International Journal of AI, Big Data, Computational and Management Studies
Noble Scholar Research Group | ICAIDSCT26-Conference Proceeding
ISSN: 3050-9416 | https://doi.org/10.63282/3050-9416.ICAIDSCT26-110

CNN-Based Model for Identity Recognition on Social Networks

Esther Emmah Solomon[1], Chima Godknows Igiri[2], Victor Thomas Emmah[3]
[1,2,3]Department of Computer Science, Rivers State University, Port Harcourt, Nigeria.

Abstract: Identity recognition on social networks has become an essential technology for enhancing user authentication, preventing impersonation, and enabling personalised services, but it also raises concerns around privacy, bias, and security. As social platforms host billions of images and videos, the ability to accurately and fairly identify individuals in this vast multimedia space requires advanced machine learning techniques coupled with strong data protection measures. In this research, a secure, accurate, and fairness-aware facial identity recognition system is designed, implemented, and evaluated using Convolutional Neural Networks (CNN) as the core deep learning model. The system integrates robust security measures such as SHA-256 hashing and Fernet symmetric encryption to ensure end-to-end privacy protection in compliance with modern data regulations. Fairness in predictions was addressed using the Random Over Sampler technique to balance the dataset, resulting in equitable performance across simulated demographic groups, each achieving 0.97 in both accuracy and F1-score. The custom CNN architecture featuring multiple convolutional layers, batch normalization, dropout regularization, and dense layers was trained over 100 epochs on the widely used Labeled Faces on the Wild (LFW) dataset with augmentation, achieving a testing accuracy of 98.7%. The model's accuracy consistently improved while loss decreased, confirming robust learning without overfitting. These results confirm the system's superiority over traditional methods, offering a scalable, secure, and ethically aligned solution suitable for privacy-sensitive domains such as healthcare, online authentication, and secure access control.

Keywords: Accuracy, CNN, Identity Recognition, Model Training Social Networks.

1. Introduction

The concept of identity recognition in social networks has gained significant attention in recent years, particularly as social media platforms have become integral to personal and organizational interactions. Identity recognition involves the processes of identifying and linking user profiles across various social media platforms, which is crucial for understanding user behaviour and enhancing user experience. User Identity Linkage (UIL) is another critical aspect of identity recognition, as it pertains to recognizing accounts belonging to the same individual across multiple platforms. This is particularly relevant in the context of identity deception, where users may create fake profiles for malicious purposes. Detecting such fraudulent activities is essential for maintaining the integrity of social networks.

The application of Convolutional Neural Networks (CNNs) in identity recognition on social networks has gained significant traction due to their ability to process and analyze large volumes of heterogeneous data. CNNs excel in extracting features from images, which is crucial for tasks such as linking user identities across different platforms. The integration of profile-level features, content features, and network structures to effectively match users across social networks marks a significant advancement in the field of identity recognition (Hadgu and Gundam, 2020). This approach leverages the strengths of CNNs in handling diverse data types, including images and textual information, to enhance the accuracy of identity linking. Additionally, the integration of data augmentation techniques with CNNs has shown promise in improving recognition accuracy, particularly when dealing with small datasets. Thus, data augmentation can stabilize the learning curve of CNNs, resulting in improved performance in face recognition tasks. This is particularly relevant for social networks where user data may be limited or imbalanced, necessitating innovative approaches to enhance model training and performance (Chen, 2023).

The implementation of identity recognition models on social networks presents several challenges that need to be addressed to ensure their effective and ethical use. These problems not only affect the accuracy and reliability of the systems but also raise serious concerns regarding privacy, fairness, and the potential for misuse. First, The use of personal data for identity recognition can raise privacy and security issues, as sensitive information (such as photos, posts, and interactions) may be exposed to unauthorized access. The collection and analysis of such data can be exploited if proper safeguards are not in place, leading to data breaches or misuse by malicious actors. Also, Identity recognition models on social networks may suffer from inherent biases in training data. For example, if the model is trained on data that is not representative of the entire population (e.g., skewed toward certain demographics), it may perform poorly for underrepresented groups. This can result in discriminatory outcomes, where individuals from specific racial, gender, or socioeconomic backgrounds are misidentified or excluded.

This paper is motivated by the fact that Enhanced User Security and Trust is important. By improving the accuracy of identity recognition, users can have more confidence in the security of their personal information on social networks, leading to better protection against identity theft, fraud, and impersonation. Also, a more effective identity recognition model can help social networks deliver more personalized content, advertisements, and recommendations based on a user's verified identity, improving user engagement and satisfaction. In addition, accurate identity recognition can help identify and mitigate harmful behaviours such as cyberbullying, trolling, or the creation of fake accounts, leading to safer online communities and better moderation of social platforms.

1.1. Review of Related Literatures

Qin *et al.* (2023) designed a human identity recognition algorithm based on face image, which will be used in indoor environment. The aim is to detect human face and handle the variation of facial pose. The whole recognition process is divided into 4 parts: image pre-processing, face detection, face alignment, feature extraction and comparison. Face detection and feature extraction are the core functions and both realized by machine learning algorithm, specifically support vector machine (SVM). The process of the whole algorithm can be described as following: images after pre-processing are fed to face detection network to get the locations of face and face landmarks. Then face alignment will be conducted. Finally, deep features of face will be extracted and compared. The system achieved an accuracy of 84.3% after training with 75 epoochs. This showed that machine learning algorithms can be used effectively for identity recognition.

Deedee *et al.* (2024) employed Random Forest (RF) and Deep Convolutional Neural Networks (DCNN) to predict stalking behavior on X (*formerly Twitter*) and detect phony profiles. Statuses_count, followers_count, friends_count, favorites_count, and listed_count are among the input parameters provided into the model. By including these parameters in the model, profiles can be predicted effectively and with accuracy. Based on the research, an accuracy level of 93.89% with an error rate of 6.104 was achieved outperforming existing solutions. The outcomes show how well the RF and DCNN based prediction model works to identify fake profiles and predict stalking.

