VeriSphere: A Privacy-Preserving Hybrid Framework for Real-Time Aadhaar Document Fraud Detection

Shaurya Dawra[1], Purvi Mohanty[2], Dheeraj Singha[3]

Department of Information Technology, SRM Institute of Science and Technology, Ramapuram, Chennai, India.

E-mail: [2]purvimohanty2503@gmail.com

Abstract

This research work presents a solution to the increasing problem of Aadhaar document forgery, which poses challenges in ensuring secure identification, in the form of VeriSphere, a complete multi-modal solution for real-time fraud detection. This framework includes a wide variety of methods such as object detection, optical character recognition (OCR), cryptography and biometrics into a unified decision-making architecture. Object detection is carried out using YOLOv8, which identifies the relevant fields in the document. EasyOCR, along with multi head attention mechanism, is used for performing text extraction from these fields. The validity of the Aadhaar number is checked using the Verhoeff algorithm. QR Code decoding and SSIM based matching provide cross modality verification. Further, other methods like Error Level Analysis (ELA) and ResNet-18 are utilized for detecting the artifacts of tampering. All the verification modules' outcomes are integrated using weighted fusion of fraud scores to reach an authenticity decision. The experimental results proved that VeriSphere offers a 97.3% accuracy, 96.8% precision, 98.1% recall and 97.4% F1-score.

Keywords: Aadhaar Document Fraud Detection, Multi-Modal Verification, Decision-Level Fusion, Digital Image Forensics, Verhoeff Checksum Validation, Privacy-Preserving Authentication.

DOI: https://doi.org/10.36548/jitdw.2026.3.008
Extended version of article received from **International Conference on Smart Innovations in Engineering & Technology (ICSIET 2026)**
Received: 15.06.2026, received in revised form: 17.07.2026, accepted: 09.08.2026, published: 20.08.2026

1. Introduction

The Aadhaar card has become one of the most popular identity cards issued in India and is mainly used to authenticate users for various services like government schemes, bank operations, healthcare organizations, education, and digital portals. With an increase in the use of Aadhaar verification services, maintaining the validity and integrity of the identity documents has become crucial. However, the developments in imaging, image editing, digital manipulation, and artificial intelligence technology have made it easier to forge the identity documents. Forgery of identity cards through modification of photographs, Aadhaar number changes, text changes, QR code changes, and cloning of identity documents is detrimental to the authentication process and results in fraud and privacy violations. This makes it necessary to develop smart fraud detection mechanisms to detect any type of tampering in the documents while ensuring the user privacy.

Existing methods for Aadhaar verification largely depend on visual analysis, optical character recognition (OCR) text extraction, and individual machine learning model. While such mechanisms can spot some forms of fraudulent behavior, they are unable to handle more complex cases that consist of contradictions between different elements of the document. For instance, a manipulated Aadhaar card can retain its visual form but have identification numbers, wrong QR data, or even a new picture inserted into it. Multiple contradictions like these cannot be spotted through one verification process. Additionally, many existing solutions do not have strong capabilities in terms of transparency and may involve transferring personal information outside the system.

In order to overcome such issues, VeriSphere, a privacy-preserving hybrid framework for real-time Aadhaar document fraud detection, is proposed in this study. The proposed system integrates field localization based on YOLOv8, information extraction based on OCR, Verhoeff checksum validation, QR code decoding and validation, biometric photo matching using Structural Similarity Index Measure (SSIM), and digital forensic analysis using error level analysis and ResNet-18 classification. In addition, a multi-head attention mechanism is adopted in the proposed framework to identify the relationship between the document fields and reinforce consistency validation. All the verification results from the above modules are integrated as a normalized fraud scoring model to provide a transparent fraud verification

result with supporting evidence. This research study intends to design an accurate and privacy-preserving Aadhaar verification system that can detect different types of forgery attempts in real-time.

2. Related Works

Identity document forgery detection has become a major challenge in the context of digital identity systems. Initial research work has been done with respect to image forensics, where copy move forgery detection has been one of the initial techniques used in detecting the duplicate segments of images and has laid the basis of document forgery detection [1][3]. With the advent of deep learning, forgery generation and detection have taken a leap forward. The ability of deep learning systems to generate document forgery has highlighted the challenges of rule-based systems [2].

