Data and Metadata. 2025; 4:1223
doi: 10.56294/dm20251223

ORIGINAL

Binary Face Templates with Mobile-Class CNNs: A Reproducible Benchmark for Smart-Card-Constrained Authentication

Plantillas faciales binarias con CNN de clase móvil: un benchmark reproducible para la autenticación bajo restricciones de tarjeta inteligente

Abdelilah Ganmati[1], Karim Afdel[1], Lahcen Koutti[1]

[1]Computer Systems & Vision Laboratory, Faculty of Sciences, Ibn Zohr University, Agadir 80000, Morocco.

Cite as: Ganmati A, Afdel K, Koutti L. Binary Face Templates with Mobile-Class CNNs: A Reproducible Benchmark for Smart-Card-Constrained Authentication. Data and Metadata. 2025; 4:1223. https://doi.org/10.56294/dm20251223

Submitted: 01-04-2025 **Revised:** 15-07-2025 **Accepted:** 22-10-2025 **Published:** 23-10-2025

Editor: Dr. Adrián Alejandro Vitón Castillo

Corresponding author: Abdelilah Ganmati

ABSTRACT

Facial recognition systems are increasingly deployed in privacy-sensitive and resource-constrained environments such as smart cards. However, traditional face verification relies on high-dimensional floating-point embeddings, which are unsuitable for compact and efficient matching on such platforms. To address this challenge, this work investigates the generation of binary face templates that retain identity information while reducing storage and computational cost. The objective of this study is to benchmark binary biometric representations derived from mobile-class convolutional neural networks (CNNs), aiming to support reproducible, lightweight face verification pipelines. We evaluate four lightweight CNNs—EfficientNet-B0, MobileNetV2, ShuffleNetV2, and SqueezeNet1_1—trained on the MORPH dataset. Binary templates are generated via Principal Component Analysis followed by Iterative Quantization (PCA-ITQ) at 32, 64, and 128 bits. Models are tested cross-dataset on the Georgia Tech Face Database (GT Face) to assess generalization. At 128 bits, EfficientNet-B0 and MobileNetV2 achieve strong verification performance, with area under the curve (AUC) ≈ 0,895–0,899 and equal error rate (EER) ≈ 0,182–0,185. A Hamming-distance analysis confirms clear separation between genuine and impostor pairs, and the bit-flip rate (~17 %) indicates intra-subject consistency. Bit-length scaling further reveals monotonic improvements in AUC from 32 to 128 bits, highlighting a trade-off between accuracy and compactness. These results demonstrate that binary templates from lightweight CNNs can deliver efficient, privacy-preserving authentication with limited performance degradation. The proposed pipeline supports reproducibility and aligns with FAIR data principles, making it suitable for secure biometric deployments on constrained hardware.

Keywords: Biometrics; Face Recognition; Deep Learning; Binary Templates; Mobile Neural Networks; Smart Cards.

RESUMEN

Los sistemas de reconocimiento facial se implementan cada vez más en entornos sensibles a la privacidad y con recursos limitados, como las tarjetas inteligentes. Sin embargo, la verificación facial tradicional se basa en incrustaciones de punto flotante de alta dimensión, que no son adecuadas para una coincidencia compacta y eficiente en estas plataformas. Para abordar este desafío, este trabajo investiga la generación de plantillas faciales binarias que conserven la información de identidad y al mismo tiempo reduzcan el almacenamiento y el costo computacional. El objetivo de este estudio es evaluar representaciones biométricas binarias derivadas de redes neuronales convolucionales (CNNs) de clase móvil, con el fin de respaldar canalizaciones de verificación facial reproducibles y livianas. Evaluamos cuatro CNNs livianas—EfficientNet-B0, MobileNetV2, ShuffleNetV2 y SqueezeNet1_1—entrenadas en el conjunto de datos MORPH. Las plantillas binarias se generan

