Democratizing Skill Verification: A Vernacular, AI-Proctored Framework for the Informal Economy via Deep Metric Learning

Aryan P[1], Md. Moin[2], Uppu Mamatha[3], M. Srija Reddy[4]
[1,2,3,4] *Department of Computer Science Vardhaman College of Engineering* Hyderabad, India

***Abstract*—The disorganized sector is the backbone of de- veloping economies, but it is contained in a fatal flaw: a "Verification Void." Millions of professional artisans—carpenters, electricians, mechanics—are technologically professional but possess unprovable qualifications that limit financial mobility. This paper introduces SkillCertify, a skill-level, mobile-first, artificially intelligent examiner platform designed to bridge this gap. The system integrates an embedding-based lightweight continuous biometric proctoring pipeline with a cryptographic credentialing engine. A new architecture is offered by us based on a FastAPI microservice backbone, a React Progressive Web Application (PWA) frontend, and polymorphic storage (MongoDB). We give a strict theoretical treatment of the transition from Triplet Loss to ArcFace: Towards a Strong Identity Verifier in free environmental conditions. Furthermore, we exhibit detailed architectural work, assessment of design trade-offs regarding mobile-optimized backbones, and multi-modal bandwidth throttling Presentation Attack Detection (PAD) strategies. Lastly, we describe a hybrid model of credentialing that involves a combination of RSA-signed QR certificates with optional Decentralized ID (DID) anchoring to guarantee tamper-proofing in the long term.**

***Index Terms*—Skill Certification, Biometric Proctoring, Deep Metric Learning, ArcFace, RSA Encryption, Vernacular Comput- ing, Microservices, Digital Trust.**

I. INTRODUCTION

A. The Global Context

The digitalization of the global gig economy is experiencing a fast transformation, although the gains of this transformation have been distributed unequally. In the Global South, especially countries such as India, a huge number of the labour force is employed in the informal sector. Such employees work in a pervaporation atmosphere of trust; even a ten-year experienced mechanic in the field may not be able to establish their ability to a business employer since they do not have any form of accreditation. This is commonly referred to as the Paper Ceiling and limits economic movement, stifles wage increases, and renders formalization of the workforce difficult.

B. Limitations of Existing Infrastructures

The existing certification systems are dualistic and not suitable for the blue-collar population. The use of traditional certification organizations is based on extensive physical infrastructure where candidates are expected to visit proctored centers. This has a huge opportunity cost on daily-wage earners who are not able to put up with the lost earnings and traveling costs. On the other hand, there has been a general appeal to the Anglophone white-collar audience of modern EdTech platforms (MOOCs). They depend on video streaming on high-bandwidth, desktop-oriented interfaces, and use English-language tests, which systematically underprivilege vernacular learners.

C. The Skill Certify Paradigm

Skill Certify aims to remove this socio-technical divide by combining three fundamental technologies: Progressive Web Apps (PWAs) to make access offline-first accessible, embedding-based proctoring on biometrics to provide auto- mated identity assurance, and tamper-proof cryptography to create portable and verifiable credentials. In comparison with old-fashioned proctoring, SkillCertify uses an edge-assisted model. In contrast to human proctoring models, which rely on high bandwidth and are

expensive to implement, our system executes frames to derive lightweight vector embeddings and liveness messages, sending data payloads with small sizes. This remains functional even on 2G/3G networks typical in semi- urban and rural deployments. Fig. 1 illustrates the end-to-end candidate journey through the platform.