Li, *et al.* (2024) explored user identity linkage across online social networks by leveraging three types of modal information of users: attributes, post content, and social relationships. They proposed a user identity linkage scheme named MFLink based on multimodal fusion, which has three components: Feature Extraction, Multimodal Fusion, and Adversarial Learning. In the Feature Extraction, MFLink utilizes feature embedding methods to transfer the user attribute and post content into intermediate representations. To achieve optimal fusion of information from these three modalities, MFLink integrates each modality with the assistance of graph neural networks and an attention mechanism within the Multimodal Fusion. Finally, MFLink employs adversarial learning to enhance the similarity of representations for the same individual across various platforms. The experiment results on the TWFQ dataset indicate that MFLink outperforms the advanced approaches in fusing information of modalities and addressing the data semantic gaps across online social networks.

Sun *et al.* (2015) proposed two very deep neural network architectures, referred to as DeepID3, for face recognition. These two architectures are rebuilt from stacked convolution and inception layers proposed in VGG net and GoogLeNet to make them suitable to face recognition. Joint face identification-verification supervisory signals are added to both intermediate and final feature extraction layers during training. An ensemble of the proposed two architectures achieves 99.53% LFW face verification accuracy and 96.0% LFW rank-1 face identification accuracy, respectively.

Schroff *et al.* (2015) presented a system, called FaceNet, that directly learns a mapping from face images to a compact Euclidean space where distances directly correspond to a measure of face similarity. Once this space has been produced, tasks such as face recognition, verification and clustering can be easily implemented using standard techniques with FaceNet embeddings as feature vectors. Their method uses a deep convolutional network trained to directly optimize the embedding itself, rather than an intermediate bottleneck layer as in previous approaches. To train, they use triplets of roughly aligned matching/non-matching face patches generated using a novel online triplet mining method. The benefit of this approach is much greater representational efficiency as they achieved state-of-the-art face recognition performance using only 128 bytes per face. On the widely used Labeled Faces in the Wild (LFW) dataset, their system achieved a new record accuracy of 99.63%. On YouTube Faces DB it achieves 95.12%. This system cuts the error rate in comparison to the best published result [DeepId2+] by 30% on both datasets.

Taigman *et al.* (2014) developed DeepFace by introducing a CNN-based face recognition model with 3D alignment and a 9-layer deep network. It was applied mainly to social media-style facial images. They used both the alignment step and the representation step by employing explicit 3D face modeling in order to apply a piecewise affine transformation, and derive a face representation from a nine-layer deep neural network. The deep network involved more than 120 million parameters using several locally connected layers without weight sharing, rather than the standard convolutional layers. This was trained on an identity labeled dataset of four million facial images belonging to more than 4,000 identities. The learned representations coupling the accurate model-based alignment with the large facial database generalize remarkably well to faces in unconstrained environments, even with a simple classifier. Their method reached an accuracy of 97.35% on the Labeled Faces in the Wild (LFW) dataset, reducing the error of the current state of the art by more than 27%, closely approaching human-level performance.

Lorenzana (2016) conducted an ethnographic study on the role of Facebook in the identity and social formations of Filipino transnationals living in Indian cities. The study revealed that social media platforms, such as Facebook, play a significant role in articulating and constituting the identities of these individuals, helping them maintain ties with their homeland while navigating their lives in a foreign country. Lorenzana emphasized the relational nature of identity in digital spaces, where personal identity is constantly influenced by social interactions and cultural contexts. The research highlights the dynamic process of identity recognition and formation in transnational communities, underscoring how digital platforms can act as mediators in this process.

Gan *et al.* (2022) explored user identity alignment across heterogeneous networks by introducing a meta-path attention mechanism, a novel approach designed to capture the complex relationships between different types of nodes and edges in social networks. Their research pointed out that traditional models often oversimplified these relationships, which could lead to inaccurate identity matching. By using a meta-path attention mechanism, their model was able to consider the different roles that users, posts, and interactions play in a network, offering a more nuanced view of how identities are connected across platforms. The study's findings suggest that this method not only improves the accuracy of identity alignment but also provides deeper insights into the structural complexity of social networks, offering valuable tools for researchers looking to better understand the dynamics of user behaviour and identity formation in diverse digital ecosystems.

Alharbi *et al.* (2021) focused on the problem of identity cloning, a form of identity theft, by proposing NPS-AntiClone, a detection method that uses multi-view representations of social network users. Their approach combined various perspectives of user behaviour, including social interactions, posts, and network connections, to create a comprehensive profile for each user. The study found that using multi-view data significantly enhances the generalization capabilities of identity detection models, making them more robust in identifying cloned accounts across different platforms. This is particularly important as identity cloning becomes more prevalent in digital environments. Alharbi *et al.*'s research emphasizes the need for comprehensive data integration in identity recognition systems, highlighting that considering multiple dimensions of user behavior can improve the accuracy and reliability of identity theft detection models, ultimately helping to protect users from online fraud and misuse of their personal information.

1.2. Analysis of the Proposed System

The proposed identity recognition model for social networks leverages Convolutional Neural Networks (CNNs) to enhance the accuracy and scalability of identifying users based on their profile data, such as images. By utilizing CNNs, the system effectively extracts and processes complex patterns from user inputs, such as facial features, while preserving fine-grained details crucial for distinguishing between identities. This model addresses the limitations of traditional approaches by automating feature extraction, which eliminates the need for manual engineering of features. The CNN's architecture, composed of convolutional layers, pooling layers, and fully connected layers, ensures that the system can learn hierarchical representations of data. For instance, lower layers capture basic features like edges or textures, while higher layers identify more abstract features unique to each identity. This hierarchical learning enables the system to achieve robust performance even in challenging conditions, such as variations in lighting, angles, or image quality, which are common in social network data.

