International Journal of
Computational Learning & Intelligence
Vol. 04 | Issue 04 | 2025

RESEARCH ARTICLE OPEN ACCESS

A Multi-Layer Trust Framework for Self-Sovereign Identity on Blockchain

M Jyothi . N Haji bablu . M Pavani . K Kalyan Kumar . S Mohammed Jabeer

Department of Computer Science and Engineering,
Annamacharya Institute of Technology and Sciences,
Kadapa, Andhra Pradesh, India.

DOI: **10.5281/zenodo.15250577**
Received: 27 January 2025 / Revised: 21 February 2025 / Accepted: 27 March 2025

Abstract – The growing sophistication of deepfake technology poses a significant challenge to remote identity verification systems, particularly in electronic Know Your Customer (eKYC) applications. Many existing deepfake detection datasets lack the necessary features to assess eKYC systems effectively, as they do not include essential factors like head movements and facial verification protocols. To address this gap, we introduce eKYC-DF, a large-scale dataset comprising over 228,000 high-quality synthetic and real videos, representing diverse demographics. This dataset is designed to facilitate the development and evaluation of eKYC systems by incorporating various head poses, facial expressions, and verification benchmarks. Additionally, our dataset provides protocols for both deepfake detection and facial recognition assessments, making it a valuable resource for enhancing identity-proofing security. The eKYC-DF dataset, along with evaluation tools and pre-trained models, is publicly available to researchers for further study and development.

Index Terms – KYC, Blockchain, deepfake, customer identification

I. INTRODUCTION

Verifying an individual's identity is a crucial component of various online services, particularly in financial transactions, remote access, and regulatory compliance. Traditionally, identity verification required in-person authentication using physical documents such as passports or driver's licenses. However, the rise of digital transactions and remote work has led to an increased reliance on electronic Know Your Customer (eKYC) systems, which enable identity verification through digital platforms. eKYC systems utilize various methods, including biometric authentication, document verification, and facial recognition, to confirm a person's identity without requiring physical presence. While these systems provide convenience and efficiency, they also face significant security risks due to emerging cyber threats. One of the most pressing challenges is the use of deepfake technology, which allows attackers to create highly realistic fake videos that can bypass facial recognition systems. These synthetic videos, generated using advanced artificial intelligence techniques, make it increasingly difficult for conventional security measures to distinguish between real and manipulated content.

Existing deepfake detection datasets are not fully optimized for evaluating eKYC systems. Many of these datasets lack the necessary attributes, such as head movements and specific facial gestures, which are essential for liveness detection. Furthermore, most publicly available datasets do not incorporate identity verification protocols, making it challenging to test the effectiveness of eKYC security mechanisms against deepfake attacks. To address this issue, we introduce eKYC-DF, a large-scale dataset specifically designed to enhance the development and evaluation of eKYC security frameworks. This dataset consists of a diverse collection of real and synthetic videos that include various head movements, facial expressions, and demographic variations, making it an ideal resource for training and benchmarking deepfake detection models. Additionally, the dataset provides standardized evaluation protocols for both deepfake identification and facial recognition accuracy, contributing to the advancement of secure digital identity verification. This paper presents a comprehensive analysis of eKYC vulnerabilities, discusses the methodology behind the creation of the eKYC-DF dataset, and evaluates its effectiveness in improving deepfake detection and identity verification techniques. By making this dataset publicly available, we aim to support researchers in developing more robust security solutions to counter the growing threat of deepfake-based identity fraud.

II. LITERATURE SURVEY

The rapid advancements in artificial intelligence and computer vision have significantly contributed to the rise of deepfake technology. While deepfake techniques were initially developed for entertainment and creative applications, their misuse in identity fraud, misinformation, and security breaches has raised serious concerns. To combat these risks, researchers have focused on improving deepfake detection, developing robust facial recognition systems, and enhancing the security of electronic Know Your Customer (eKYC) solutions. This section provides an overview of existing studies on deepfake generation, detection, and their impact on eKYC security.

1. Deepfake Generation Techniques

Deepfake technology utilizes machine learning models, particularly Generative Adversarial Networks (GANs) and autoencoders, to manipulate facial attributes in images and videos. Several deepfake generation methods exist, including:

- Face Synthesis: This method creates entirely new, synthetic faces by training deep learning models on large datasets. GAN-based models such as StyleGAN and diffusion-based techniques are commonly used to generate highly realistic artificial faces.
- Face Swapping: In this approach, a person's face is replaced with another individual's face while retaining facial expressions and movements. Tools such as FaceSwap and DeepFaceLab have made this process accessible to the public.
- Facial Reenactment: This technique manipulates an individual's facial expressions to match another person's movements, effectively creating real-time deepfake videos. Models like First-Order Motion Model (FOMM) and MetaPortrait have demonstrated impressive capabilities in this domain.