Object detection technology has improved the process of extracting important document features such as ID numbers, photos, and QR codes. In particular, real-time object detection models, such as YOLO, allow for fast and accurate region detection, which helps to verify the extracted information [6]. Likewise, deep learning models have contributed to advancements in the field of OCR, and sequence-based models, as well as transformer models, help in increasing the efficiency of text recognition from documents by integrating complex layout structures [7][8]. The visual verification process relies on structural similarity measurements that help to identify whether there is any face image manipulation [4]. Additionally, deep convolutional neural networks, especially residual architectures, help to increase the efficiency of the fraud detection model [5].

However, most of the research works are limited to the individual verification methods like image forensics, OCR verification, object detection, or deep learning-based classification. A comprehensive approach that incorporates visual analysis, text extraction, cryptographic authentication, biometrics, forensic verification, and context consistency along with protecting the user privacy has not yet been investigated. The proposed VeriSphere framework attempts to overcome this challenge by developing an architecture that integrates multiple methods for verification and authentication to detect the fraudulent Aadhaar documents.

3. Methodology

The proposed VeriSphere framework perfroms Aadhaar document authentication in real-time by using a combination of visual inspection, text verification, cryptographic validation, biometric matching, and digital forensics. In contrast to traditional verification systems that only makes use of one evidence source, the proposed framework makes use of various evidence sources to increase the fraud detection accuracy while ensuring user privacy. The workflow of the entire process is shown in Figure 1, which demonstrates various stages from document collection to fraud detection. All of these stages provide independent evidence about the authenticity of the document.

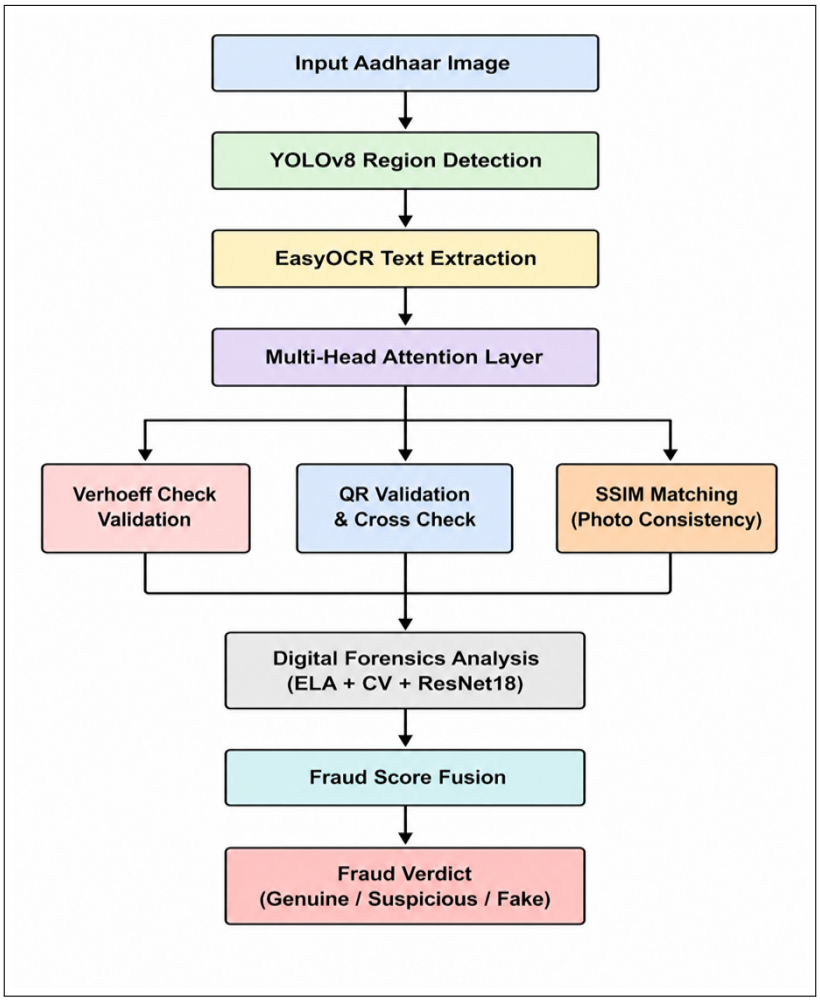

Figure 1. Architecture of the proposed verisphere framework for aadhaar fraud detection