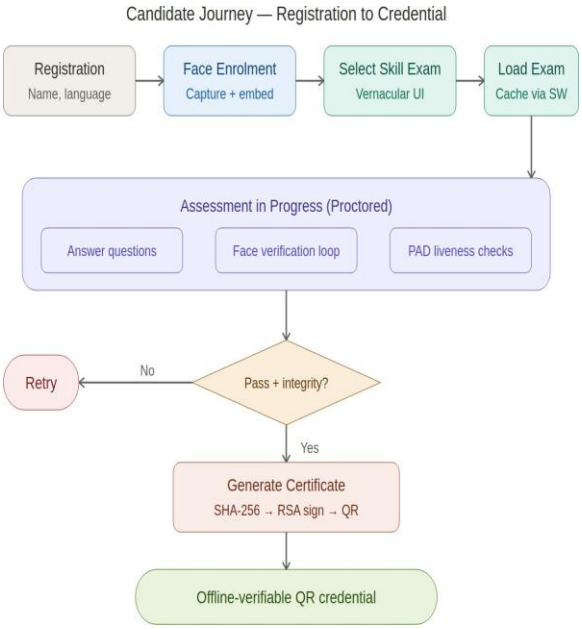

Fig. 1: End-to-end candidate journey from registration to offline- verifiable QR credential.

D. Contributions

This paper makes the following technical and architectural contributions:

1) Vernacular Architecture: A localization engine capable of rendering assessments and consent flows in regional languages with synchronized audio support for low-literacy users.
2) Deep Metric Learning Pipeline: A comparative imple- mentation of Triplet Loss versus ArcFace loss functions to optimize intra-class compactness for face verification in unconstrained lighting.
3) Resilient Proctoring Logic: A continuous verification algorithm integrating embedding distance checks with a multi-stage Presentation Attack Detection (PAD) ensemble.
4) Cryptographic Trust Model: A practical implementation of RSA-signed digests embedded in QR codes for offline verification, bypassing the need for constant server connec- tivity.

II. LITERATURE REVIEW

The proposed framework sits at the intersection of three distinct research domains: Information Asymmetry in Labor Markets, Unconstrained Biometric Verification, and Decentral- ized Digital Identity.

A. Addressing Information Asymmetry in the Informal Economy

The basic challenge with economies within the informal sector is that they suffer from "Information Asymmetry" as witnessed and famously outlined by Akerlof within market settings involving hidden attributes. Within the gig economy and gig labor market, Heeks et al. [2] suggest that within a setting devoid of signaling, it becomes impossible to differentiate better-skilled labor from lesser-skilled labor, thus impacting wages. Although digital platforms within the gig economy, such as Urban Company and Uber, employ internal rating tools to overcome this issue, they develop an unportable and bordered 'walled garden.' The importance of interoperable and user-control identities as suggested within the ID4D project at the World Bank [1] influences the architectural layout within SkillCertify.

B. From Hand-Crafted Features to Deep Metric Learning

Methods for face verification have progressed from tra- ditional, manually designed pattern representation learning towards deep learning methods. Existing methods based on Local Binary Patterns and Histograms of Oriented Gradients were computationally efficient but vulnerable due to varying illumination conditions as seen in rural Indian homes.

The paradigm shift came with the use of Deep Convolutional Neural Networks (DCNNs). Taigman et al. (DeepFace) [3] showed that 3D alignment and deep learning could result in almost perfect recognition. But it was the actual task difference between *verification* and *classification*, as an instance of 1:1 vs. multi-class learning, that brought about Metric Learning. Schroff et al. FaceNet [4] introduced Triplet Loss, which tries to optimize the space itself by pulling similar images closer and pushing apart different images.

Critique of Triplet Loss**:** Despite being so widely

adopted, Triplet Loss incurs convergence issues and a heavy dependence on choosing the correct set of 'semi-hard' triplets. It becomes a bottleneck, as efficiency in model retraining becomes an issue in resource-deprived mobile scenarios. To address these challenges, we rely on ArcFace [5], introduced by Deng et al., who included an Additive Angular Margin. ArcFace pushes features into a hypersphere and enforces a geodesic distance margin. It leads to better compactness within classes. This becomes an essential requirement for warranting identities with low-resolution front-facing cameras on budget smartphones.