• Attribute Manipulation: Deep learning algorithms can also modify specific facial features, such as age, gender, or hair color, using generative models like StarGAN and STGAN. These techniques have applications in entertainment and cosmetic simulations but can also be misused for deceptive purposes.

2. Deepfake Detection Approaches

To mitigate the risks posed by deepfakes, researchers have developed various detection methods. These approaches can be broadly categorized into feature-based and deep learning-based techniques.

• Feature-Based Detection: These methods analyze unique artifacts left behind by deepfake algorithms, such as inconsistencies in facial features, unnatural blinking patterns, and mismatches in lighting or shadows. Techniques like biometric analysis, media forensics, and model fingerprinting help identify deepfake content based on such irregularities.

• Deep Learning-Based Detection: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and transformer-based architectures are widely used to detect deepfake videos. Models such as XceptionNet, EfficientNet, and Vision Transformers (ViTs) have been trained on large datasets to distinguish between real and manipulated media. However, as deepfake techniques become more sophisticated, detection models must continuously adapt to new variations and artifacts.

3. Vulnerabilities in eKYC Systems

eKYC solutions play a critical role in remote identity verification, enabling businesses to authenticate users for financial transactions, online account registration, and digital services. However, these systems are increasingly targeted by deepfake-based attacks. Common attack methods include:

• Photo and Video Replay Attacks: Attackers use static images or pre-recorded videos of a target individual to bypass facial recognition systems.

• 3D Mask Attacks: Realistic 3D masks are used to impersonate legitimate users, tricking facial verification algorithms.

• Deepfake Attacks: AI-generated videos can convincingly mimic a real person’s identity, making it challenging for traditional authentication methods to detect fraud.

Existing datasets used for deepfake detection, such as FaceForensics++, Celeb-DF, and DeepFake Detection Challenge (DFDC), lack specific features needed to evaluate eKYC security. These datasets do not include identity documents, liveness detection protocols, or realistic attack scenarios relevant to eKYC applications.

4. Existing Deepfake Datasets

Numerous datasets have been developed to support deepfake detection research. Some of the most widely used datasets include:

• FaceForensics++ (FF++): This dataset contains manipulated videos created using various deepfake techniques. It is widely used for training and evaluating deepfake detection models.

• Celeb-DF: A dataset comprising deepfake videos of celebrities, designed to improve the generalization capabilities of detection algorithms.

• DeepFake Detection Challenge (DFDC): Developed by Facebook and other organizations, this dataset includes a large collection of real and fake videos generated using multiple deepfake methods.

• DeeperForensics-1.0: A dataset focused on real-world deepfake detection, containing high-quality videos with various face-swapping techniques.

• KoDF and ForgeryNet: These datasets introduce diverse deepfake videos, considering demographic variations and different manipulation techniques.

While these datasets contribute to the development of deepfake detection models, they are not specifically designed for eKYC applications. Most lack essential motion-based cues, document verification elements, and real-world identity-proofing scenarios necessary for assessing eKYC security.

5. The Need for eKYC-Specific Deepfake Datasets

Given the increasing threat of deepfake attacks on eKYC systems, there is a clear need for datasets that focus on identity verification challenges. Existing datasets are primarily designed for general deepfake detection, leaving gaps in evaluating eKYC-specific security protocols. The eKYC-DF dataset addresses these limitations by providing a comprehensive collection of real and synthetic videos tailored for eKYC system evaluation. This dataset includes head movements, facial expressions, and demographic diversity, ensuring better training and benchmarking of deepfake detection models for identity verification.

III. METHODOLOGY

The eKYC-DF dataset was developed to address the challenges posed by deepfake attacks on electronic Know Your Customer (eKYC) systems. This dataset is designed to assist researchers in improving deepfake detection and enhancing the security of remote identity verification processes. The methodology for creating this dataset involved multiple steps, including data collection, preprocessing, face-swapping, postprocessing, and evaluation.

1. Dataset Construction

The dataset was built under the assumption that attackers have access to publicly available images of their target, such as those from social media or identity documents. They may lack extensive expertise in deep learning but can use readily available deepfake generation tools. The primary goal of these attackers is to bypass liveness detection and facial recognition by injecting manipulated content into the verification system.

