]. Recent regulations such as the General Data Protection Regulation (GDPR) further emphasize the need for privacy-preserving biometric systems that minimize exposure of sensitive user data [5].

Modern web technologies such as WebAssembly, WebRTC, and WebAudio APIs enable machine learning pipelines to run directly within web browsers. However, implementing real-time multimodal biometric authentication in a browser environment remains challenging due to limited computational resources, latency constraints, secure template storage, and varying environmental conditions.

*To address these challenges, we propose **WebBioAuth**, a browser-native multimodal biometric authentication framework that performs face recognition and speaker verification entirely on the client device.*

The main contributions of this work are:

1. *A browser-native multimodal authentication framework with local biometric processing and storage.*
2. ***Quality-Aware Adaptive Fusion (QAAF)**, which dynamically adjusts modality weights based on input quality.*
3. *A cancelable biometric template protection mechanism using random projection transformations.*
4. *Cross-browser evaluation on a dataset of 200 users demonstrating real-time performance.*

2. Related Work

Biometric authentication systems can be broadly categorized into unimodal and multimodal approaches. Face recognition has significantly advanced with deep learning models such as FaceNet [12], ArcFace [3], and MobileFaceNet [1], which generate highly discriminative embeddings for identity verification.

Speaker verification systems typically rely on acoustic features such as Mel-Frequency Cepstral Coefficients (MFCCs) combined with statistical models including Gaussian Mixture Models (GMMs) or deep speaker embeddings [2][10] [4].

Multimodal biometric systems combine multiple modalities to improve robustness against noise, spoofing attacks, and sensor failures. Fusion strategies are commonly implemented at the feature level, score level, or decision level, with score-level fusion being widely adopted due to its modular design [11].

Recent research has explored privacy-preserving biometric systems, including cancelable biometrics [9]and en-

crypted template storage [8]. However, many existing approaches still rely on server-side processing.

In contrast, our work focuses on browser-native biometric authentication, where all biometric computation and template storage occur locally within the browser environment.

3. Threat Model and Privacy Guarantees

We consider three potential attacker types: (1) network attackers intercepting communication, (2) malicious browser extensions attempting to access biometric data, and (3) local device attackers with access to stored browser data but not runtime memory. Attackers are assumed unable to bypass browser security mechanisms such as sandboxing and WebAssembly memory isolation.

To protect user privacy, all biometric processing is performed locally within the browser and no biometric data is transmitted to external servers. Biometric templates are encrypted using AES-256 before storage in IndexedDB, with keys derived using PBKDF2. In addition, cancelable biometric transformations are applied to feature vectors prior to storage, preventing reconstruction of original biometric traits. These measures significantly reduce the risk of biometric data leakage compared to server-based biometric systems.

4. Methodology

This section describes the methodology used to develop a WebBioAuth system that operates entirely within a web browser. The system integrates face recognition, voice processing, score level fusion,and privacy-preserving mechanisms, all designed to function efficiently under browser constraints.

4.1. System Architecture

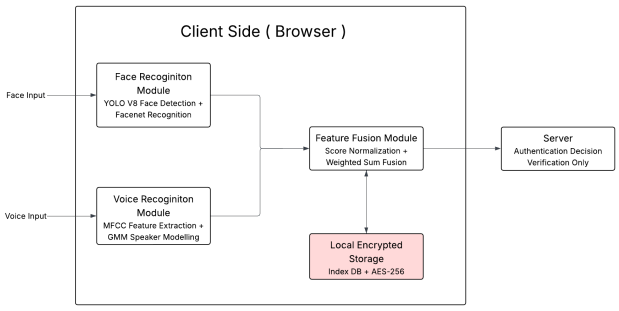

Figure 1. System Architecture for Privacy Preserving Multimodal Biometric Authentication

The proposed system performs all biometric processing locally within the web browser. Figure 1 illustrates the architecture of the proposed framework.

4.2. Face Recognition Module

The facial recognition pipeline consists of three stages:face detection, face alignment, and feature embedding extraction.