The composition of the dataset is presented in Table 1, and it consists of both genuine and fraudulent Aadhaar card samples that were collected for different verification cases in the real world. The collection contains mainly genuine cards, whereas the rest of the samples include

various types of document manipulations such as photo manipulation, QR code manipulation, text alteration, forgery, and low-quality scans. It is structured to challenge the proposed system against easy and complexfraud cases.

The dataset obtained from the Aadhaar Images Dataset [11] and improvement with controlled forgery creation helps to create an equal distribution of visual, textual, cryptographic, and biometric errors. Stratified split 80/20 was chosen to train and test the data, thereby creating a good distribution of all types of forgeries during the process. In addition, every image in the dataset was labeled with six fields (Annotation Classes).

Table 1. Summary of the aadhaar document fraud detection dataset

Attribute	Value
Source	Aadhaar Images Dataset (Kaggle) [11] + Synthetic Forgeries
Total Images	2,000
Genuine Samples	1,200
Forged Samples	800
Annotation Classes	6 (Aadhaar Number, Name, DOB, Address, Face, QR Code)
Dataset Split	80% Training, 20% Testing
Forgery Categories	Photo Replacement, QR Substitution, Text Manipulation, Composite Forgery, Low-Quality Scans

The first verification stage involves document field localization by using YOLOv8. Precise extraction of each document field is very critical since the verification will depend on access to the data available in the field. YOLOv8 detects each field and creates bounding boxes based on the Complete Intersection over Union (CIoU) loss function.:

$$L_{\text{box}} = 1 - IoU + \frac{\rho^2(b, \hat{b})}{c^2} + av \tag{1}$$

where,

- ρ : denotes the distance between predicted and actual bounding-box centers
- c : represents the diagonal length of the smallest enclosing box

- αv : penalizes aspect-ratio mismatch.

Documents that show extremely low localization confidence are regarded as suspicious since forgeries tend to have structural irregularities that impact the detection process. After field extraction, the textual content is obtained using EasyOCR. The OCR module has a convolutional recurrent architecture which can detect textual patterns from different layouts and qualities of documents. Text recognition is modeled as

$$P\left(y \mid x\right) = \prod_{t=1}^{T} P\left(y_t \mid h_t\right) \tag{2}$$

where,

- h_t : represents the contextual sequence state generated from the extracted visual features.

Additional semantic validation is performed through contextual embedding comparison:

$$S_e(w) = \cos\left(E_f(w), C_e\right) \tag{3}$$

$$S_{combined}(w) = S_e(w) + S_c(w) \tag{4}$$

This stage extracts textual content also evaluates confidence and semantic consistency across critical identity fields. Inconsistencies detected during OCR verification contribute to the overall fraud assessment process.

To further enhance document validation, the relationship between the extracted information is modeled through a multi-head attention network. The attention-based model learns dependencies between the different identity attributes like the name, date of birth, and Aadhaar number. The attention computation is defined as

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5}$$

where Q, K, and V denote query, key, and value representations. The contextual representation facilitates identification of any abnormalities in the relationship of fields which might suggest tampering. Once the contextual encoding takes place, the Verhoeff checksum

algorithm verifies the integrity of the Aadhaar number. Given that Aadhaar numbers have a mathematical checksum format, small changes can be easily caught. The validation process is expressed as

$$c \leftarrow d\left(p\left(i \bmod 8, a_i\right), c\right) \tag{6}$$

where,

- d : denotes the multiplication table
- p : represents the permutation operation.