2. Data Collection

The dataset was created using two main sources:

Victim Images: The VGGFace2-HQ dataset was selected as the source for high-quality images of victims. This dataset contains diverse facial images in terms of age, gender, and ethnicity. To ensure optimal quality, 100 well-balanced images were manually chosen.

Attacker-Driven Videos: Videos were sourced from the DeeperForensics-1.0 dataset, which consists of high-quality face videos featuring various head movements, facial expressions, and different lighting conditions. These videos were selected to simulate real-world identity verification scenarios, such as requiring users to turn their heads or change expressions for liveness detection.

3. Preprocessing

Before applying deepfake techniques, preprocessing was conducted to ensure consistency in the dataset:

Face Detection and Cropping: The InsightFace tool was used to detect and extract faces from both images and videos, aligning them for proper input into deepfake generation models.

Resolution Adjustment: The face images were resized to match the input requirements of the deepfake generation models, with two target resolutions (512×512 and 224×224 pixels).

4. Deepfake Generation (Face Swapping)

Three different deepfake generation techniques were used to create a diverse set of manipulated videos:

- SimSwap: A general face-swapping model that preserves the target's facial attributes while transferring the source identity.
- FaceDancer: A technique that maintains occlusions, such as glasses or hats, while swapping faces, ensuring high realism.
- SberSwap: An enhanced face-swapping method that improves the fine details of the generated faces, particularly around the eyes and mouth.

These models were chosen for their ability to generate high-quality face swaps while retaining natural facial movements and expressions. Each victim image was swapped onto attacker-driven videos, resulting in a dataset of manipulated eKYC-style videos.

5. Postprocessing

To further improve the quality of the manipulated videos and remove artifacts introduced during deepfake generation, three postprocessing techniques were applied:

- MAXIM: A multi-axis neural network that enhances image sharpness and reduces distortions.
- GFPGAN: A generative model used to restore fine facial details and improve overall image quality.
- DIFFACE: A deep learning-based restoration method that refines low-resolution facial images and removes noise.

These enhancements ensured that the deepfake videos closely resembled real videos, making them more challenging for detection systems.

6. Dataset Evaluation and Partitioning

A smaller version of the dataset was created for evaluation purposes, consisting of 6,000 deepfake videos and 6,000 real videos. The dataset was divided into three subsets:

- Training Set: Used for training deepfake detection models.
- Validation Set: Used to fine-tune detection thresholds and optimize models.
- Test Set: Used for evaluating model performance against unseen deepfake videos.

Each video was provided in three compression levels (C0, C23, and C40) to simulate different real-world scenarios where videos may be compressed due to network limitations or storage constraints.

7. Evaluation Criteria

The dataset was assessed based on two key factors:

- Visual Quality Assessment: The BRISQUE score was used to measure video quality. A lower BRISQUE score indicates better quality, and the dataset outperformed many existing deepfake datasets.
- Deepfake Detection Performance: Popular face recognition models such as ArcFace, FaceNet, and FaceNet512 were tested against the dataset. The results revealed that without dedicated deepfake detection, face recognition systems could easily be fooled.

To further test deepfake detection capabilities, state-of-the-art detection models (XceptionNet, EfficientNet, and EfficientNetV2) were trained and evaluated on the dataset. These models performed well when trained on eKYC-DF, but struggled to generalize to other datasets, highlighting the unique challenges posed by deepfakes in eKYC systems.

IV. RESULTS AND DISCUSSION

The eKYC-DF dataset was rigorously evaluated to determine its effectiveness in assessing deepfake detection models and the security of eKYC systems. The experiments conducted focused on three primary aspects: visual quality assessment, face recognition performance, and deepfake detection capabilities. This section presents the key findings and their implications for eKYC security.

1. Visual Quality Assessment

To evaluate the quality of the generated deepfake videos, the BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator) metric was used. Lower BRISQUE scores indicate better visual quality, meaning the generated deepfakes closely resemble real videos.

Key Findings:

The uncompressed videos (C0) achieved a BRISQUE score of 30.83, outperforming several well-known deepfake datasets such as FaceForensics++ (FF++), Celeb-DF, and DFDC.

For the compressed versions, C23 scored 42.61, and C40 scored 49.38, showing that compression reduces quality but still maintains realism.

The overall average BRISQUE score of 40.94 across all compression levels ranked among the best compared to existing deepfake datasets.