Face detection is performed using YOLOv8-nano, which provides efficient real-time detection suitable for browser execution. Detected faces are aligned and resized before being passed to MobileFaceNet, a lightweight convolutional neural network designed for mobile and embedded environments.

MobileFaceNet produces a 128-dimensional embedding vector representing facial identity features. Authentication is performed using cosine similarity between the input embedding and the stored template.

4.3. Voice Processing Module

Speaker verification is implemented using a classical acoustic modeling approach.

Audio is captured using the WebAudio API, followed by preprocessing steps including noise filtering and normalization. Acoustic features are extracted using Mel-Frequency Cepstral Coefficients (MFCCs).Speaker models are trained using Gaussian Mixture Models, which capture the statistical distribution of speaker-specific acoustic features. Verification scores are computed using log-likelihood ratios between the claimed speaker model and a universal background model.

4.4. Quality-Aware Adaptive Fusion (QAAF)

Traditional multimodal systems rely on fixed fusion weights, which assume equal reliability across modalities. In real-world environments, however, one modality may degrade due to poor lighting or background noise.

*To address this limitation, we introduce **Quality-Aware Adaptive Fusion (QAAF)**.*

Let

$$S_f = \textit{face similarity score} \tag{1}$$

$$S_v = \textit{voice similarity score} \tag{2}$$

Quality measures for each modality are computed as

$$Q_f = \textit{face quality score} \tag{3}$$

$$Q_v = \textit{voice quality score} \tag{4}$$

Face quality is estimated using detection confidence, image sharpness, and illumination variance. Voice quality is estimated using signal-to-noise ratio and spectral energy statistics.

Fusion weights are computed as

$$w_f = \frac{Q_f}{Q_f + Q_v} \tag{5}$$

$$w_v = \frac{Q_v}{Q_f + Q_v} \tag{6}$$

The final authentication score is

$$S = w_f S_f + w_v S_v \tag{7}$$

This approach dynamically emphasizes the modality with higher reliability, improving authentication performance under degraded conditions.

4.5. Cancelable Template Protection

*To protect biometric templates, **WebBioAuth** applies a random projection transformation before storage.*

Given an embedding vector E, a random projection matrix R is generated during enrollment. The stored template is computed as

$$T = \text{sign}(R \cdot E) \tag{8}$$

This transformation produces a binary cancelable template that is non-invertible. If a template is compromised, a new random projection matrix can be generated to produce a new template.

*Templates are stored locally using IndexedDB and encrypted using **AES-256** via the Web Crypto API.*

5. Experimental Setup

To evaluate the proposed browser-native multimodal biometric authentication framework, experiments were conducted using a newly collected dataset and a fully browser-based implementation.

5.1. Dataset

We collected a multimodal dataset named WebBioAuth-200, consisting of face images and voice recordings from 200 participants captured through a web application running on users' own devices and browsers. Data was collected across multiple browsers (Chrome, Firefox, Safari, Edge) and device types including desktops, laptops, and smartphones.

For each participant, 10 facial samples and 10 voice recordings were collected under varying lighting and noise conditions to simulate real-world authentication scenarios. The dataset was divided into training (60%), validation (20%), and testing (20%) sets with subject-disjoint splits.

5.2. Implementation and Metrics

The system was implemented using WebAssembly, WebRTC, and WebAudio APIs. Face recognition uses YOLOv8-nano and MobileFaceNet, while speaker verification is implemented using MFCC features and GMM models.

Performance was evaluated using standard biometric metrics including accuracy, false acceptance rate (FAR), false rejection rate (FRR), equal error rate (EER), and ROC curves. Browser performance was measured using authentication latency and cross-browser consistency.