Document authenticity is confirmed by the process of checksum validation since any manipulation carried out visually will fail the cryptographic validation process. The next step involves the decoding of the localized QR code through a cascading multi-backend decoding process. Comparison is then carried out between the data recovered and the OCR extracted content. QR decoding is performed as

$$\text{Fields } = GZIP^{-1}\left(\text{ hex 2 bytes } \left(BigInt\left(QR_{\text{raw}}\right)\right)\right) \tag{7}$$

while the cross-verification score is calculated as

$$S_{QR} = \frac{1}{N} \sum_{i=1}^{N} 1\{\text{ match } \left(OCR_i, QR_i\right)\} \tag{8}$$

where,

- N : denotes the number of verified fields.

The verification results are summarized in Table 2, the documents have been classified based on agreement rates between the OCR and QR information. Verification process is very efficient in discovering QR substitution attacks and covert identity changes.

Table 2. QR–OCR consistency verification outcomes and fraud interpretation

Consistency Level	Verification Condition	Fraud Assessment
High Consistency	All critical fields match	Genuine
Moderate Consistency	Partial field agreement observed	Suspicious

Low Consistency	Significant mismatch across fields	Fake

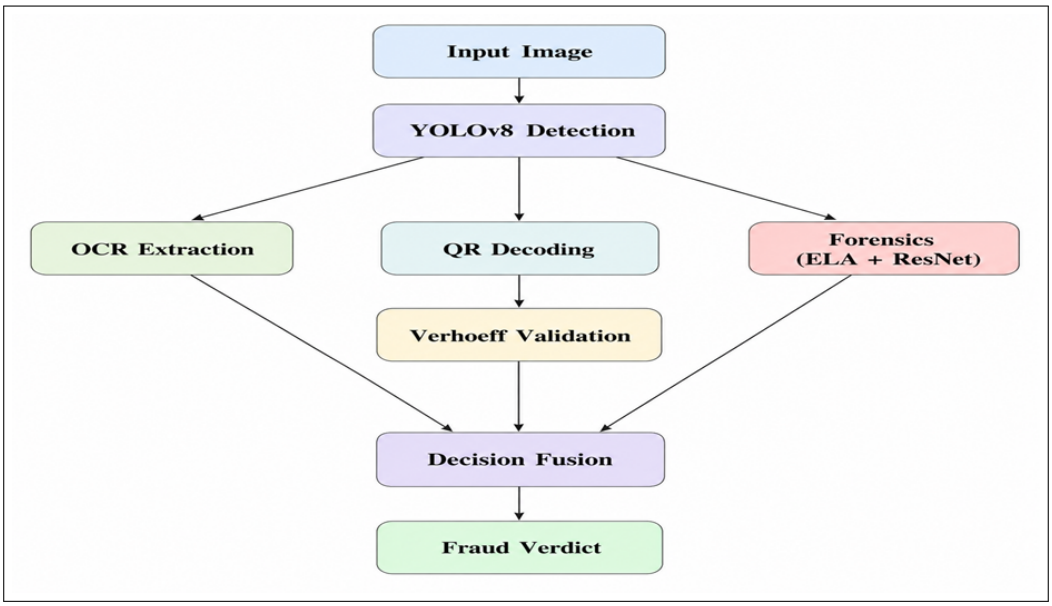

Figure 2. Multi-modal analysis and decision fusion architecture of verisphere

Figure 2 shows the relationship between OCR verification, QR validation, and forensic analysis takes place. The difference from a traditional serial architecture lies in the fact that these verification modules work separately from each other and provide evidence for the final decision-making process. This helps because manipulation that goes unnoticed in one module can still be detected in another one.

Biometric consistency verification is then performed by comparing the face image printed on the Aadhaar card with the photograph embedded within the QR payload. Structural Similarity Index Measure (SSIM) is employed to quantify facial consistency:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{9}$$

where, μ, σ^2, and σ_{xy} denote mean intensity, variance, and covariance, respectively. Significant reductions in similarity scores indicate possible photograph replacement or identity substitution attempts.