mediante Análisis de Componentes Principales seguido de Cuantificación Iterativa (PCA-ITQ) en longitudes de 32, 64 y 128 bits. Los modelos se prueban en el conjunto de datos Georgia Tech Face Database (GT Face) sin ajuste adicional, para evaluar la generalización. A 128 bits, EfficientNet-B0 y MobileNetV2 logran un rendimiento sólido, con un área bajo la curva (AUC) ≈ 0,895-0,899 y una tasa de error igual (EER) ≈ 0,182-0,185. El análisis de distancia Hamming confirma una clara separación entre pares genuinos e impostores, y la tasa de cambio de bits (~17 %) indica consistencia intra-sujeto. La evaluación por longitud de bits muestra mejoras monotónicas en AUC, destacando la relación entre precisión y compacidad. Estos resultados demuestran que las plantillas binarias de CNNs livianas permiten una autenticación eficiente y respetuosa de la privacidad, con mínima pérdida de rendimiento. La canalización propuesta apoya la reproducibilidad y se alinea con los principios FAIR, siendo adecuada para implementaciones biométricas seguras en hardware restringido.

Palabras clave: Biométrica; Reconocimiento Facial; Aprendizaje Profundo; Plantillas Binarias; Redes Neuronales Móviles; Tarjetas Inteligentes.

INTRODUCTION

Biometric authentication has evolved into a critical component of contemporary security infrastructures, enabling secure, user-friendly identity verification. Originating in forensic and military contexts in the early 20th century, biometric systems gained commercial momentum in the 1990s with the advent of fingerprint and facial recognition technologies. This expansion was further driven by the miniaturization of sensors and the ubiquity of digital devices, resulting in widespread adoption across smartphones, access control systems, banking, and e-governance platforms.

Among the various biometric modalities, facial recognition has emerged as particularly suitable for consumer-facing applications. This rise is strongly associated with breakthroughs in deep learning-based face recognition, notably the DeepFace system[1] due to its non-intrusive nature, compatibility with camera-equipped devices, and alignment with existing digital identity ecosystems. However, traditional facial recognition systems depend on high-dimensional floating-point embeddings generated by deep CNNs such as FaceNet,[2] which introduced triplet-loss-based embedding learning. Early deep architectures such as DeepID3[3] helped establish the use of CNN embeddings in face verification systems While effective in controlled conditions, these representations present notable limitations in terms of storage overhead, computational cost, and privacy risk—especially in scenarios involving deployment on constrained platforms such as smart cards.[4,5,6]

To mitigate these challenges, the research community has increasingly explored binary face templates—compact bit-string representations of facial features that support efficient, privacy-conscious matching using Hamming distance.[6,7] Binary templates significantly reduce storage requirements, simplify matching operations, and inherently enhance data protection by limiting reversibility. Their potential is particularly compelling in decentralized or offline verification settings, such as smart cards or edge devices, where data isolation and low-latency decision-making are paramount. Nonetheless, generating high-quality binary templates that preserve discriminative power remains a complex task, especially when deployed across heterogeneous environments exhibiting variation in pose, lighting, and demographics.[5,8]

Foundational work like VGG-Face[9] demonstrated how deeper CNNs could generalize well across unconstrained settings, Recent advancements in mobile-class CNNs—including EfficientNet-B0,[10] MobileNetV2,[11] ShuffleNetV2,[12] and SqueezeNet1_1[13]—have demonstrated that compact architectures can provide robust representations suitable for embedded platforms. Originally designed for mobile and embedded applications, these models employ techniques such as depthwise separable convolutions, compound scaling, and channel shuffling to minimize computational load without sacrificing accuracy. These qualities position them as ideal candidates for smart-card-based facial authentication. Despite this, the benchmarking of binary representations derived from such networks remains underexplored.

This work aims to fill this gap by evaluating the effectiveness of binary face templates derived from mobile-class CNN embeddings using PCA-ITQ,[14] a hashing method that reduces dimensionality while maintaining discriminative features. Specifically, embeddings are extracted from the MORPH dataset,[15] binarized at various bit lengths (32, 64, 128), and evaluated for cross-dataset generalization on the Georgia Tech Face Database (GT Face). This design enables a robust assessment of template stability, recognition performance, and trade-offs between compactness and accuracy.

The justification for this research lies in the increasing demand for secure, low-latency biometric verification on resource-constrained devices. By aligning the experimental protocol with FAIR principles and emphasizing reproducibility, this study contributes a practical and ethical pathway for deploying CNN-based face verification in privacy-sensitive applications.