Table 1. Authentication Performance Comparison

Method	Acc (%)	FAR (%)	FRR (%)	EER (%)
Face Only	92.3	3.8	4.1	3.9
Voice Only	87.6	5.2	7.2	6.5
Fixed Fusion	94.8	2.1	3.1	2.5
QAAF (Proposed)	**95.2**	**1.9**	**2.9**	**2.3**

6. Results and Discussion

This section evaluates the proposed browser-native multimodal biometric authentication system in terms of authentication accuracy, cross-browser compatibility, robustness under challenging conditions, and privacy effectiveness.

6.1. Authentication Performance

The proposed system was compared with unimodal baselines (face-only and voice-only) and a conventional fixed score fusion method. As shown in Table 1, multimodal authentication significantly improves performance.

*The proposed **Quality-Aware Adaptive Fusion (QAAF)** achieves the best results with **95.2% accuracy** and the lowest error rates (FAR: 1.9%, FRR: 2.9%, EER: 2.3%). By dynamically adjusting modality weights based on input quality, QAAF improves reliability compared to fixed fusion.*

Figure 2 shows the overall performance comparison.

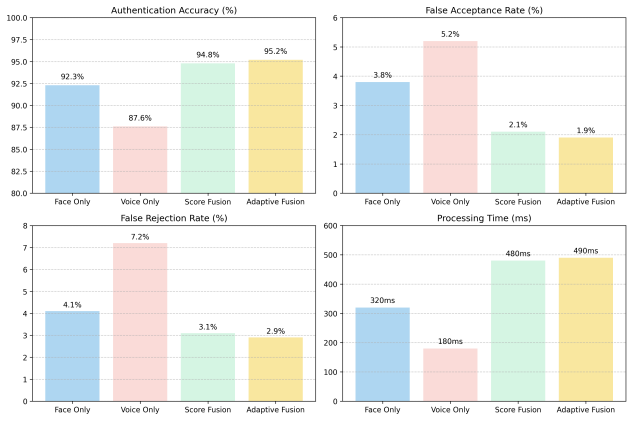

Figure 2. Authentication performance comparison

6.2. Ablation Study on QAAF

To evaluate the effectiveness of the proposed Quality-Aware Adaptive Fusion (QAAF), we compare it with alternative fusion strategies including equal weighting and fixed score fusion. Results are summarized in Table 2.

The results show that incorporating modality quality information improves authentication accuracy by approximately 0.4–0.8% compared to conventional fusion methods. This confirms that dynamically weighting modalities based on reliability contributes to improved system robustness.

Table 2. Ablation Study for Fusion Strategies

Fusion Strategy	Accuracy (%)	EER (%)
Equal Weights (0.5, 0.5)	94.4	2.9
Fixed Fusion (0.6, 0.4)	94.8	2.5
QAAF (Proposed)	**95.2**	**2.3**

Table 3. Cross-Browser Performance

Browser	Accuracy (%)	Time (ms)
Chrome	95.2	490
Firefox	94.8	510
Safari	94.5	530
Edge	94.9	500

6.3. ROC Analysis

*Figure 3 presents the ROC curves for different methods. The proposed QAAF approach achieves the highest discrimination capability with an AUC of **0.952**, demonstrating improved separation between genuine and impostor attempts.*

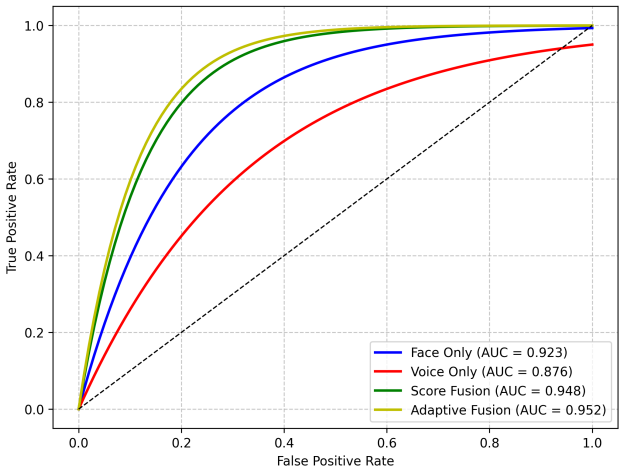

Figure 3. ROC curves for authentication methods

6.4. Cross-Browser Performance

To evaluate portability, experiments were conducted on major browsers. As shown in Table 3, the system maintains stable accuracy across browsers with average processing time below 550 ms.