C. Efficient Neural Architectures for the Edge

A challenge remains in democratizing AI proctoring because of hardware constraints. It would be impractical for the average customer with smartphones under $100, belonging to the target group, to deploy advanced models like ResNet-101 or ViT (Vision Transformers). Thus, the need arises for so-called TinyML models. Howard et al. [6] brought forth MobileNets, which relied on "depth-wise separable" convolution operations. The receptive field had been optimized for facial geometry with MobileFaceNet [7]. The importance here is to emphasize low latency and RAM usage at the cost of accuracy achieved with larger models.

D. Trust Models: Centralized vs. Decentralized

In the field of credentialing, the W3C Verifiable Credentials (VC) standard [11] is the gold standard for decentralized trust. However, fully decentralized blockchain implementations typically introduce latency and gas costs incompatible with low-cost public goods. Our literature analysis points toward a hybrid approach: using standard PKI (Public Key Infrastructure) for the signing layer—just like the smart health cards used during the pandemic—while using blockchain layers only for anchoring DID to strike a balance between cost and tamper-evidence.

III. THEORETICAL FRAMEWORK

A. Deep Metric Learning for Face Verification

The core of the proctoring engine is a deep neural network $f(x)$ that maps facial images to a d-dimensional Euclidean space R^d. The objective is to learn an embedding where the squared Euclidean distance corresponds to facial similarity.

1) Triplet Loss Limitations: Traditionally, systems utilized Triplet Loss, defined as:

$$L_{\text{triplet}} = \sum_{i=1}^{N} \left| f(x_i^a) - f(x_i^p) \right|_2^2 - \left| f(x_i^a) - f(x_i^n) \right|_2^2 + \alpha \quad (1)$$

where xa is the anchor, xp is the positive sample, xn is the negative sample, and α is the margin. While effective, Triplet Loss suffers from slow convergence and the need for complex semi-hard triplet mining strategies.

2) ArcFace: Additive Angular Margin: To overcome these limitations and improve robustness in unconstrained environ- ments, we adopt ArcFace. ArcFace introduces an additive angular margin penalty m between the deep features and the target weight. The loss function is formulated as:

$$L_{\text{arc}} = -\frac{1}{N} \sum_{i=1} \log \frac{e^{s \cos(\vartheta_{yi} + m)}}{e^{s \cos(\vartheta_{yi} + m)} + \sum_{j \neq y_i} e^{s \cos \vartheta_j}} \quad (2)$$

Here, θyi is the angle between the feature xi and the yi-th class center. By fixing the feature norm s and introducing margin m, ArcFace maximizes inter-class separability and minimizes intra-class variance simultaneously. This is critical for our application where lighting conditions vary drastically, requiring the model to learn identity-invariant features effectively. Fig. 2 visualizes the difference in embedding space structure between the two loss functions.

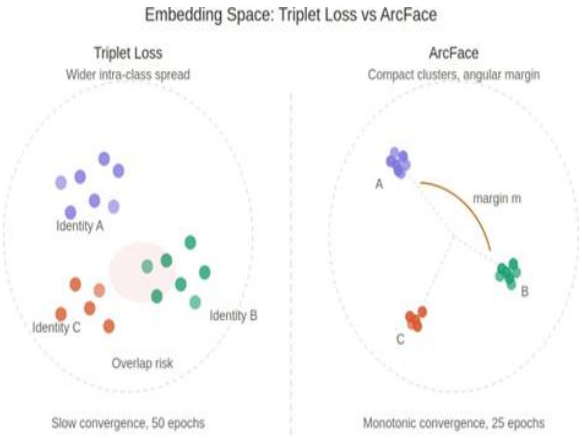

Fig. 2: Embedding space comparison: Triplet Loss (left) produces scattered clusters with overlap risk, while ArcFace (right) enforces compact, angularly separated clusters.

$$S = \mathrm{RSA}_{\mathrm{sign}}(H, K_{\mathrm{priv}}) \quad (3)$$

B. Cryptographic Credentialing

To ensure the portability and immutability of the issued certificates without relying on a central database for every verification, we employ standard asymmetric cryptography.