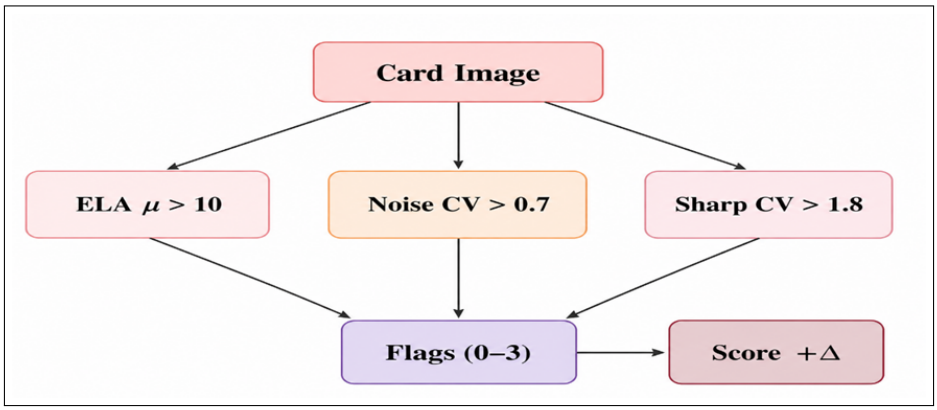

Figure 3. Digital forensics module for tampering evidence extraction

As shown in Figure 3, three complementary forensic techniques are employed.

Error Level Analysis identifies recompression artifacts generated during editing:

$$ELA(I) = \left| I - JPEG_{q=95}(I) \right| \tag{10}$$

Noise consistency is evaluated through the coefficient of variation:

$$CV = \frac{\sigma(I)}{\mu(I)} \tag{11}$$

while deep forensic classification is performed using a fine-tuned ResNet-18 model:

$$P(\text{ fake } \mid I) = \text{ softmax } (\text{ ResNet } 18(I))_{\text{fake}} \tag{12}$$

The results of the detection system are combined to determine the probability of manipulation, thus providing more evidence for fraud identification. Finally, all verification outputs are combined using a normalized fraud score function. The overall fraud score is computed as

$$S_{\text{fraud}} = w_1(1-V) + w_2\left(1-S_{QR}\right) + w_3\left(1-\text{ photo }_{\text{match}}\right) + w_4 P(\text{ fake } \mid I) + w_5 f \tag{13}$$

where the weights satisfy $w_1 + w_2 + w_3 + w_4 + w_5 = 1$, ensuring a normalized decision space. According to the calculated score, the document will be labeled either authentic, suspect, or counterfeit. This multi-modal fusion technique allows the proposed approach to identify

any manipulation of Aadhaar documents while preserving explainability, privacy, and real-time performance.

Algorithm 1: VeriSphere multi-modal aadhaar fraud detection

Input: Aadhaar card image I

Output: Fraud verdict V {Genuine, Suspicious, Fake}

1. Load Aadhaar image I
2. Apply YOLOv8 detector to localize document fields
3. Extract regions: Aadhaar Number, Name, DOB, Address, Face Photograph, QR Code
4. Perform OCR extraction using EasyOCR
5. Compute OCR confidence score S_OCR
6. Encode extracted textual fields using Multi-Head Attention mechanism
7. Validate Aadhaar number using Verhoeff checksum algorithm
8. If checksum validation fails, assign Verhoeff fraud flag
9. Decode QR code using ZXing → Pyzbar → OpenCV decoder chain
10. Compare OCR fields with QR payload
11. Compute QR consistency score S_QR
12. Extract QR embedded photograph
13. Compare QR photograph with card photograph using SSIM
14. Compute biometric matching score S_SSIM
15. Perform Error Level Analysis (ELA)
16. Compute Noise Consistency using Coefficient of Variation (CV)
17. Apply ResNet-18 forensic classifier
18. Generate forensic score S_F
19. Aggregate forensic flags: ELA, Noise CV, Sharpness CV
20. Compute normalized fraud score:

$S_{\text{fraud}} = w_1(1 - V_c) + w_2(1 - S_{\text{QR}}) + w_3(1 - S_{\text{SSIM}}) + w_4 S_F + w_5 f$