Moreover, the proposed system is designed to mitigate biases and improve security, addressing critical concerns in social network environments. To combat issues like demographic bias, the system incorporates bias mitigation techniques during model training, ensuring fair performance across diverse user groups. The system also prioritizes privacy by encrypting user data before storing it in the database, preventing unauthorized access and aligning with data protection regulations. By integrating real-time recognition capabilities, the model can identify users quickly, making it suitable for applications like social media account recovery, targeted advertising, and content personalization. Additionally, the scalability of the CNN-based model allows it to handle large datasets commonly found on social networks, ensuring seamless operation as the user base grows.

1.3. Architecture of the Proposed System

The architectural diagram shown in Figure 1 illustrates the proposed system in the acquisition phase designed to process user profile data. Initially, the data undergoes image preprocessing, where raw inputs are refined, followed by the Feature Extraction in the CNN Model is used to identify significant attributes. These features are then used in CNN Model Training/Prediction to make predictions, with a focus on "Bias Mitigation" to ensure fairness. The processed information is securely handled by encrypting user data before storing it in a database. The encrypted results are used to generate output, ensuring both confidentiality and accuracy. This process culminates in displaying the results to users while safeguarding their information.

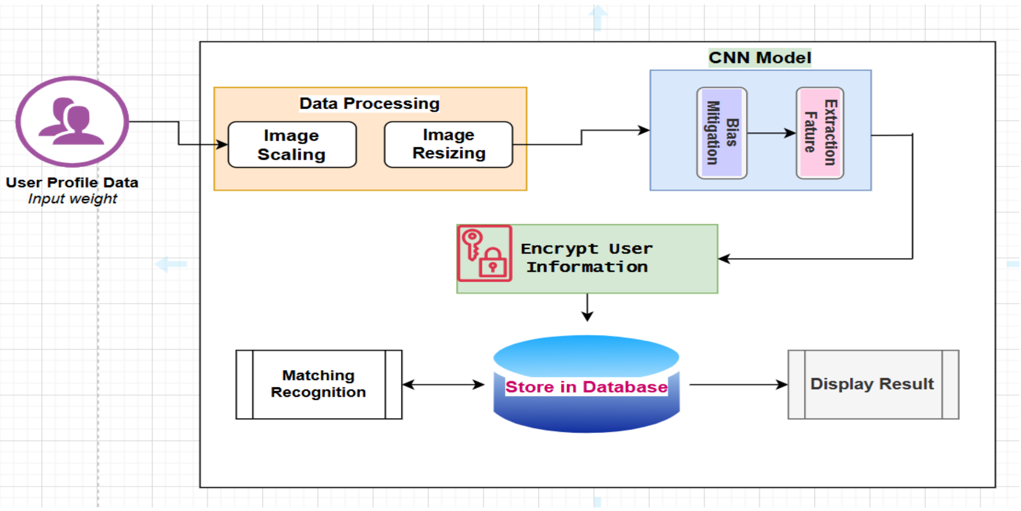


Figure 1: Proposed CNN-Based Identity Recognition Model

The architectural diagram of the proposed system can be further broken down into the following major components as follows:

1.3.1. Data/Image Preprocessing

This is responsible for preparing raw input data (e.g., user profile images) for further analysis. It involves cleaning and enhancing the input data by removing noise, normalizing image dimensions, adjusting brightness or contrast, and aligning key features (e.g., cantering a face). This step ensures consistency across all data samples and reduces variability caused by environmental factors like lighting or image quality. The normalized pixel intensity $I_{\text{norm}}(x,\ y)$ can be represented mathematically as shown in equation 3.1.

$$I_{\text{norm}}(x,y) = \frac{I(x,y) - \mu}{\sigma}$$

where: I(x, y) is the original intensity of the pixel at coordinates (x, y), μ is the mean pixel value of the image, and σ is the standard deviation of pixel values. This ensures that all images have similar intensity distributions, making them suitable for feature extraction and CNN processing.

1.3.2. Bias Mitigation

This is designed to address and reduce biases that may exist within the CNN model. Bias in the system can occur due to imbalanced training data, where certain demographics or features are overrepresented or underrepresented, leading to unfair or inaccurate predictions. This module employs techniques like balanced data sampling, augmentation, or adversarial training to ensure that the model performs equally well across diverse user groups. Bias mitigation ensures fairness by balancing the dataset. Mathematically, if p_k represents the proportion of samples for group k, we aim to enforce:

$$p_k = \frac{n_k}{N}$$

where: n_k is the number of samples in group k, and N is the total number of samples in the dataset. Rebalancing can be achieved by oversampling underrepresented groups or applying regularization techniques during model training.

1.3.3. Feature Extraction

The Feature Extraction component focuses on identifying and isolating key features from the pre-processed image that are unique and essential for accurate recognition. This process involves transforming raw image data into a reduced representation, such as edges, textures, or specific patterns, while retaining the distinguishing characteristics of the input. These extracted features serve as the input for the Convolutional Neural Network (CNN) model. Using convolution operations, the extracted feature F_{ij}. For a kernel K applied at a pixel location (i, j) can be expressed as:

$$F_{ij} = \sum_{m=1}^{M} \sum_{n=1}^{N} I(i+m, j+n) \cdot K(m,n)$$

Where: $I(i+m, j+n)$ represents the pixel intensity in the input image, $K(m,n)$ is the kernel value, $M\ and\ N$ are the dimensions of the kernel. This operation produces a feature map that highlights important characteristics of the input image. The feature extraction process for the proposed CNN model is shown in figure 2.