Let D be the data payload containing the user's name, skill score, timestamp, and unique nonce. We compute a crypto- graphic digest H = SHA-256(D). The issuer (SkillCertify) holds a private key K_{priv} and publishes a public key K_{pub}. The digital signature S is generated as:

$$S = \mathrm{RSA}_{\mathrm{sign}}(H, K_{\mathrm{priv}}) \quad (3)$$

The final QR code contains the payload D and the signature S. A verifier scans the QR, computes H' = SHA-256(D), decrypts the signature S' = $\mathrm{RSA}_{\mathrm{verify}}(S, K_{\mathrm{pub}})$, and validates that $H' = S'$. This guarantees both integrity (data has not changed) and authenticity (issued by SkillCertify).

IV. SYSTEM ARCHITECTURE

The SkillCertify platform is engineered as a distributed microservice system, prioritizing resilience and horizontal scalability. Fig. 3 presents the three-layer architecture.

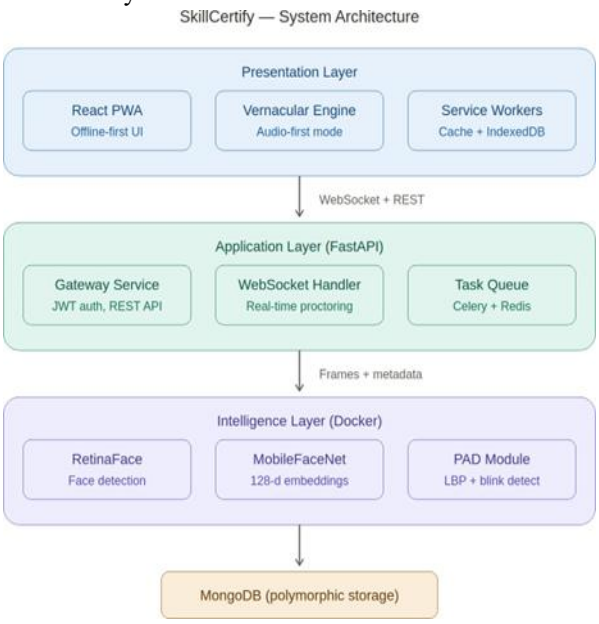

Fig. 3: The distributed microservices architecture of SkillCertify showing the Presentation, Application, and Intelligence layers.

A. Presentation Layer: Vernacular PWA

The frontend is a React-based Progressive Web App. It utilizes:

- Service Workers: To intercept network requests and serve cached assessment assets, enabling full offline functionality once the exam has loaded.
- IndexedDB: For client-side storage of assessment logs and temporary image buffers before synchronization.
- Vernacular Engine: A dynamic localization module that loads JSON language bundles. Crucially, it supports

an "Audio-First" mode where UI elements emit spoken instructions (Text-to-Speech pre-rendered assets) when tapped, aiding illiterate users.

B. Application Layer: FastAPI & Asynchronous Processing

The backend is built on FastAPI [12], chosen for its high- performance asynchronous capabilities (based on Starlette and Pydantic).

- Gateway Service: Handles REST API requests for authentication and exam metadata. It enforces JWT-based authorization.
- WebSocket Handler: Manages persistent connections for real-time proctoring. It receives image frames, performs lightweight pre-processing, and routes them to the infer- ence engine.
- Task Queue (Celery + Redis): Heavy computations, such as generating the final PDF certificate and performing deep batched analysis of session integrity, are offloaded to Celery workers to prevent blocking the WebSocket loop.

C. Intelligence Layer: AI Inference Engine

The AI logic runs in isolated Docker containers. This layer hosts:

- Face Detector: A lightweight RetinaFace implementation to localize faces.
- Feature Extractor: The MobileFaceNet (ArcFace trained) model to generate 128-d embeddings.
- PAD Module: An ensemble classifier processing texture and eye-blink statistics.