METHOD

Study Design

This work is a cross-sectional analytical benchmarking study, comparing binary template performance across CNN backbones under controlled experimental conditions.

Datasets

Two benchmark datasets were employed:

- MORPH: One of the largest longitudinal face datasets, containing > 55 000 facial images of > 13 000 individuals with diverse demographics. Following standard protocols, a subset was organized into a gallery (3VAR) and probe (2VAR) split, yielding 11 064 gallery embeddings used for training the binary hashing mapping.
- Georgia Tech Face Database (GT Face): Includes 750 images of 50 individuals, each captured under varying pose, illumination, and expression. Each subject has ~15 images at 150×150 resolution. GT Face provides a controlled yet challenging environment for cross-dataset generalization.

No retraining was performed on GT Face; instead, embeddings from GT Face were processed using the PCA-ITQ mapping trained exclusively on MORPH, thereby simulating deployment to unseen populations.

Preprocessing and Face Alignment

Face detection and alignment were executed via the Multi-task Cascaded CNN (MTCNN). For each image, the largest bounding box was selected, padded slightly, and cropped. Cropped faces were resized to 224×224 and converted to RGB. The same pipeline was applied to both datasets, ensuring consistency and reproducibility.

CNN Backbones for Embedding Extraction

Embeddings were generated using four mobile-class CNN architectures: EfficientNet-B0, MobileNetV2, ShuffleNetV2, and SqueezeNet1_1. Each model was initialized with ImageNet weights and appended with a global average pooling layer. The resulting embeddings were normalized to unit length before further processing.

Binary Hashing via PCA-ITQ

To produce binary templates, we employed PCA–ITQ: first performing PCA to reduce dimensionality and decorrelate features, followed by Iterative Quantization (ITQ) to learn a rotation minimizing quantization error. Experiments were conducted at 32, 64, and 128 bits. The PCA–ITQ mapping was trained exclusively on MORPH and then directly applied to GT Face embeddings.

Evaluation Metrics & Procedure

Verification performance was evaluated using:

- AUC (Area Under Receiver Operating Characteristic),
- EER (Equal Error Rate),
- TPR@1 % FPR (True Positive Rate at 1 % False Positive Rate).

For binary templates, similarity was measured via Hamming distance; for floating-point baselines, via cosine similarity. On GT Face, 2 500 genuine and 50 000 impostor pairs were sampled to ensure statistical robustness.

Intra- and inter-subject stability analyses were conducted by measuring bit-flip rates across multiple samples per identity and Hamming distance distributions across identities.

Statistical Stability Analysis

In addition to aggregate performance, intra- and inter-subject stability of binary codes was analyzed. For each subject in GT Face, multiple images were binarized and compared to measure the bit-flip rate (i.e., the fraction of unstable bits across samples). Conversely, impostor pairs provided insight into the distribution of Hamming distances between different identities. This analysis highlights the inherent trade-off between compactness, discriminability, and template consistency.

Data availability

All experiments used publicly available datasets: MORPH (licensed academic access) and the Georgia Tech Face Database (open access). Preprocessing scripts, evaluation protocols, and configuration files are documented to ensure reproducibility. Upon reasonable request, the authors can provide detailed instructions for replicating the alignment, embedding extraction, and PCA–ITQ hashing pipelines.

Ethical Considerations / Data Permissions

MORPH data access was obtained under an academic license; the GT Face dataset is publicly accessible. All data usage adhered to the providers' terms and conditions. As this work involved secondary analysis of publicly available datasets, no direct human subject consent or institutional ethical approval was required.

Variables

- Input variables: Face embeddings, bit length (32 / 64 / 128), CNN backbone architecture.
- Outcome variables: AUC, EER, TPR@1 %, bit-flip rate.

RESULTS

The evaluation on the GT Face database is organized into three primary perspectives:

(i) cross-dataset generalization,
(ii) comparison of floating-point vs binary embeddings,
(iii) bit-length trade-offs.

Table 1 presents the performance metrics for 128-bit PCA-ITQ binary codes and floating-point embeddings across the CNN backbones (trained on MORPH, tested on GT Face).