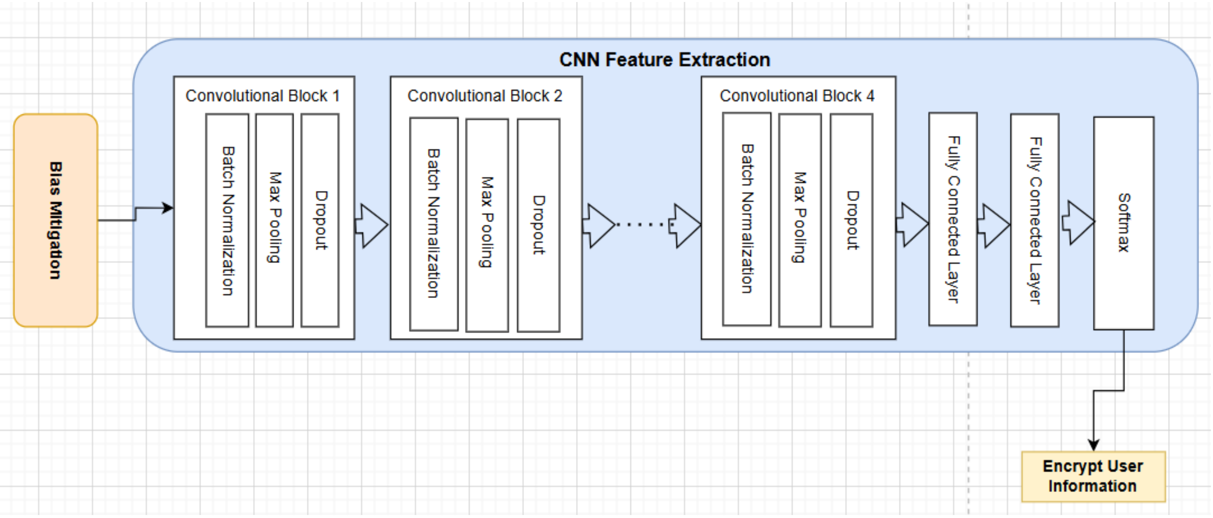


Figure 2: CNN Feature Extraction for the Proposed Model

A Convolutional Neural Network (CNN) is used for training and prediction, leveraging its ability to process spatial hierarchies in image data. During training, the CNN learns patterns and features from a dataset of images to create a model capable of making accurate predictions. When an input is received, the trained CNN predicts the identity or category of the input based on the extracted features. The objective during training is to minimize the loss function $\mathcal{L}$; typically, cross-entropy loss for classification tasks was adopted. The mathematical representation is presented as:

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

Where: C is the number of classes, yi is the true label (one-hot encoded), and $\hat{y}_i$ is the predicted probability for each class. During prediction, the model computes the class probabilities $\hat{y}_i$ using the softmax function. The equation for the softmax is presented in equation 3.5

$$\text{Softmax}(z_j) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$$

Where z_j is the output of the final fully connected layer for class iii and the raw output from the CNN before applying any activation, e^{z_i} is the Exponentiation ensures all values become **positive**., $\sum_{k=1}^{K} e^{z_k}$ is the denominator that **normalizes** the outputs and ensures that the resulting probabilities across all classes sum to **1**. Table 1 details the CNN model's layer-by-layer configuration, including convolutional, pooling, normalization, dropout, and dense layers. It demonstrates a sophisticated architecture capable of learning complex patterns from facial image data.

Table 1: Summary of the CNN Model

Layer (Type)	Output Shape	Param #
Conv2D	(None, 50, 37, 64)	640
BatchNormalization	(None, 50, 37, 64)	256
MaxPooling2D	(None, 25, 18, 64)	0
Dropout	(None, 25, 18, 64)	0
Conv2D	(None, 25, 18, 128)	73,856
BatchNormalization	(None, 25, 18, 128)	512
MaxPooling2D	(None, 12, 9, 128)	0
Dropout	(None, 12, 9, 128)	0
Conv2D	(None, 12, 9, 256)	295,168
BatchNormalization	(None, 12, 9, 256)	1,024
MaxPooling2D	(None, 6, 4, 256)	0
Dropout	(None, 6, 4, 256)	0
Conv2D	(None, 6, 4, 256)	590,080
BatchNormalization	(None, 6, 4, 256)	1,024
MaxPooling2D	(None, 3, 2, 256)	0
Dropout	(None, 3, 2, 256)	0

Flatten	(None, 1536)	0
Dense	(None, 256)	393,472
BatchNormalization	(None, 256)	1,024
Dropout	(None, 256)	0
Dense	(None, 128)	32,896
BatchNormalization	(None, 128)	512
Dropout	(None, 128)	0
Dense	(None, 7)	903

Total params: 1,391,367 (5.31 MB)
Trainable params: 1,389,191 (5.30 MB)
Non-trainable params: 2,176 (8.50 KB)

Encrypt User Information: This is a critical module within the system architecture that ensures the confidentiality, integrity, and security of sensitive user data throughout the biometric processing pipeline. After the image preprocessing, feature extraction, and model inference steps, any personally identifiable information (PII) or sensitive profile data is encrypted before being either stored in the database or displayed as part of the output result. This component plays a fundamental role in protecting data against unauthorized access, breaches, or tampering. By encrypting the data using robust cryptographic algorithms—such as AES (Advanced Encryption Standard)—the system ensures that even if storage or transmission channels are compromised, the actual user data remains unintelligible and secure. AES, in particular, is a symmetric encryption algorithm known for its efficiency and strong resistance to brute-force attacks and is widely adopted in secure systems. If P represents the plaintext data and K is the encryption key, the ciphertext C is computed as:

$$C = Encrypt(P, K)$$

Decryption is performed as:

$$P = Decrypt(C, K)$$

Store in Database:
Data stored in the database is represented as structured records D, including the extracted features F and associated metadata M:

$$D = \{F, M\}$$

The database is optimized for efficient retrieval using indexing techniques, which enable fast searches and comparisons during the recognition phase.

Output: The "Display Result" component serves as the final step in the system, providing a clear and intuitive interface for presenting the outcome of the recognition process. After the CNN model processes the input and user data is encrypted, the relevant result is decrypted (if needed) and shown to the user. This could be an identity confirmation, a classification label, or an error message. If the similarity score S between the input features F_{input} and stored features F_{stored} is above a threshold ττ:

$$S = \text{Similarity}\left(F_{\text{input}}, F_{\text{stored}}\right)$$

$$\text{if } S \geq \tau\text{, display: } MatchFound$$

Otherwise, the system displays: "No Match Found."