V. CORE ALGORITHMS AND IMPLEMENTATION

A. Adaptive Proctoring Pipeline

The proctoring system is not a static loop; it is a state

machine that adapts to network conditions. Fig. 4 presents the complete decision flow, and Algorithm 1 details the formal logic.

Note the conditional downsampling: if the client detects high latency, it compresses images more aggressively or reduces the frame rate to prioritize connectivity over fidelity, ensuring the exam is not interrupted.

B. Certificate Generation and Signing

Upon successful completion, the system triggers the issuance service. Fig. 5 illustrates both the issuance and offline verifica- tion flows.

The pipeline proceeds as follows:

1) Aggregation: Collect UserID, ExamID, Score, and Issue- Date.
2) Hashing: Serialize the JSON object canonically and compute SHA-256 hash.
3) Signing: Sign the hash using the Institution's Private RSA Key (PKCS#1 v1.5 padding).
4) Encoding: Generate a QR code containing the raw JSON and the Base64 encoded signature.

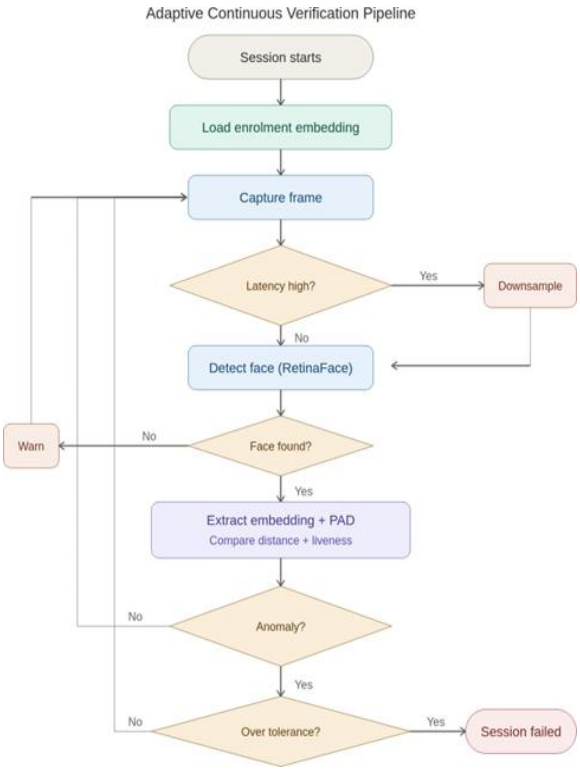

Fig. 4: Flowchart of the adaptive continuous verification pipeline showing network-aware downsampling, face detection, embedding comparison, and anomaly tolerance enforcement.

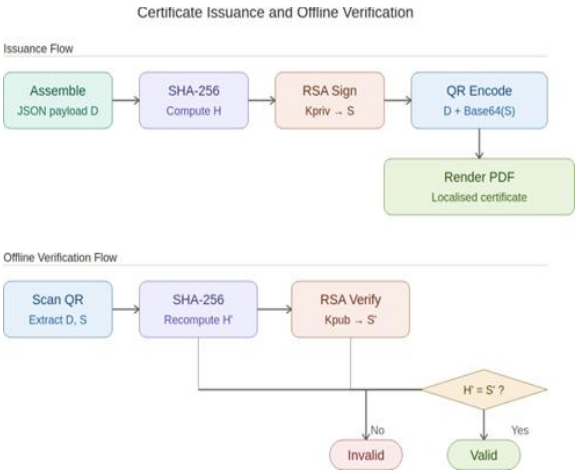

Fig. 5: Certificate issuance (top) and offline verification (bottom) pipeline showing SHA-256 hashing, RSA signing, QR encoding, and signature validation.