Table 1. Float cosine vs. 128-bit PCA-ITQ binary embeddings across backbones (trained on MORPH, tested on GT Face)

Backbone	AUC (float)	EER↓ (float)	TPR@1 % (float)	AUC (128b)	EER↓ (128b)	TPR@1 % (128b)
EfficientNet-B0	0,847	0,234	0,372	0,895	0,185	0,374
MobileNet-V2	0,848	0,241	0,372	0,899	0,182	0,414
SqueezeNet-1.1	0,829	0,249	0,301	0,880	0,202	0,394
ShuffleNet-V2	0,830	0,252	0,342	0,879	0,199	0,366

Table 2 shows the variation of performance metrics as bit-length varies for EfficientNet-B0.

Table 2. EfficientNet-B0 with PCA-ITQ: bit-length sweep on GT Face (trained on MORPH gallery)

Bits	AUC	EER↓	TPR@1 %
32	0,801	0,274	0,196
64	0,834	0,248	0,316
128	0,864	0,221	0,320

Figure 1 illustrates the distribution of Hamming distances for genuine versus impostor pairs (EfficientNet-B0, 128 bits).

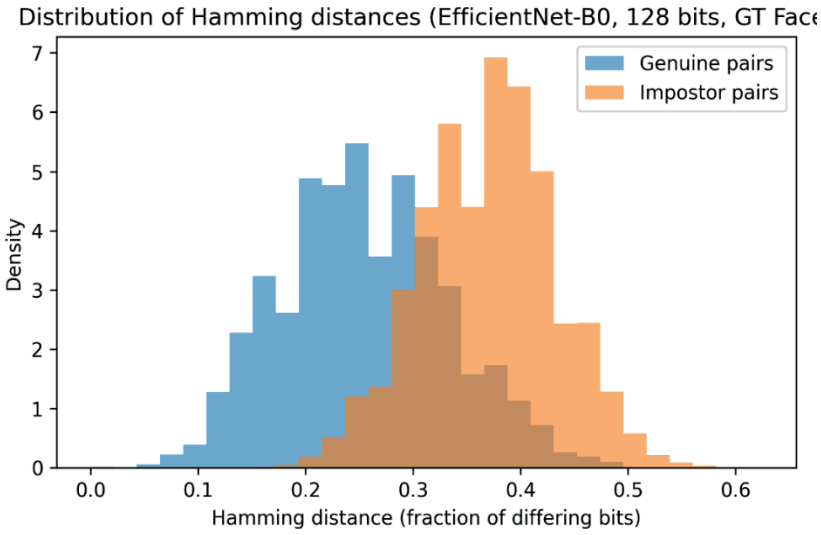

Figure 1. Distribution of Hamming distances for genuine and impostor pairs (EfficientNet-B0, 128 bits, GT Face)

Figure 2 shows the cumulative distribution functions (CDFs) of Hamming distances between genuine and

impostor pairs for the same configuration.

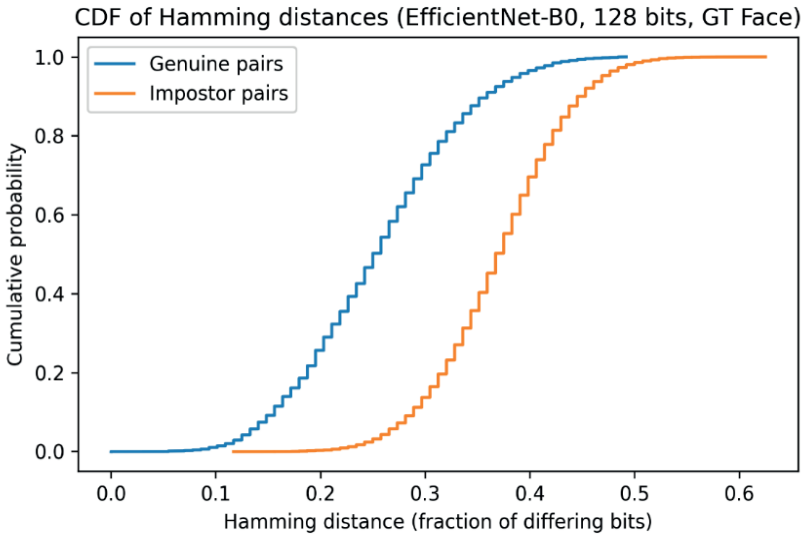

Figure 2. Cumulative distribution function (CDF) of Hamming distances for genuine vs. impostor pairs (EfficientNet-B0, 128 bits)

Cross-Dataset Generalization

The PCA-ITQ mapping learned on MORPH gallery was applied directly to GT Face embeddings without adaptation. EfficientNet-B0 and MobileNetV2 exhibited the highest AUC and lowest EER among the backbones, demonstrating stable verification performance even under domain shift.