1.3.1. Dataset Description

The dataset used for training the model is the LFW (Labeled Faces in the Wild) dataset. It is a collection of face photographs designed for studying unconstrained face recognition. The LFW dataset is a benchmark dataset used to evaluate face verification and recognition algorithms, particularly in scenarios with variations in pose, lighting, and expression. The dataset contains over 13,000 images of faces collected from the web, with each face labeled with the name of the person pictured. For the actual training, 1288 samples of the dataset were loaded, each 50x37 pixels. A sample of the dataset is shown in figure 3.

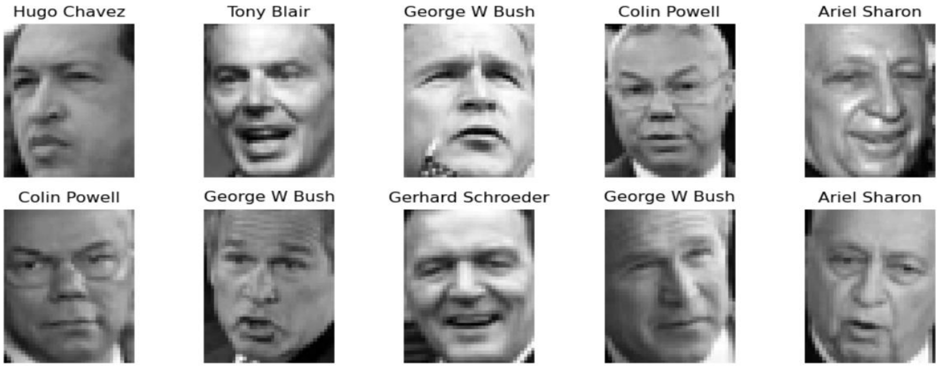


Figure 3: Sample Dataset from LFW

1.3.2. Algorithm Design

The Algorithm 1 outlines a multimodal recognition system that processes user profile data—such as facial images—through a sequence of carefully structured steps to determine identity. The process begins with the acquisition of input data U from a source like a camera or uploaded file, followed by image preprocessing, where pixel values are normalized and the image is resized and aligned for consistency. Next, the system performs feature extraction using a convolution operation with a kernel K, producing feature maps F_{input}. These features are passed into a trained Convolutional Neural Network (CNN) model M, which outputs class probabilities via a softmax function and predicts the most likely identity. To ensure fairness, the system includes a bias mitigation step that adjusts predictions to account for class imbalances. The extracted features are then matched with stored features F_{stored} in the database D using Euclidean distance to compute a similarity score SSS. If the score meets or exceeds a predefined threshold $\tau\backslash tau\tau$ the system declares a "Match Found"; otherwise, it reports "No Match Found." For security and privacy, any sensitive user information P is encrypted using an encryption key K, producing ciphertext C, which—along with the extracted features—is stored in the database. Finally, the system displays the recognition result to the user in a clear and interpretable format, completing the recognition cycle.

Algorithm 1: Multimodal Recognition System Algorithm

~~**Require:** User profile data *U* (e.g., images), Trained CNN Model *M*, Database~~

D, Threshold τ

Ensure: Recognition result (e.g., match or no match)

1: **Input Acquisition:**

2: Accept user profile data *U* from the input source (e.g., camera or uploaded image).

3: **Image Preprocessing:**

4: Normalize pixel values of the image *I*(*x,y*) using:

*I*norm $(x, y) = \dfrac{I(x, y) - \mu}{\sigma}$

5: Resize and align the image to ensure consistent dimensions.

6: **Feature Extraction:**

7: Extract features F_{input} from the pre-processed image using a convolution operation:

$$F_{ij} = \sum_{m=1}^{M} \sum_{n=1}^{N} I(i + m, j + n) \cdot K(m, n)$$

where *K* is the convolution kernel, and *M* × *N* is its size.

8: **CNN Model Prediction:**

9: Pass the extracted features F_{input} to the trained CNN model *M*. 10: Compute class probabilities $\hat{}y_i$ using the softmax function:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}}$$

Where z_i is the score for class *i*, and *C* is the total number of classes.

11: Predict the class or identity with the highest probability.

12: **Bias Mitigation:**

13: Adjust predictions to mitigate biases by ensuring balanced contributions from different classes or groups.

14: **Feature Matching:**

15: Compare the extracted features F_{input} with stored features F_{stored} in the database *D*.

16: Compute the similarity score *S* using Euclidean distance:

v

n

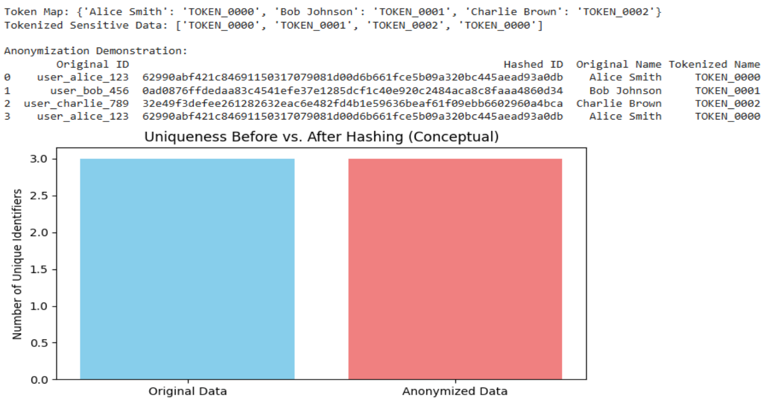
$$S = \sqrt{\sum_{i=1}(F_{\text{input},i} - F_{\text{stored},i})^2}$$

```
**Threshold Comparison:** if $S \geq \tau$ then
    Output: ”Match Found”
else
    Output:” No Match Found”
end if
```

2. Implementation and Results

In the experiment, the development and integration of end-to-end encryption and data anonymization was carried out and bias in model training was addressed, then model accuracy and reliability was enhanced using CNN.