21. If S_fraud < 0.15, Verdict ← Genuine
22. Else if 0.15 ≤ S_fraud < 0.40, Verdict ← Suspicious
23. Else, Verdict ← Fake
24. Return Verdict and fraud explanations

Algorithm 1 outlines the overall workflow of the proposed VeriSphere framework for Aadhaar fraud detection. The process begins with YOLOv8-based localization of key

document fields, followed by OCR extraction of textual information. All these fields are subjected to the multi-head attention model to identify contextual relationship. The veracity of the Aadhaar number is authenticated using the Verhoeff checksum method, while decoding of the QR code helps in matching machine-readable data with the extracted one. Further, biometric authentication involves comparing the card photograph with the image present in the QR code via SSIM. On the other hand, digital forensics involve the use of Error Level Analysis (ELA) technique, noise consistency test, and ResNet-18-based classification of images to detect any fraudulent activity. Finally, all these module outputs are fused together through a fraud score normalization technique.

4. Results and Discussions

4.1 Experimental Setup

This VeriSphere model was implemented through the combination of object detection, OCR, cryptography validation, biometric verification, and forensic investigation into one document authentication workflow. The implementation was optimized for real-time processing but still maintains strong detection capabilities regardless of the variety of document manipulations performed. The implementation and training setup used for this proposed framework is presented in Table 3.

Table 3. Experimental configuration of the proposed verisphere framework

Parameter	Configuration
Programming Language	Python 3.11
Deep Learning Framework	PyTorch 2.2
Object Detection Model	YOLOv8
OCR Engine	EasyOCR
Forensic Classification Model	ResNet-18
Training–Testing Split	80%: 20%
Data Augmentation	Rotation, Scaling, Brightness Adjustment, Perspective Transformation

Hardware	Intel Core i9 Processor, 32 GB RAM, NVIDIA RTX 3090 GPU

4.2 Performance Evaluation

The detection accuracy of the proposed VeriSphere framework was extensively analyzed using classification measures in order to distinguish genuine from forged Aadhaar documents. As indicated by the results presented in Table 4, the proposed framework demonstrated remarkable classification accuracy in incorporating several verification methods into one comprehensive solution: visual, textual, cryptographic, biometric, and forensic. The classification accuracy of the proposed framework demonstrated no preference to any of the two categories and showed remarkable performance in both aspects, which is confirmed by the high recall rate demonstrating the ability of the proposed framework to detect manipulated documents with minimum number of false negatives, and precision preventing false fraud alarms. Moreover, the F1-score reflects the dual potential of the proposed framework for detection sensitivity and predictive power. Additionally, the inference latency is low enough to make the proposed framework applicable for real-time Aadhaar document verification.

Table 4. Performance evaluation of the proposed verisphere framework

Metric	Value
Accuracy (%)	97.3
Precision (%)	96.8
Recall (%)	98.1
F1-Score (%)	97.4
Average Inference Latency (s)	1.2

4.3 Web Verification Dashboard

A Web-based verification dashboard has been designed to serve as an interactive interface to display the outputs generated by the proposed VeriSphere framework in a combined manner. The dashboard involves both document-level fraud score and module-wise verification proof such that the verification process can be understood without checking each model output

individually.

As shown in Figure 4, the interface displays the fraud score, the OCR confidence, the SSIM similarity score for biometrics, and the status of the Verhoeff checksum verification as key indicators for verification purposes. In this particular test case, the fraud score is above the decision threshold set, whereas the verification using the checksum and biometrics reveals inconsistencies. The dashboard shows the number of extracted fields and the QR-OCR consistency score as well, making it possible to detect discrepancies on the field-level basis. The forensic analysis module shows the tampering artifacts found by the combination of ELA, noise-consistency, and ResNet-18 analyses. An example comparison panel is also displayed on the interface to provide the user with a visual representation of the actual and fake fields features.