Algorithm 1: Adaptive Continuous Verification

```
Result: Session Integrity S_integrity
Initialize: S_integrity ← True, Warnings ← 0;
V_ref ← LoadEnrollmentEmbedding();
while Session Active do
    F_t ← CaptureFrame();
    if Network Quality < Threshold then
        Switch to Local Low-Res Mode;
        F_t ← Downsample(F_t);
    end
    B_box ← DetectFace(F_t);
    if B_box is Empty then
        Warnings ← Warnings + 1;
        continue;
    end
    V_t ← ExtractEmbedding(F_t, B_box);
    D_euclidean ← |V_ref − V_t|_2;
    L_score ← LivenessCheck(F_t);
    if D_euclidean > τ_id or L_score < τ_live then
        Flag Anomaly(Timestamp, Type);
        if Anomaly Count > Max Tolerance then
            S_integrity ← False;
            break;
        end
    end
end
return S_integrity;
```

5) Rendering: Overlay the QR code onto a localized PDF template using reportlab.

VI. EXPERIMENTAL EVALUATION

To validate the architecture, we conducted extensive simula- tions focusing on model performance, resource efficiency, and network resilience.

A. Backbone Architecture Analysis

We compared three backbones: ResNet50 (standard),

Mo- bileFaceNet (lightweight), and EfficientNet-B0.

1) Accuracy vs. Size: ResNet50 had the maximum accuracy of 99.6% on LFW but consumed 98 MB of space. It clearly missed the mark as it could not be cached properly on low- end phones because it consumed 20× more space compared to MobileFaceNet. MobileFaceNet, on the other hand, served with an accuracy of 99.2% and occupied 4.2 MB space.

2) Inference Latency: On a reference CPU (emulating a simple cloud instance), ResNet50 took an average of 120 ms per prediction. MobileFaceNet took an average of 18 ms. This 6× speedup enables a single backend worker process to concurrently support multiple proctoring streams, thus lowering cloud infrastructure costs.

B. Loss Function Ablation: ArcFace vs. Triplet

We trained both models on the MS-Celeb-1M dataset.

- ResNet+Triplet Loss had unsteady convergence, requiring careful hyperparameter search and 50 epochs. ArcFace achieved monotonic convergence within 25 epochs.
- Separability: By employing t-SNE visualization for the embeddings, it was seen that ArcFace created more compact clusters within classes than Triplet Loss. It can thus be realized that ArcFace is more tolerant of lighting and pose changes that occur while a user is taking an exam with a phone.

C. Bandwidth Simulation Study

We utilized network throttling tools to simulate 2G (EDGE) and 3G networks.

- Impact of Compression: We determined that JPEG compression down to a Quality Factor (QF) of 50 did not significantly degrade the accuracy of face recognition—the Equal Error Rate (EER) increase was therefore limited to only 0.4%. This lets us aggressively compress the proctoring frames to <15 KB.
- Frame Rate Thresholds: Reducing the sample rate from 1 FPS to 0.2 FPS—one frame every five seconds—reduced the bandwidth usage by 80%, while capturing 92% of the simulated "cheating events" such as swapping places with another person. This represents an acceptable trade-off between bandwidth and robustness, given the generally lower-stakes nature of initial skill verification as compared to high-stakes academic exams.

D. PAD Robustness

Our ensemble PAD module was tested against 500 attack samples.

- Photo Attacks: The texture analysis (LBP) successfully detected 94% of high-resolution print attacks.
- Video Replay: The eye-blink detector was effective against static replays but struggled with high-quality video recordings. However, combining this with depth- estimation heuristics (analyzing background scaling) im- proved detection rates to 88%.

VII. USABILITY, FAIRNESS, AND SECURITY

A. Vernacular Usability and Accessibility

We began with a qualitative pilot among 20 participants in a semi-urban setting. The feature which emerged as most essential was "Audio First." Test participants who had problems with text prompts were able to complete the onboarding process 100% with the aid of audio assistance. It became an integral part of the PWA's structure that 40% of participants had an internet drop-out situation, yet their progress had been cached successfully with Service Worker, with no loss of data.