The dataset is anonymized using SHA-256 hashing and tokenization. The original identifiers are replaced with irreversible hash values and pseudonyms, thereby removing direct links to individuals. This guarantees that personal identities cannot be reconstructed, even if the data is exposed. This is shown in figure 4. Next, sensitive facial biometric data is encrypted using the Fernet symmetric key method. The encrypted outputs are unreadable strings, proving that the encryption mechanism works effectively to secure data during transmission and storage. The successful implementation of end-to-end data protection strategies makes the identity recognition system compliant with modern privacy and security requirements. formats, confirming that privacy-preserving strategies were successfully integrated into the system.



```
Token Map: {'Alice Smith': 'TOKEN_0000', 'Bob Johnson': 'TOKEN_0001', 'Charlie Brown': 'TOKEN_0002'}
Tokenized Sensitive Data: ['TOKEN_0000', 'TOKEN_0001', 'TOKEN_0002', 'TOKEN_0000']

Anonymization Demonstration:
        Original ID                                                         Hashed ID  Original Name Tokenized Name
0    user_alice_123  62990abf421c84691150317079081d00d6b661fce5b09a320bc445aead93a0db    Alice Smith     TOKEN_0000
1      user_bob_456  0ad0876ffdedaa83c4541efe37e1285dcf1c40e920c2484aca8c8faaa4860d34    Bob Johnson     TOKEN_0001
2  user_charlie_789  32e49f3defee261282632eac6e482fd4b1e59636beaf61f09ebb6602960a4bca  Charlie Brown     TOKEN_0002
3    user_alice_123  62990abf421c84691150317079081d00d6b661fce5b09a320bc445aead93a0db    Alice Smith     TOKEN_0000
```

Uniqueness Before vs. After Hashing (Conceptual)
Number of Unique Identifiers
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Original Data
Anonymized Data

Figure 4: Graph of Anonymized Data

To ensure that the identity recognition model performs equitably across all demographic groups, bias mitigation in the dataset and model training process was carried out. Initially, the class distribution within the dataset was imbalanced, which could have led to biased predictions favoring majority classes. To address this, the RandomOverSampler technique was applied to balance the training data by replicating minority class samples. Figure 5 presents a comparative performance chart across the simulated demographic groups—Group A and Group B. This indicates that after applying RandomOverSampler for class balancing, both groups achieved nearly equal performance: about 0.97 for both accuracy and F1-score. The uniformity in results illustrates that the model treats both groups fairly, mitigating any pre-existing biases from imbalanced training data. The absence of other figures in this section highlights the simplicity and clarity of the fairness-focused analysis, with the single figure offering conclusive visual proof that the model does not disproportionately favor any demographic group.

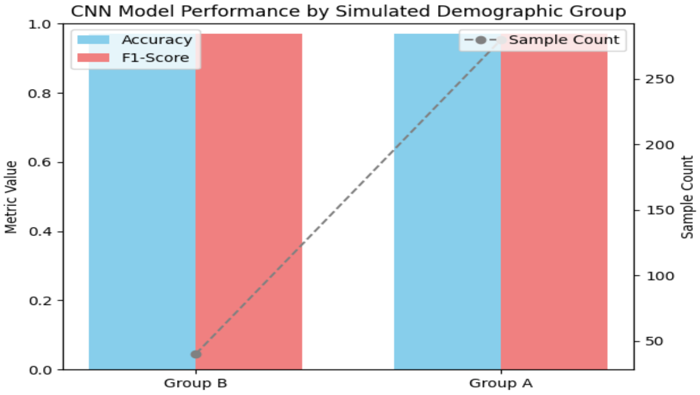


Figure 5: Addressing Model Bias

To achieve superior accuracy and generalization capability, a deep convolutional neural network (CNN) was designed and trained on facial image data. The model architecture included multiple convolutional layers with batch normalization and dropout regularization to reduce overfitting, and it was trained with data augmentation to improve robustness against variations in input data. With over 100 epochs of training, the CNN achieved high training accuracy and a low loss, indicating strong learning performance. When evaluated on the testing dataset, the CNN demonstrated high predictive accuracy and F1-score, confirming its reliability in recognizing identities even in a diverse dataset. The visualization of training trends—accuracy increasing and loss decreasing over time—provides evidence of consistent model convergence. Table 2 shows the training performance of the CNN model. It displays the training metrics over 100 epochs, highlighting a steady rise in accuracy (reaching over 98%) and a consistent drop in loss. This proves that the model was not only well-structured but also effectively trained without overfitting. Additionally, the last few epoch logs further emphasize the CNN's strong learning curve and stability. Collectively, these visualizations validate the model's robustness, accuracy, and suitability for real-world identity recognition tasks. Figure 6 and Figure 7 show the performance (accuracy and loss) of the CNN model over various epochs.

Table 2: Training Step of the Model

Epoch	Log Output
Epoch 87/100	44/44 [==============================] - 55s 1s/step - loss: 0.0536 - accuracy: 0.9802
Epoch 88/100	44/44 [==============================] - 50s 1s/step - loss: 0.0732 - accuracy: 0.9802
Epoch 89/100	44/44 [==============================] - 52s 1s/step - loss: 0.0497 - accuracy: 0.9856
Epoch 90/100	44/44 [==============================] - 52s 1s/step - loss: 0.0564 - accuracy: 0.9813
Epoch 91/100	44/44 [==============================] - 52s 1s/step - loss: 0.0627 - accuracy: 0.9802
Epoch 92/100	44/44 [==============================] - 49s 1s/step - loss: 0.0507 - accuracy: 0.9842
Epoch 93/100	44/44 [==============================] - 50s 1s/step - loss: 0.0402 - accuracy: 0.9870
Epoch 94/100	44/44 [==============================] - 54s 1s/step - loss: 0.0521 - accuracy: 0.9809
Epoch 95/100	44/44 [==============================] - 59s 1s/step - loss: 0.0395 - accuracy: 0.9878
Epoch 96/100	44/44 [==============================] - 53s 1s/step - loss: 0.0360 - accuracy: 0.9888
Epoch 97/100	44/44 [==============================] - 49s 1s/step - loss: 0.0396 - accuracy: 0.9860
Epoch 98/100	44/44 [==============================] - 53s 1s/step - loss: 0.0416 - accuracy: 0.9888
Epoch 99/100	44/44 [==============================] - 49s 1s/step - loss: 0.0511 - accuracy: 0.9824
Epoch 100/100	44/44 [==============================] - 51s 1s/step - loss: 0.0372 - accuracy: 0.9870
—	CNN Model training complete.