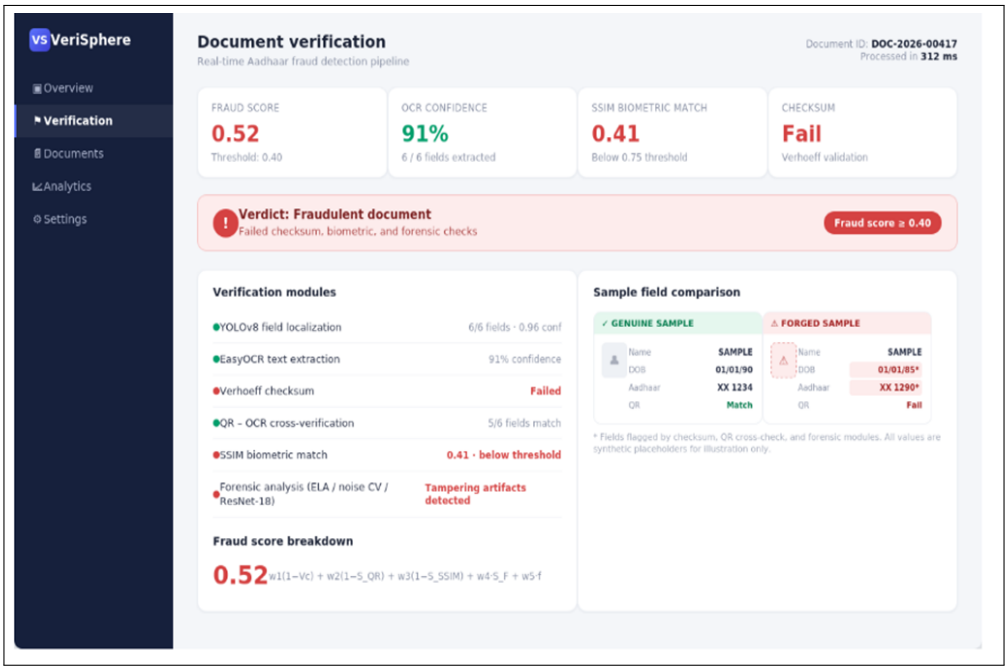

Figure 4. Web-based verification dashboard of the VeriSphere framework

4.4 Ablation Study

An ablation analysis was conducted to understand the role played by each individual verification module through the gradual elimination of vital parts from the suggested VeriSphere architecture without changing anything else about the architecture. The performance differences are shown in Figure 5. The complete framework had the best detection performance, which proves the effectiveness of integrating multiple verifications into a single decision mechanism.

Elimination of the Verhoeff checksum verification system led to decreased performance in detection of tampered samples due to its importance in ensuring the validity of the Aadhaar number. Likewise, removal of the QR code verification system hampered the performance of detecting the tampered information especially where there was any change in the text. Biometric verification elimination led to decreased capability of identifying photograph manipulation tampering attacks while removal of the digital forensics component lowered the sensitivity towards image tampering and compression.