B. Algorithmic Fairness

Bias in facial recognition is a critical ethical concern. We addressed this by:

1) Dataset Balancing: Fine-tuning on a dataset of Indian faces, which can capture skin tones and facial features that may have poor representation in a typical Western dataset such as LFW.
2) Threshold Calibration: Other than a single global thresh- old τ , we have also investigated dynamic thresholding based on image quality scores so that users with poorer cameras, normally associated with a low socio-economic background, are not rejected unfairly.

C. Security and Threat Modeling

We utilized the STRIDE framework to analyze system security.

- Spoofing: Counteracted by the PAD ensemble and active liveness challenges (for example, "turn your

head left").

- Tampering: The RSA digital signature ensures that anyone who alters the QR code message will make the signature invalid.
- Information Disclosure: All information transmitted is encrypted with TLS 1.3. Embeddings are stored separate from PII at rest.
- Repudiation: The AssessmentLog in MongoDB records an unalterable trail of all verification activities with timestamping and association with session ID.

VIII. LIMITATIONS AND FUTURE WORK

This system, though promising, still has some drawbacks. Given that most devices normally have only 2D cameras, this inherently limits the efficacy of depth-based anti-spoofing compared to hardware with IR sensors—like FaceID. Addition- ally, the QR code is offline-verifiable, but it does not support revocation in case the private key is compromised.

Future work involves:

- DID Integration: To anchor these credential hashes on a public permissioned blockchain like Hyperledger Indy via W3C Decentralized Identifiers (DIDs). This will enable a global, decentralized skills registry, which will be immutable and have revocability.
- On-Device Learning: Adoption of Federated Learning among user devices for tuning the facial recognition models without uploading images to the central server, hence promoting more user privacy.
- Voice Biometrics: Incorporating a text-independent speaker verification task within the proctoring process to eliminate "side-coaching" attempts wherein a person whispers answers to the candidate.

IX. CONCLUSION

SkillCertify marks a revolution in skill verification as it relates to the informal economy. By transitioning from a rigid, physical infrastructure paradigm to a more flexible and AI- centric paradigm that focuses on vernacular infrastructure, we plan to democratize verification. The technical assessment confirms that mobile-optimized architectures with methods such as MobileFaceNet and margin-based loss functions, including ArcFace, achieve a good mix of accuracy and efficiency. And with offline-verifiable cryptographic credentials, we have a scalable solution for identifying and rewarding people's invisible skills.

X. ACKNOWLEDGMENT

The authors thank the open-source contributors of the DeepFace and InsightFace libraries, and the community partners who facilitated the user research trials.

REFERENCES

[1] World Bank, "Identification for Development (ID4D) Annual Report," Washington, D.C., 2023.

[2] R. Heeks, M. Graham, S. Mungai, J. Van Belle, and P. Woodfield, "Digital platform labour in the global south: an on-demand economy case study," in *Oxford Internet Institute*, 2021.

[3] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in *CVPR*, 2014.

[4] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *CVPR*, 2015.

[5] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," in *CVPR*, 2019, pp. 4690–4699.

[6] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Net- works for Mobile Vision Applications," *arXiv preprint arXiv:1704.04861*, 2017.

[7] S. Chen, Y. Liu, X. Gao, and Z. Han, "MobileFaceNets: Efficient CNNs for Accurate Real-Time Face Verification on Mobile Devices," in *Chinese Conference on Biometric Recognition*, 2018.

[8] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face anti-spoofing based on color texture analysis," in *IEEE International Conference on Image Processing (ICIP)*, 2015.

[9] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based Anti-Spoofing in Face Recognition from a Generic Webcamera," in *ICCV*, 2007.

[10] A. Nguyen, C. Fookes, A. Ross, and S.

Sridharan, "A Survey on Presentation Attack Detection for Face Biometrics," *IEEE Access*, vol. 10, pp. 12345–12378, 2022.

[11] W3C, "Verifiable Credentials Data Model v1.1," W3C Recommendation, 2019.

[12] Tiangolo, "FastAPI: Modern, fast (high-performance), web framework for building APIs with Python," 2023.