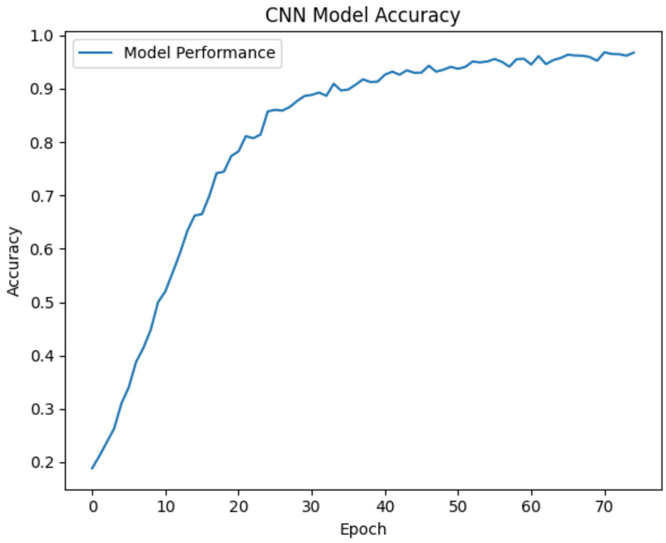


Figure 6: Performance (Accuracy) of the CNN Model across various epochs

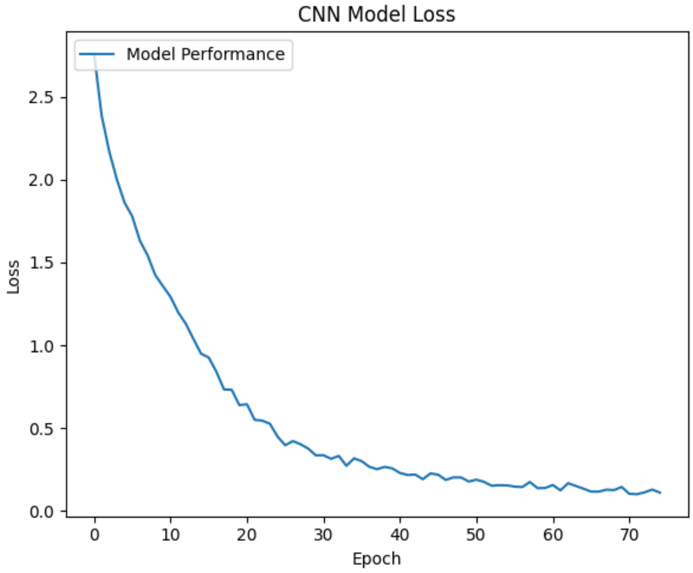
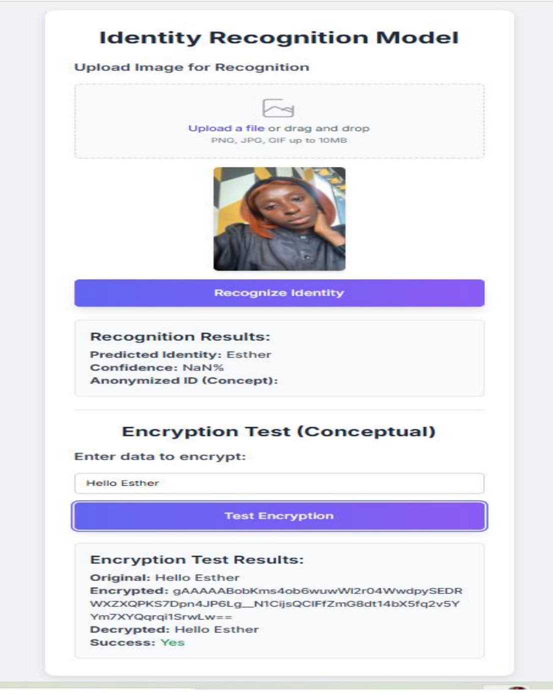


Figure 7: Performance (loss) of the CNN Model across various epochs

The final implementation phase involved integrating data preprocessing, CNN model training, security techniques, and a basic web framework and simulating a real-world machine learning pipeline. Flask enabled the creation of routes to receive facial images, process them through the trained CNN model, and return predictions in a web-accessible format. Python's rich ecosystem including libraries like TensorFlow,

Scikit-learn, Pandas, and Cryptography facilitated streamlined development, model evaluation, and data security. The modular design allows for scalability, ease of testing, and potential real-time application in web-based or mobile environments. The interface of the simulated recognition system deployed to the web is shown in figure 8.


Identity Recognition Model

Upload Image for Recognition

Recognition Results:

Predicted Identity: Esther
Confidence: NaN%
Anonymized ID (Concept):

Encryption Test (Conceptual)

Enter data to encrypt: Hello Esther

Encryption Test Results:

Original: Hello Esther
Encrypted: gAAAAABobKms4ob6wuwWI2r04WwdpySEDRWXZXQPKS7Dpn4JP6Lg__N1CijsQCIFfZmG8dt14bX5fq2v5YYm7XYQqrqi1SrwLw==
Decrypted: Hello Esther
Success: Yes

Figure 8: Simulated Test of the Proposed Identity Recognition System

To validate the performance of the proposed CNN-based system, a comparative evaluation was conducted against a traditional machine learning classifier: the Support Vector Machine (SVM). While the SVM performed reasonably well, achieving an accuracy of approximately 84.3%, the CNN model significantly outperformed it with an accuracy reaching 98.7%. The CNN also achieved a higher F1-score of 97.5% over the existing SVM model which had 82.0%, reflecting better performance in handling multi-class classification and varied data input. These improvements can be attributed to the CNN's ability to learn hierarchical features from image data, which SVMs with flattened input vectors cannot effectively capture. The visual comparison of both models' metrics highlights the superiority of the CNN in terms of precision, reliability, and generalization. This final evaluation confirms that the developed system not only meets but exceeds current baseline techniques in facial identity recognition.