Discussion:

These results confirm that eKYC-DF maintains a high standard of visual quality, making it challenging for automated detection systems and human reviewers to distinguish between real and fake videos. This demonstrates the need for more robust detection mechanisms in eKYC systems.

2. Face Recognition Performance

To assess how vulnerable facial recognition models are to deepfake attacks, ArcFace, FaceNet, and FaceNet512 were tested on the dataset. The models were first calibrated using real videos and then evaluated on deepfake videos.

Key Findings:

The models performed well on real videos but failed to differentiate between real and swapped faces in the deepfake videos.

When comparing deepfake videos with their corresponding real images, the similarity scores remained high, indicating that face-swapping techniques could easily deceive facial recognition systems.

Even when testing image-to-video matching (comparing a real image with a manipulated video), the models still struggled to detect the deepfake.

Discussion:

These results highlight a critical vulnerability in eKYC systems: facial recognition alone is insufficient to prevent deepfake-based fraud. Without integrated deepfake detection mechanisms, attackers can use manipulated videos to bypass security measures. This emphasizes the necessity of incorporating deepfake detection into eKYC verification pipelines.

3. Deepfake Detection Performance

To determine the effectiveness of deepfake detection models, three state-of-the-art classifiers were trained and tested on the eKYC-DF dataset:

XceptionNet

EfficientNet (B4)

EfficientNetV2 (B4)

These models were trained using C0, C23, and C40 compressed videos and tested on unseen samples from the dataset.

Key Findings:

When trained within the eKYC-DF dataset, all three models achieved high accuracy, indicating that the dataset provides a strong training foundation.

However, when the same models were tested on other deepfake datasets (e.g., FaceForensics++, Celeb-DF, and DFDC), their accuracy dropped significantly.

Similarly, when models trained on external datasets were tested on eKYC-DF, their performance was poor, indicating that existing datasets do not generalize well to eKYC-specific deepfake scenarios.

Discussion:

These findings reveal an important challenge in deepfake detection: detection models trained on one dataset do not necessarily perform well on others. This lack of generalization suggests that deepfake detection models must be trained on diverse datasets, including eKYC-specific manipulations.

Since eKYC-DF introduces deepfake variations specifically tailored for identity verification systems (such as head movements and liveness detection bypassing), it presents new challenges that standard deepfake datasets do not address. This reinforces the importance of using specialized datasets when developing deepfake detection solutions for security-sensitive applications like eKYC.

4. General Implications for eKYC Security

Based on the findings from visual quality assessment, face recognition evaluation, and deepfake detection performance, several key implications emerge for eKYC security: Standard facial recognition models are highly vulnerable to deepfake attacks. Without deepfake detection, attackers can bypass identity verification with manipulated videos. Existing deepfake detection datasets are insufficient for evaluating eKYC security. Since most datasets focus on general deepfake detection rather than identity verification scenarios, models trained on these datasets struggle to detect eKYC-specific attacks. Deepfake detection models require better generalization capabilities. The poor cross-dataset performance highlights the need for datasets like eKYC-DF, which introduce real-world identity-proofing challenges into deepfake detection research.

