iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.82207

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Personal Data Leak Detection System

Asish Mundappallil, Deepak Ravi, Kevin Sebastian
Computer Science Mar Baselios Christian College of Engineering and Technology Kuttikanam, Kerala, India

Abstract: ***The increasing number of data breaches and unauthorized data exposures has raised serious concerns regarding personal data security in the digital era. Sensitive user information, including email credentials, personal identifiers, and biometric data, is frequently exposed due to insecure systems, large-scale data breaches, and improper data handling practices. Such exposures often remain unnoticed by users, leading to significant privacy risks and potential misuse of personal information. This paper proposes an AI-based Personal Data Leak Detection System designed to identify and analyze potential data leaks associated with individual users. The system integrates intelligent data processing techniques, biometric analysis, and pattern recognition methods to detect exposed information across structured and unstructured datasets. Facial recognition models are utilized to identify potential biometric leaks, while additional analytical modules evaluate user-related data exposure and associated risks. The system is implemented using a Flutter-based frontend and a FastAPI backend, with Firebase authentication ensuring secure and isolated user access. Experimental observations indicate that the proposed system effectively detects user-related data leaks and enhances user awareness regarding privacy risks. The results demonstrate the potential of AI-driven approaches in strengthening digital privacy and improving cybersecurity mechanisms.***

Keywords: Artificial Intelligence, Data Leak Detection, Cybersecurity, Facial Recognition, Deepfake Detection, Privacy Protection, Biometric Security, Data Breach Analysis.

I. INTRODUCTION

The rapid growth of digital technologies and online platforms has led to an unprecedented increase in the collection and storage of personal data. Users routinely share sensitive information such as email addresses, personal identifiers, and biometric data across various applications and services. However, this widespread data usage has also resulted in a significant rise in data breaches, unauthorized access, and information leaks. In many cases, users remain unaware that their data has been exposed, making them vulnerable to identity theft, fraud, and privacy violations.

Traditional approaches to data security primarily focus on prevention through encryption and access control mechanisms. While these methods are essential, they often fail to address the problem of already leaked data. Existing systems that attempt to detect data breaches are generally limited in scope and often rely on manual checks or predefined databases. Furthermore, such systems rarely incorporate advanced analysis techniques capable of identifying complex patterns in leaked data or detecting exposure of biometric information.

Recent advancements in Artificial Intelligence (AI) and machine learning have enabled the development of intelligent systems capable of analyzing large volumes of data and identifying meaningful patterns. Techniques such as facial recognition, natural language processing, and deep learning-based media analysis have shown promising results in various domains of cybersecurity. These technologies can be leveraged to detect potential data leaks, analyze user-specific information, and identify manipulated or synthetic media content.

This paper proposes an AI-based Personal Data Leak Detection System designed to identify and analyze potential data leaks associated with individual users. The system integrates multiple components, including biometric recognition, data pattern analysis, and media verification, to provide a comprehensive approach to personal data security. By combining these techniques, the system aims to enhance user awareness, improve detection accuracy, and provide a reliable solution for identifying data exposure in modern digital environments.

II. RELATED WORK

Several research efforts have explored the use of machine learning and Artificial Intelligence techniques for detecting data breaches and managing sensitive information. Early approaches to data leak detection primarily relied on rule-based systems and traditional database matching methods, which were limited in their ability to identify complex or unstructured data exposures. These systems often required manual intervention and lacked scalability when dealing with large volumes of data.

©IJRASET: All Rights are Reserved

With the advancement of machine learning, more sophisticated techniques have been introduced for analyzing and detecting data leaks. Studies have utilized classification algorithms and pattern recognition methods to identify exposed credentials such as email addresses and passwords. Additionally, Natural Language Processing (NLP) techniques have been applied to analyze textual data and detect sensitive information embedded within large datasets. While these approaches improved detection accuracy, they often focused on specific types of data and did not provide a comprehensive solution.

In the domain of biometric security, facial recognition systems based on deep learning models such as convolutional neural networks and architectures like ArcFace have demonstrated high accuracy in identity verification tasks. These systems generate feature embeddings that allow efficient comparison and matching of facial data. However, most existing implementations are designed for authentication purposes rather than identifying leaked biometric data across external datasets.

Recent developments in deepfake detection have introduced hybrid models that combine multiple techniques to identify synthetic or manipulated media. These approaches analyze visual artifacts, inconsistencies in facial features, and metadata signatures to distinguish between real and AI-generated content. Although effective, many of these systems operate independently and are not integrated into broader data security frameworks.

Despite these advancements, existing solutions typically focus on a single aspect of the problem, such as credential leak detection, biometric authentication, or media verification. There is a lack of unified systems that combine these capabilities into a single platform for comprehensive personal data protection. The proposed system addresses this gap by integrating data leak detection, biometric analysis, and media verification into a cohesive AI-based framework.

III. PROPOSED SYSTEM

The proposed AI-based Personal Data Leak Detection System is designed to provide a comprehensive solution for identifying and analyzing potential data leaks associated with individual users. The system integrates multiple intelligent modules that work together to detect exposed personal information, analyze biometric data, and identify manipulated or synthetic media. Unlike traditional approaches that focus on a single aspect of data security, the proposed system combines data leak detection, facial recognition, and media analysis into a unified framework.