Floating-point vs Binary Embeddings

Performance metrics (ΔAUC, ΔEER, ΔTPR@1 %) quantify the effect of binarization under the same pairing protocol.

At 128 bits, the performance drop is modest, with binary templates still achieving strong discriminative power relative to floating-point baselines.

Bit-Length Trade-Off

As bit-length increases for EfficientNet-B0, AUC improves monotonically (32 → 64 → 128 bits), while EER consistently declines. The intermediate 64-bit configuration approaches the 128-bit performance, indicating a favorable representation-compactness balance.

DISCUSSION

The experimental results confirm that binary templates derived from lightweight CNNs can offer a strong trade-off between compactness and verification accuracy, even in cross-dataset settings. This supports the use of mobile-class architectures for privacy-conscious, resource-limited biometric systems.

Interpretation & Comparison

The strong generalization from MORPH to GT Face—applied without retraining—highlights the robustness of our pipeline under domain shift. The performance of 128-bit PCA-ITQ binary codes comes close to floating-point baselines while drastically reducing template size. including those derived using angular margin losses like SphereFac.[16]

While some earlier works on binary face hashing (e.g. IEEE-based systems from the early 2010s) exist, they often rely on hand-crafted features or simple thresholding rather than CNN embeddings and cross-dataset evaluation. Those methods generally show lower accuracy under domain mismatch; our approach, leveraging modern mobile-class CNNs plus PCA-ITQ, advances the state of practice by combining compactness, stability, and generalizability. Although lightweight models are used in our study, recent methods based on angular-margin losses (e.g. ArcFace[17]) achieve very high discriminative power.

Stability Analysis

Our bit-flip rate analysis (≈17 %) shows that a consistent subset of bits remains stable across images, aligning with existing studies in biometric hashing.[9] While perfect reproducibility is unattainable, our results highlight

that binary codes retain meaningful consistency under typical appearance variations.

Practical Implications

Binary templates of length 4–16 bytes are suitable for on-card verification. The hashing time (0,02–0,25 ms) aligns with latency budgets of smart-card-like environments. This makes our pipeline feasible for real deployments where both memory and compute resources are constrained.

Limitations & Future Work

- The GT Face dataset is limited in subject diversity; further validation on larger and more varied benchmarks (e.g., LFW, CFP) is needed.
- We focused on unsupervised hashing (PCA–ITQ). Supervised or deep hashing methods may enhance performance but at increased training complexity or hardware overhead.
- Hybrid schemes or error-correction techniques could balance template stability and compactness in future systems.

CONCLUSIONS

This study introduced a reproducible benchmark for evaluating binary face templates generated from mobile-class CNNs, tailored for deployment on resource-constrained platforms such as smart cards. By combining lightweight embeddings with PCA–ITQ binarization, we addressed key trade-offs between verification accuracy, template size, and computational efficiency.

The proposed approach emphasizes practicality, with binary templates offering fast matching, compact storage, and acceptable stability for real-world biometric authentication. Our evaluation protocol, based on public datasets and transparent methodology, supports broader reproducibility and cross-dataset validation.

Beyond empirical performance, this benchmark contributes to the ongoing discourse on privacy-preserving and efficient AI-driven authentication. Future work will extend this framework to diverse populations and explore advanced hashing strategies, fostering more trustworthy and deployable biometric systems.

REFERENCES

1. Taigman Y, Yang M, Ranzato M, Wolf L. DeepFace: closing the gap to human-level performance in face verification. Proc. CVPR. 2014. DOI: 10.1109/CVPR.2014.220

2. Schroff F, Kalenichenko D, Philbin J. FaceNet: a unified embedding for face recognition and clustering. Proc. CVPR. 2015. DOI: 10.1109/CVPR.2015.7298682

3. Sun Y, Liang D, Wang X, Tang X. DeepID3: face recognition with very deep neural networks. arXiv preprint arXiv:1502.00873; 2015.