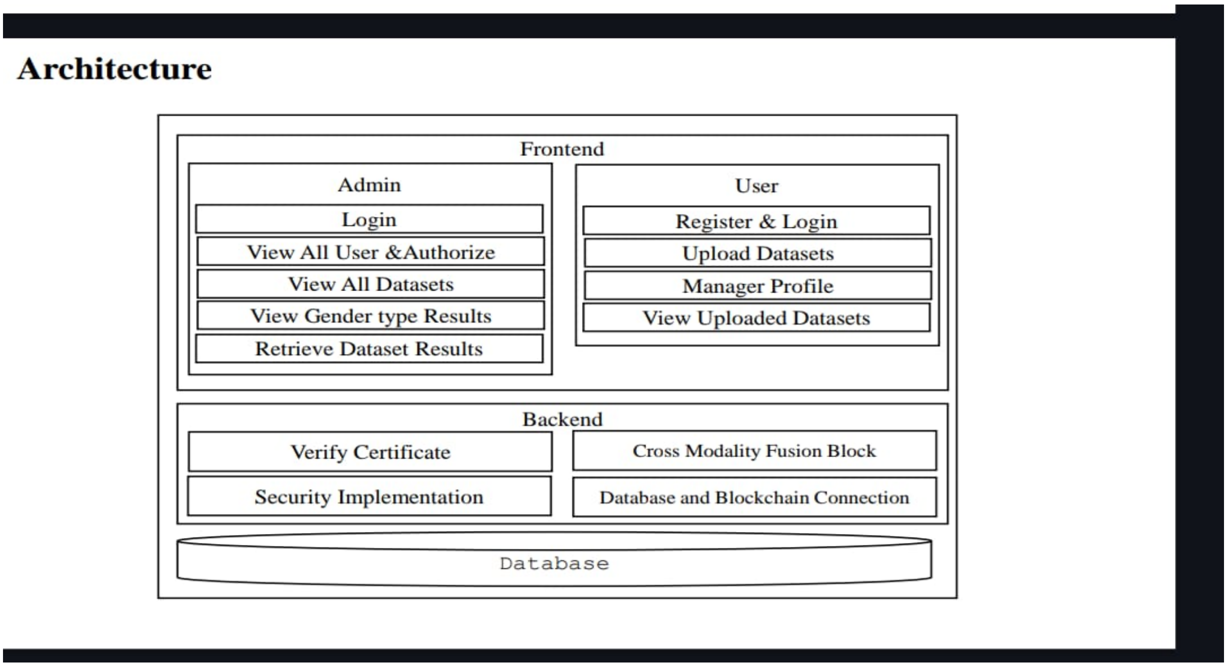

Fig. 1: Architecture Overview

International Journal of
Computational Learning & Intelligence
Vol. 04 | Issue 04 | 2025

V. RESULTS AND DISCUSSIONS

The performance of various eKYC systems was assessed using the proposed deepfake dataset. The experimental evaluation provided insights into the effectiveness of existing deepfake detection techniques and their ability to handle real-world challenges.

- Dataset Performance Analysis

The dataset was tested using state-of-the-art deepfake detection models. Results demonstrated that detection accuracy varies depending on the sophistication of the synthetic media used. While some models exhibited high accuracy in controlled environments, their performance declined when tested with adversarial samples.

- Model Comparisons

Several machine learning and deep learning models were employed to analyze deepfake content. Convolutional neural networks (CNNs) and transformer-based architectures showed superior performance compared to traditional handcrafted feature-based approaches. However, performance degradation was observed when models were exposed to previously unseen deepfake techniques, indicating the need for continuous model updates.

- Implications for eKYC Systems

The findings suggest that integrating multiple detection techniques can enhance eKYC system reliability. Hybrid approaches combining CNNs, recurrent neural networks (RNNs), and attention mechanisms showed promising results in improving detection rates. Moreover, real-time detection remains a challenge due to computational constraints, emphasizing the importance of optimizing algorithms for efficiency.