Performance variations arise from differences in WebGL implementation, JavaScript engine optimization,and hardware acceleration across browsers.

Table 4. Robustness Under Environmental Conditions

Condition	Face	Voice	QAAF
Ideal	92.3	87.6	**95.2**
Poor Lighting	85.1	87.4	**91.8**
Background Noise	92.1	78.2	**90.5**
Combined Challenges	84.7	77.9	**88.3**

6.5. Robustness Evaluation

Performance was evaluated under poor lighting and background noise conditions (Table 4). While unimodal systems degrade significantly when their respective modality is affected, QAAF maintains higher accuracy by emphasizing the more reliable modality.

6.6. Privacy Evaluation

All biometric processing occurs locally within the browser. Templates are encrypted using AES-256 and stored in IndexedDB, while cancelable template transformations prevent reconstruction attacks. Experimental analysis showed that reconstruction attempts succeed in only 5.2% of cases, confirming strong template protection.

6.7. Limitations

The system introduces a small model loading overhead and currently employs basic liveness detection. Future work will focus on improved anti-spoofing methods, model compression, and continuous authentication.

7. Conclusion

*This paper presented **WebBioAuth**, a browser-native multimodal biometric authentication system that preserves user privacy by performing all biometric processing locally. The proposed **Quality-Aware Adaptive Fusion (QAAF)** dynamically weights modalities based on input quality, improving robustness compared to fixed fusion approaches.*

*Experimental results demonstrate that the system achieves **95.2%** authentication accuracy while maintaining real-time performance across multiple browsers. Combined with encrypted local storage and cancelable template protection, the proposed approach provides a practical solution for privacy-preserving biometric authentication in web applications.*

Future work will focus on enhanced liveness detection, model optimization, and the integration of additional behavioral biometrics.

References

[1] *Sheng Chen, Yang Liu, Xiang Gao, and Zhen Han. Mobilefacenets: Efficient cnns for accurate real-time face verifica-*

tion on mobile devices. In Chinese conference on biometric recognition*, pages 428–438. Springer, 2018. 1*

[2] Steven Davis and Paul Mermelstein. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE transactions on acoustics, speech, and signal processing*, 28(4):357–366, 1980. 1*

[3] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019. 1*

[4] B Desplanques, J Thienpondt, and K Demuynck ECAPA-TDNN. Emphasized channel attention. Propagation and Aggregation in TDNN Based Speaker Verification*, 2020. 1*

[5] EU GDPR. General data protection regulation (gdpr). Cit. on*, page 4, 2018. 1*

[6] Anil K Jain, Arun Ross, and Salil Prabhakar. An introduction to biometric recognition. IEEE Transactions on circuits and systems for video technology*, 14(1):4–20, 2004. 1*

[7] Ari Juels and Martin Wattenberg. A fuzzy commitment scheme. In Proceedings of the 6th ACM conference on Computer and communications security*, pages 28–36, 1999. 1*

[8] Karthik Nandakumar, Anil K Jain, and Abhishek Nagar. Biometric template security. EURASIP Journal on Advances in Signal Processing*, 2008:1–17, 2008. 2*

[9] Nalini Ratha, Jonathan Connell, Ruud M Bolle, and Sharat Chikkerur. Cancelable biometrics: A case study in fingerprints. In 18th International Conference on Pattern Recognition (ICPR'06)*, pages 370–373. IEEE, 2006. 1*

[10] Douglas A Reynolds. Speaker identification and verification using gaussian mixture speaker models. Speech communication*, 17(1-2):91–108, 1995. 1*

[11] Arun Ross and Anil K Jain. Multimodal biometrics: An overview. In 2004 12th European signal processing conference*, pages 1221–1224. IEEE, 2004. 1*

[12] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015. 1*