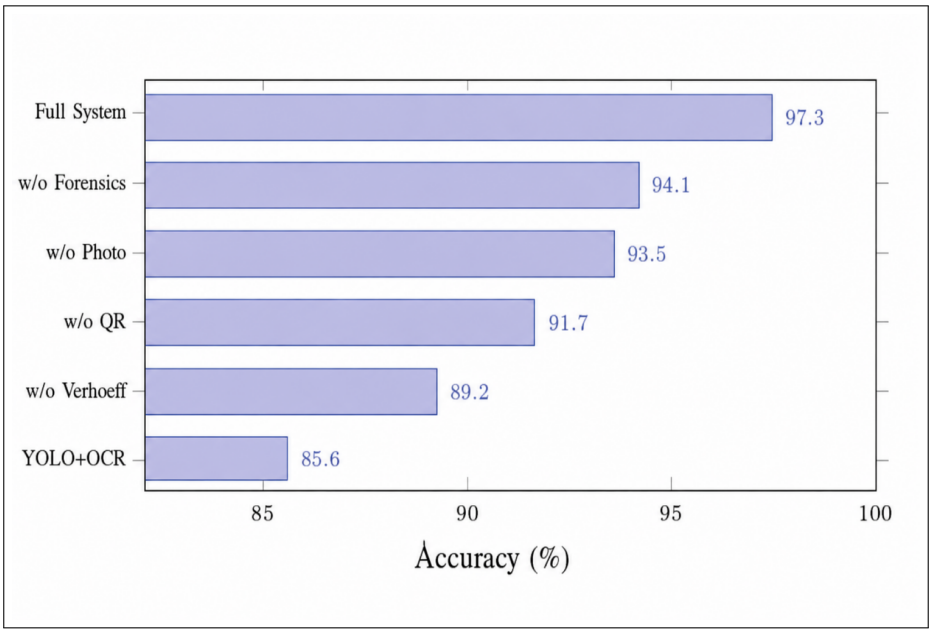

Figure 5. Ablation study of the VeriSphere framework

Overall, the performance reductions identified across the ablation configurations confirm that each module provides complementary evidence for the final fraud decision. The results demonstrate that integrating visual, textual, cryptographic, biometric, and forensic information through decision-level fusion improves the robustness and reliability of VeriSphere against diverse Aadhaar document forgery scenarios while supporting privacy-preserving authentication.

5. Conclusion and Future Work

In this research, VeriSphere was introduced as a privacy-preserving hybrid approach for real-time detection of fraudulent Aadhaar documents via object detection, optical character recognition, cryptographic verification, biometric consistency checks, and digital image forensics. The proposed system successfully detects various manipulation attacks

on documents while maintaining verification efficiency. Through experimental evaluations, the multi-modal verification approach was found to have 97.3% accuracy, 96.8% precision, 98.1% recall, and a 97.4% F1-score, which confirms the ability of the approach to provide efficient fraud detection with minimal inference latency. An ablation study further validates the contribution of each verification module in adding complementary information for increasing robustness of the system. Future research will be directed towards expanding the scope of the proposed solution to support more government-issued ID cards, using transformer-based vision-and-language models for better cross-modal reasoning, improving resistance to future AI-generated ID card forgery attacks, and implementing privacy-preserving federated learning for edge model deployment.

References

[1] Mahfoudi, Gaël, Frédéric Morain-Nicolier, Florent Retraint, and Marc Pic. ”CMID: A New Dataset for Copy-Move Forgeries on ID Documents.” IEEE International Conference on Image Processing (ICIP) 2021: 3028-3032.

[2] Zhao, Lin, Changsheng Chen, and Jiwu Huang. ”Deep Learning-Based Forgery Attack on Document Images.” IEEE Transactions on Image Processing 2021, vol 30: 7964-7979.

[3] Fridrich, Jessica, David Soukal, and Jan Lukas. ”Detection of Copy-Move Forgery in Digital Images.” In Proceedings of Digital Forensic Research Workshop 2003, vol. 3, no. 2: 652-63.

[4] Wang, Zhou, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. ”Image Quality Assessment: From Error Visibility to Structural Similarity.” IEEE Transactions on Image Processing 2004, vol. 13, no. 4: 600-612.

[5] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. ”Deep Residual Learning for Image Recognition.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016: 770-778.

[6] Redmon, Joseph, and Ali Farhadi. ”Yolov3: An Incremental Improvement.” arXiv preprint 2018, arXiv:1804.02767.

[7] Shi, Baoguang, Xiang Bai, and Cong Yao. ”An End-To-End Trainable Neural Network

for Image-Based Sequence Recognition and its Application to Scene Text Recognition." IEEE Transactions on Pattern Analysis and Machine Intelligence 2016, vol. 39, no. 11: 2298-2304.

[8] Xu, Jianjun, Yuxin Wang, Hongtao Xie, and Yongdong Zhang. "OTE: Exploring Accurate Scene Text Recognition Using One Token." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024: 28327-28336.

[9] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is All You Need." Advances in Neural Information Processing Systems 2017, vol. 30: 1-11.

[10] Raganato, Alessandro, and Jörg Tiedemann. "An Analysis of Encoder Representations in Transformer-Based Machine Translation." In Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP: 287-297.

[11] Varun Kumar Gera, Aadhar Images Dataset, Kaggle, 2023. Available: https://www.kaggle.com/datasets/varunkumargera/aadhar-images