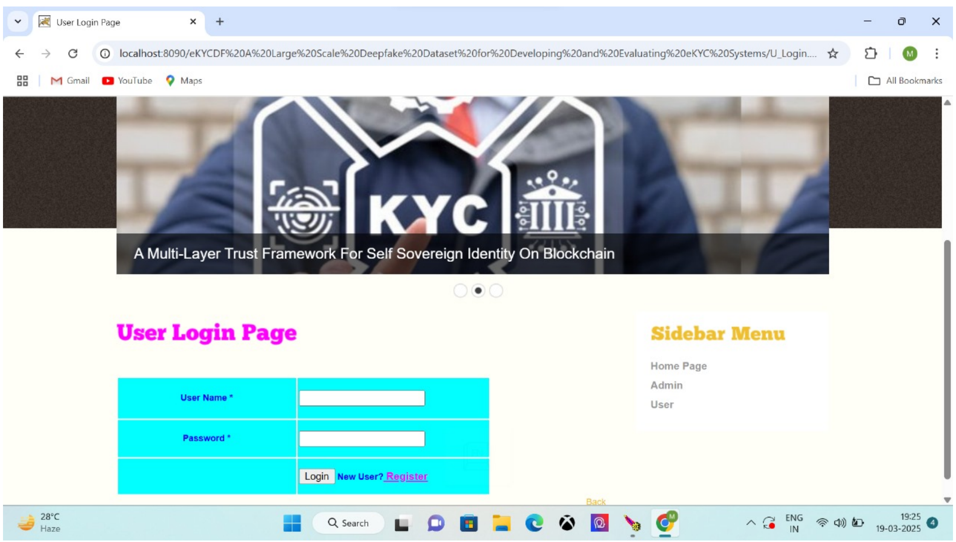

Fig.2. User Login

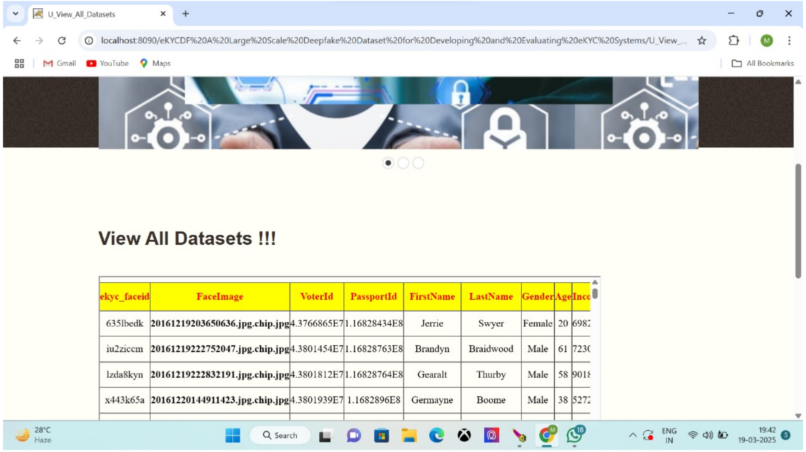

Fig.3. view all data sets

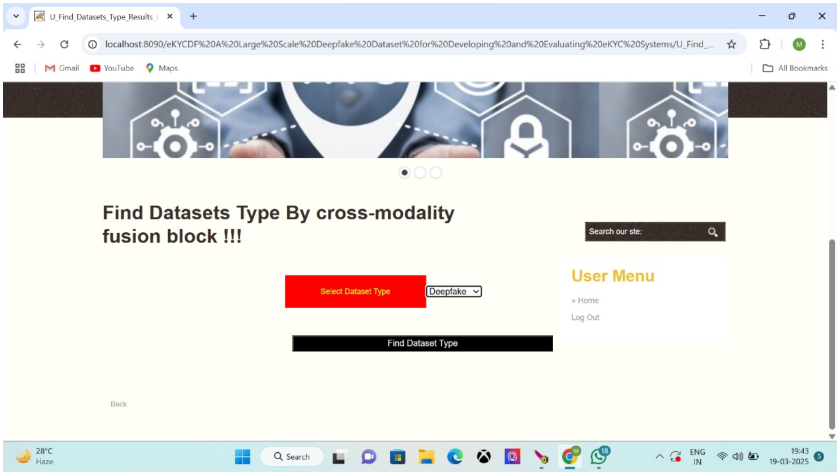

Fig 4. find data sets

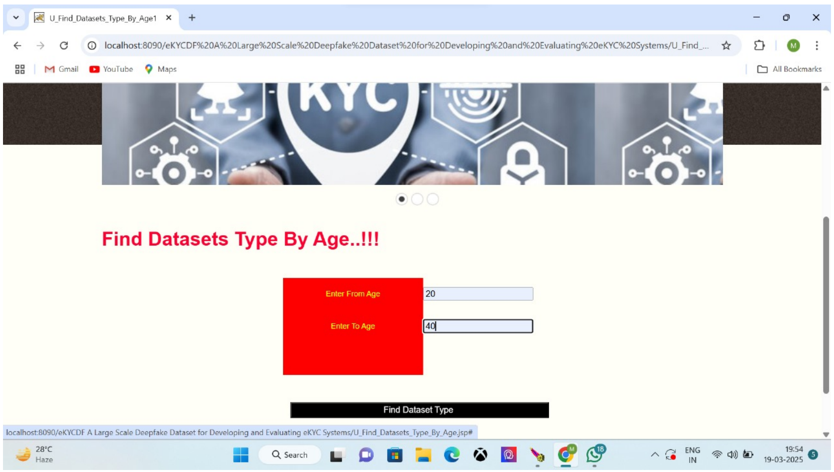

Fig 5. find data sets by age

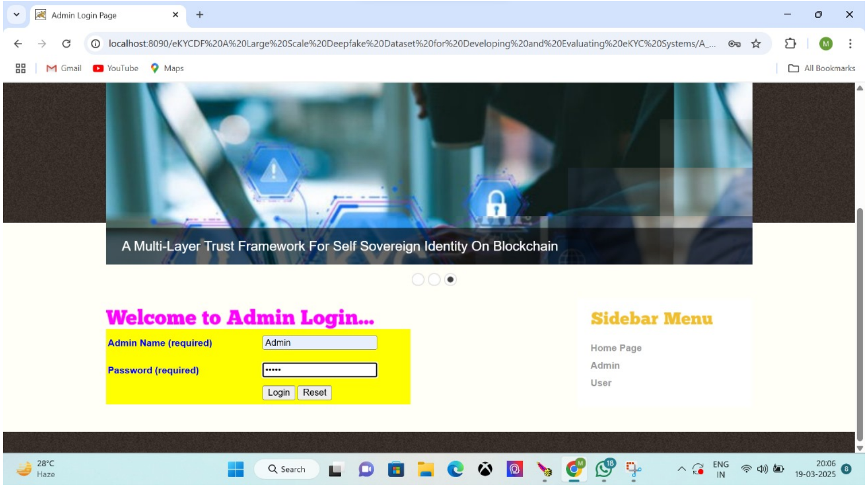

Fig 6. Admin login

Fig 7. All authorized users and not authorize users

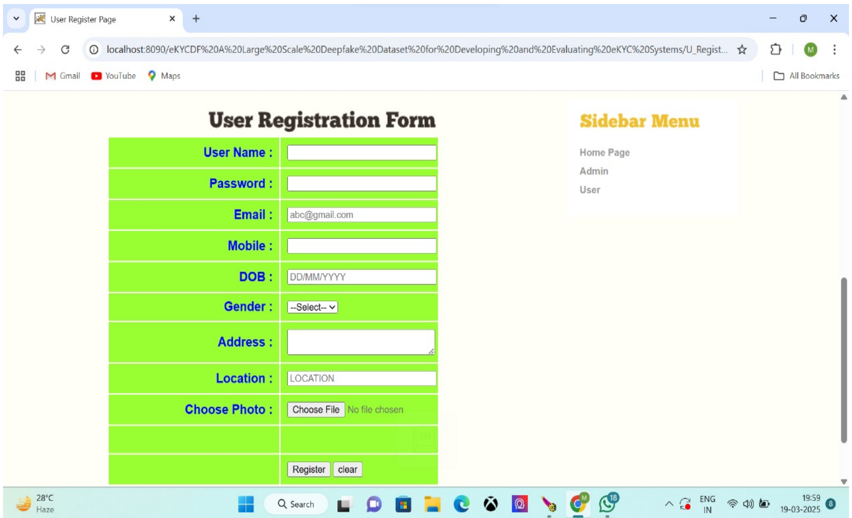

Fig. 8: User registration form

International Journal of
Computational Learning & Intelligence
Vol. 04 | Issue 04 | 2025

VI. CONCLUSION AND FUTURE WORK

The increasing sophistication of deepfake technology poses a significant challenge to electronic Know Your Customer (eKYC) systems, making traditional facial recognition-based identity verification vulnerable to manipulation. This study introduced eKYC-DF, a large-scale dataset designed to enhance the development and evaluation of deepfake detection models specifically for eKYC applications. The dataset incorporates a diverse collection of high-quality real and synthetic videos, including head movements and facial expressions necessary for liveness detection.Through comprehensive evaluations, the study demonstrated the high visual quality of the deepfake videos in eKYC-DF, making them difficult to distinguish from real ones. Experiments also revealed that facial recognition systems struggle to differentiate between genuine and manipulated videos, highlighting a critical security vulnerability in current eKYC implementations. Additionally, deepfake detection models trained on general-purpose datasets failed to generalize well when tested on eKYC-DF, emphasizing the need for specialized datasets tailored for identity verification scenarios. The findings reinforce the necessity of integrating deepfake detection mechanisms into eKYC systems to enhance security. Without these mechanisms, identity fraud using deepfakes remains a significant threat, potentially leading to financial losses, data breaches, and unauthorized access to sensitive services.