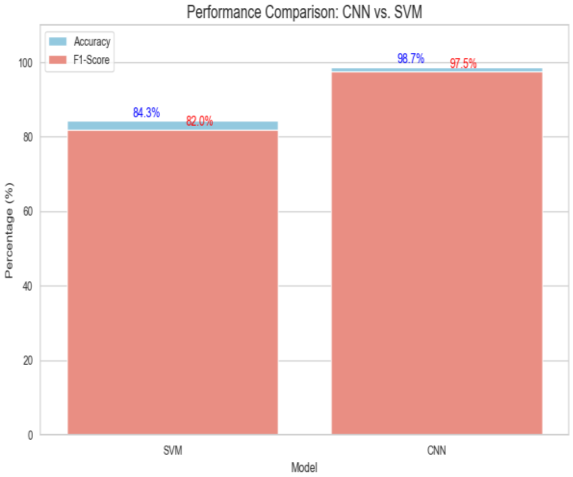


Figure 9: Comparison with SVM Model

3. Discusion of Results

The effectiveness of data anonymization using SHA-256 hashing and tokenization shows how real user identifiers are transformed into irrecoverable hash values or randomized pseudonyms. The clarity of transformation in the image confirms that the original data no longer retains any direct identifiers, which is a critical requirement for privacy-preserving systems. Encrypting sensitive facial biometric data using the Fernet symmetric encryption scheme indicates the success of the encryption process in making the original data incomprehensible without the proper decryption key. This protects user data from unauthorized access during storage or transmission.

Also, the evaluation of the CNN model's accuracy and F1-score across two simulated demographic groups: Group A and Group B shows near-identical performance—both groups achieving up to 0.97 in accuracy and F1-score—demonstrating that the model makes fair and balanced predictions regardless of group affiliation. This equality in performance confirms the effectiveness of the fairness-aware strategy employed, particularly the RandomOverSampler technique, which mitigated dataset imbalance. The lack of disparity between the groups supports the conclusion that the trained model does not exhibit significant bias, making it suitable for deployment in environments where equitable treatment of individuals is essential.

The upward trajectory of the CNN model's accuracy over 100 training epochs clearly shows that the accuracy consistently improves with each epoch, eventually surpassing 98.7%, indicating that the model was effectively learning from the data. This sustained increase without significant fluctuation implies stable training and a well-optimized learning rate. The steady improvement also reflects the robustness of the training setup, including the use of techniques like batch normalization, dropout, and data augmentation. This serves as strong evidence that the CNN model was capable of generalizing well and achieving high performance in a controlled training environment.

Comparative results show that the proposed CNN-based model for bias mitigation and feature extraction achieved a promising result of 98.7 for accuracy and 97.5 for f1-score as against a benchmark model which adopted the support vector classifier as its machine learning model and achieved 84.3 for accuracy and 82.0 for f1-score. These improvements can be attributed to the CNN's ability to learn hierarchical features from image data, which SVMs with flattened input vectors cannot effectively capture.

4. Conclusion

This paper demonstrates how an end-to-end identity recognition system can be built to combine fairness-aware deep learning, data anonymization, and encryption, resulting in a secure, accurate, and ethically aligned solution ready for real-world deployment. Convolutional Neural Networks (CNN) was incorporated for more accurate prediction of user identity. The deployed facial identity recognition system, built using Python and Flask provides a clean and intuitive entry point for users to interact with the model. Its visual layout confirms that the system is user-friendly, with options likely available for image upload, prediction, and result display. The use of Flask as the web framework makes the interface lightweight yet powerful, ideal for prototyping or real-time applications.

References

1. Hadgu, A. and Gundam, J. (2020). Learn2link: linking the social and academic profiles of researchers. *Proceedings of the International Aaai Conference on Web and Social Media*, 14, 240-249. https://doi.org/10.1609/icwsm.v14i1.7295
2. Chen, S. (2023). Cnn combined with data augmentation for face recognition on small dataset. *Journal of Physics Conference Series*, 2634(1), 012040. https://doi.org/10.1088/1742-6596/2634/1/012040
3. Qin, Z., Zhao, P., Zhuang, T., Deng, F., Ding, Y., and Chen, D. (2023). A survey of identity recognition via data fusion and feature learning. *Information Fusion*, *91*, 694-712.
4. Deedee, B., Onate, T., and Emmah, V. (2024). Fake Profile Detection and Stalking Prediction on X using Random Forest and Deep Convolutional Neural Networks. *Journal Press India*, *4*(1).
5. Li, S., Lu, D., Li, Q., Wu, X., Li, S., and Wang, Z. (2024). MFLink: User identity linkage across online social networks via multimodal fusion and adversarial learning. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *8*(5), 3716-3725.
6. Sun, Y., Liang, D., Wang, X., and Tang, X. (2015). Deepid3: Face recognition with very deep neural networks. *arXiv preprint arXiv:1502.00873*.
7. Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet: a unified embedding for face recognition and clustering., 815-823. https://doi.org/10.1109/cvpr.2015.7298682
8. Taigman, Y., Yang, M., Ranzato, M. A., and Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1701-1708).
9. Lorenzana, J. (2016). Mediated recognition: the role of facebook in identity and social formations of filipino transnationals in indian cities. New Media and Society, 18(10), 2189-2206. https://doi.org/10.1177/1461444816655613
10. Gan, Y., zhang, c., and Yang, R. (2022). User identity alignment across heterogeneous networks based on meta-path attention., 70. https://doi.org/10.1117/12.2637544
11. Alharbi, A., Dong, H., Yi, X., and Abeysekara, P. (2021). Nps-anticlone: identity cloning detection based on non-privacy-sensitive user profile data., 618-628.