4. Drozdowski P, Struck F, Rathgeb C, Busch C. Benchmarking binarisation schemes for deep face templates. Proc. IEEE Int. Conf. on Image Processing (ICIP). 2018. DOI: 10.1109/ICIP.2018.8451291

5. Das A, Sengupta A, Saqib M, Pal U, Blumenstein M. More realistic and efficient face-based mobile authentication using CNNs. Proc. IJCNN. 2018. DOI: 10.1109/IJCNN.2018.8489070

6. Adebayo O, Boulgouris NV. Binary face representation for privacy-aware authentication. IEEE Trans. Inf. Forensics Security. 2012;7(6):1780–1785. DOI: 10.1109/TIFS.2012.2207726

7. Pandey R, Zhou Y, Urala Kota B, Govindaraju V. Deep secure encoding for face template protection. Proc. CVPR Workshops. 2016. DOI: 10.1109/CVPRW.2016.17

8. Mai G, Cao K, Lan X, Yuen PC. SecureFace: face template protection. IEEE Trans. Inf. Forensics Security. 2021;16:223–238. DOI: 10.1109/TIFS.2020.3009590

9. Parkhi OM, Vedaldi A, Zisserman A. Deep face recognition. Proc. BMVC. 2015. DOI: 10.5244/C.29.41

10. Tan M, Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks. Proc. ICML. 2019. DOI: 10.48550/arXiv.1905.11946

11. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C. MobileNetV2: Inverted residuals and linear bottlenecks. CVPR. 2018. DOI: 10.48550/arXiv.1801.04381

12. Ma N, Zhang X, Zheng H-T, Sun J. ShuffleNet V2: Practical guidelines for efficient CNN architecture design. ECCV. 2018. DOI: 10.1007/978-3-030-01264-9_8

13. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. arXiv preprint arXiv:1602.07360; 2016. DOI: 10.48550/arXiv.1602.07360

14. Gong Y, Lazebnik S, Gordo A, Perronnin F. Iterative Quantization: A Procrustean approach to learning binary codes. CVPR. 2013. DOI: 10.1109/CVPR.2013.214

15. Ricanek K, Tesafaye T. MORPH: A Longitudinal Image Database of Normal Adult Age-Progression. FGR. IEEE; 2006. DOI: 10.1109/FGR.2006.78

16. Liu W, Wen Y, Yu Z, Li M, Raj B, Song L. SphereFace: Deep hypersphere embedding for face recognition. CVPR. 2017. DOI: 10.1109/CVPR.2017.713

17. Deng J, Guo J, Xue N, Zafeiriou S. ArcFace: Additive angular margin loss for deep face recognition. IEEE TPAMI. 2022;44(10):5962–5979. DOI: 10.1109/TPAMI.2021.3084828

FUNDING

The authors did not receive funding for the development of this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTION

Conceptualization: Abdelilah Ganmati, Karim Afdel, Lahcen Koutti.
Data curation: Abdelilah Ganmati, Karim Afdel, Lahcen Koutti.
Formal analysis: Abdelilah Ganmati, Karim Afdel, Lahcen Koutti.
Research: Abdelilah Ganmati, Karim Afdel, Lahcen Koutti.
Methodology: Abdelilah Ganmati, Karim Afdel, Lahcen Koutti.
Project management: Abdelilah Ganmati, Karim Afdel, Lahcen Koutti.
Resources: Abdelilah Ganmati, Karim Afdel, Lahcen Koutti.
Software: Abdelilah Ganmati, Karim Afdel, Lahcen Koutti.
Supervision: Abdelilah Ganmati, Karim Afdel, Lahcen Koutti.
Validation: Abdelilah Ganmati, Karim Afdel, Lahcen Koutti.
Visualization: Abdelilah Ganmati, Karim Afdel, Lahcen Koutti.
Writing – original draft: Abdelilah Ganmati, Karim Afdel, Lahcen Koutti.
Writing – revision and editing: Abdelilah Ganmati, Karim Afdel, Lahcen Koutti.