REFERENCES

1. Serengil, S. I., & Ozpinar, A. (2021, October). HyperExtended LightFace: A facial attribute analysis framework. In *Proceedings of the International Conference on Engineering and Emerging Technologies (ICEET)* (pp. 1–4).
2. Carta, K., Barral, C., El Mrabet, N., & Mouille, S. (2022, March). Video injection attacks on remote digital identity verification solution using face recognition. In *13th International Multi-Conference on Complexity, Informatics and Cybernetics (IMCIC)* (pp. 92–97).
3. Nanda, A., Shah, S. W. A., Jeong, J. J., Doss, R., & Webb, J. (2023). Towards higher levels of assurance in remote identity proofing. *IEEE Consumer Electronics Magazine, 13*(1), 1–8.
4. Do, T.-L., Tran, M.-K., Nguyen, H. H., & Tran, M.-T. (2022). Potential attacks of DeepFake on eKYC systems and remedy for eKYC with DeepFake detection using two-stream network of facial appearance and motion features. *Social Networking and Computational Science, 3*(6), 1–17.
5. Dagar, D., & Vishwakarma, D. K. (2022). A literature review and perspectives in deepfakes: Generation, detection, and applications. *International Journal of Multimedia Information Retrieval, 11*(3), 219–289.
6. Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Niessner, M. (2019, October). FaceForensics++: Learning to detect manipulated facial images. In *IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 1–11).
7. Akhtar, Z. (2023). Deepfakes generation and detection: A short survey. *Journal of Imaging, 9*(1), 18.
8. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022, June). High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 10674–10685).
9. Ahmed, S. T., Kumar, V. V., Singh, K. K., Singh, A., Muthukumaran, V., & Gupta, D. (2022). 6G enabled federated learning for secure IoMT resource recommendation and propagation analysis. *Computers and Electrical Engineering*, *102*, 108210.
10. Ahmed, S. T., Singh, D. K., Basha, S. M., Abouel Nasr, E., Kamrani, A. K., & Aboudaif, M. K. (2021). Neural network based mental depression identification and sentiments classification technique from speech signals: A COVID-19 Focused Pandemic Study. *Frontiers in public health*, *9*, 781827.
11. Ahmed, S. T., Vinoth Kumar, V., Mahesh, T. R., Narasimha Prasad, L. V., Velmurugan, A. K., Muthukumaran, V., & Niveditha, V. R. (2024). FedOPT: federated learning-based heterogeneous resource recommendation and optimization for edge computing. *Soft Computing*, 1-12.
12. Chen, R., Chen, X., Ni, B., & Ge, Y. (2020, October). SimSwap: An efficient framework for high fidelity face swapping. In *Proceedings of the 28th ACM International Conference on Multimedia* (pp. 2003–2011).

13. Dwaram, J. R., & Madapuri, R. K. (2022). Crop yield forecasting by long short-term memory network with Adam optimizer and Huber loss function in Andhra Pradesh, India. *Concurrency and Computation: Practice and Experience, 34*(27). https://doi.org/10.1002/cpe.7310
14. Fathima, A. S., Basha, S. M., Ahmed, S. T., Khan, S. B., Asiri, F., Basheer, S., & Shukla, M. (2025). Empowering consumer healthcare through sensor-rich devices using federated learning for secure resource recommendation. *IEEE Transactions on Consumer Electronics*
15. Fathima, A. S., Basha, S. M., Ahmed, S. T., Mathivanan, S. K., Rajendran, S., Mallik, S., & Zhao, Z. (2023). Federated learning based futuristic biomedical big-data analysis and standardization. *Plos one*, *18*(10), e0291631.
16. Fathima, A. S., Reema, S., & Ahmed, S. T. (2023, December). ANN based fake profile detection and categorization using premetric paradigms on instagram. In *2023 Innovations in Power and Advanced Computing Technologies (i-PACT)* (pp. 1-6). IEEE.
17. Kumar, A., Satheesha, T. Y., Salvador, B. B. L., Mithileysh, S., & Ahmed, S. T. (2023). Augmented Intelligence enabled Deep Neural Networking (AuDNN) framework for skin cancer classification and prediction using multi-dimensional datasets on industrial IoT standards. *Microprocessors and Microsystems*, *97*, 104755.
18. Madapuri, R. K., & Senthil Mahesh, P. C. (2017). HBS-CRA: Scaling impact of change request towards fault proneness: Defining a heuristic and biases scale (HBS) of change request artifacts (CRA). *Cluster Computing, 22*(S5), 11591–11599. https://doi.org/10.1007/s10586-017-1424-0
19. Reddy, B. S. H. (2025). Deep learning-based detection of hair and scalp diseases using CNN and image processing. *Milestone Transactions on Medical Technometrics, 3*(1), 145–5. https://doi.org/10.5281/zenodo.14965660
20. Reddy, B. S. H., Venkatramana, R., & Jayasree, L. (2025). Enhancing apple fruit quality detection with augmented YOLOv3 deep learning algorithm. *International Journal of Human Computations & Intelligence, 4*(1), 386–396. https://doi.org/10.5281/zenodo.14998944
21. Ren, J., Xu, C., Chen, H., Qin, X., Li, C., & Zhu, L. (2023). Towards flexible, scalable, and adaptive multi-modal conditioned face synthesis. *arXiv preprint arXiv:2312.16274*.