The system operates by collecting user-specific data through a secure and authenticated interface. Users can provide input in the form of email identifiers, biometric data such as facial images or videos, and media files for analysis. The collected data is processed using advanced Artificial Intelligence techniques to determine whether any portion of the user's information has been exposed or misused. For biometric analysis, facial recognition models are used to generate feature embeddings, which are compared against stored data representations to identify potential matches. This enables the system to detect whether a user's biometric data has been compromised or appears in unauthorized datasets.

In addition to biometric detection, the system incorporates data pattern analysis techniques to identify exposed credentials and personal identifiers. The analysis process examines both structured and unstructured data to determine whether user-related information is present in known leak datasets. Furthermore, the system includes a media analysis module that evaluates uploaded images and videos to detect signs of manipulation or synthetic generation using deep learning-based models.

To ensure responsible usage and prevent misuse, the system enforces controlled data acquisition mechanisms. Biometric inputs are restricted to real-time capture through the application interface, preventing the use of pre-existing media from device storage. This design choice helps maintain user consent and reduces the risk of unauthorized scanning of third-party data.

The final output of the system includes detection results, risk evaluation, and user notifications. By integrating multiple detection techniques into a single platform, the proposed system provides a more effective and user-centric approach to personal data security, enhancing awareness and enabling timely action against potential data exposure.

IV. SYSTEM ARCHITECTURE

The system architecture of the proposed AI-based Personal Data Leak Detection System is designed as a multi-layered framework that ensures secure data handling, efficient processing, and accurate detection of potential data leaks. The architecture consists of several interconnected layers, including the user interface layer, authentication layer, processing layer, detection layer, and data storage layer, each responsible for a specific function within the system.

The user interface layer is developed using Flutter and serves as the primary interaction point for users. It allows users to input data, capture biometric information, and upload media files for analysis. The interface is designed to provide a smooth and secure user experience while enabling access to different system functionalities such as data leak checking, facial scanning, and media analysis.

The authentication layer ensures secure access to the system using Firebase-based authentication mechanisms.

Each user is assigned a unique identity, and all requests to the backend are verified using secure tokens. This layer enforces user-specific data isolation, ensuring that personal information and biometric data are not shared or accessible across different users.

The processing layer is responsible for handling and preparing the input data for analysis. It performs operations such as data preprocessing, feature extraction, and transformation. For biometric data, facial images and video frames are processed to generate feature embeddings. For textual data, relevant patterns and identifiers are extracted for further evaluation.

The detection layer forms the core of the system and integrates multiple analytical modules. The facial recognition module compares generated embeddings with stored representations using similarity measures to identify potential matches. The data analysis module evaluates structured and unstructured datasets to detect exposed personal information. Additionally, the media analysis module utilizes deep learning techniques to identify manipulated or synthetic content in uploaded images and videos.

The data storage layer is responsible for securely storing user-related data and processed information. Lightweight database systems such as SQLite, along with cloud-based services like Firebase, are used to maintain data efficiently. Sensitive data is handled with appropriate security measures to ensure privacy and integrity.

Overall, the layered architecture enables modularity, scalability, and secure data processing, allowing the system to efficiently detect and analyze personal data leaks while maintaining strict user data protection.

V. IMPLEMENTATION

The proposed AI-based Personal Data Leak Detection System is implemented using a combination of modern frontend and backend technologies to ensure efficient processing and secure data handling. The user interface is developed using Flutter, providing a responsive and platform-independent environment for user interaction. The application enables users to securely input personal data, capture biometric information in real time, and upload media files for analysis. The frontend communicates with the backend through authenticated API requests, ensuring that all data exchanges are secure and user-specific.

The backend is developed using the FastAPI framework in Python, which provides an asynchronous and high-performance environment for handling multiple requests and processing large data inputs. Firebase Authentication is integrated into the system to manage user identity and enforce secure access control. Each request from the frontend includes a verified authentication token, which is validated at the backend to ensure that only authorized users can access the system's functionalities. This mechanism also enables strict isolation of user data, preventing unauthorized access or data leakage across different user accounts.

For biometric analysis, the system utilizes deep learning-based facial recognition techniques to generate feature embeddings from captured images or video frames. These embeddings represent unique facial characteristics and are stored in a serialized format for efficient comparison. During the detection process, similarity measures are used to compare user-generated embeddings with stored data representations to identify potential matches. This approach allows the system to detect whether a user's biometric data appears in unauthorized datasets.

The system also incorporates data analysis techniques to identify potential exposure of user-related information such as email addresses and other identifiers. Pattern matching and data processing methods are applied to analyze both structured and unstructured data sources. Additionally, the media analysis component employs deep learning-based models to evaluate uploaded images and videos for signs of manipulation or synthetic generation. Multiple analysis techniques are combined to improve detection reliability and reduce false positives.

All processed data and results are stored using lightweight database solutions such as SQLite, along with cloud-based services provided by Firebase. Sensitive information is handled with appropriate security measures to ensure data integrity and privacy. Overall, the implementation integrates multiple technologies and AI techniques into a cohesive system capable of detecting and analyzing personal data leaks efficiently and securely.

VI. RESULTS AND DISCUSSION

The data analysis module was capable of detecting sensitive information such as email identifiers and personal data patterns through integration with external breach databases and internally stored datasets. The system utilizes APIs such as Have I Been Pwned (HIBP) to identify known data breaches associated with user-provided information. Additionally, pattern recognition techniques are applied to analyze available data sources and identify potential exposure scenarios. The effectiveness of detection depends on the availability and quality of accessible data, which can be further enhanced by expanding the dataset over time.

VII. CONCLUSION AND FUTURE WORK

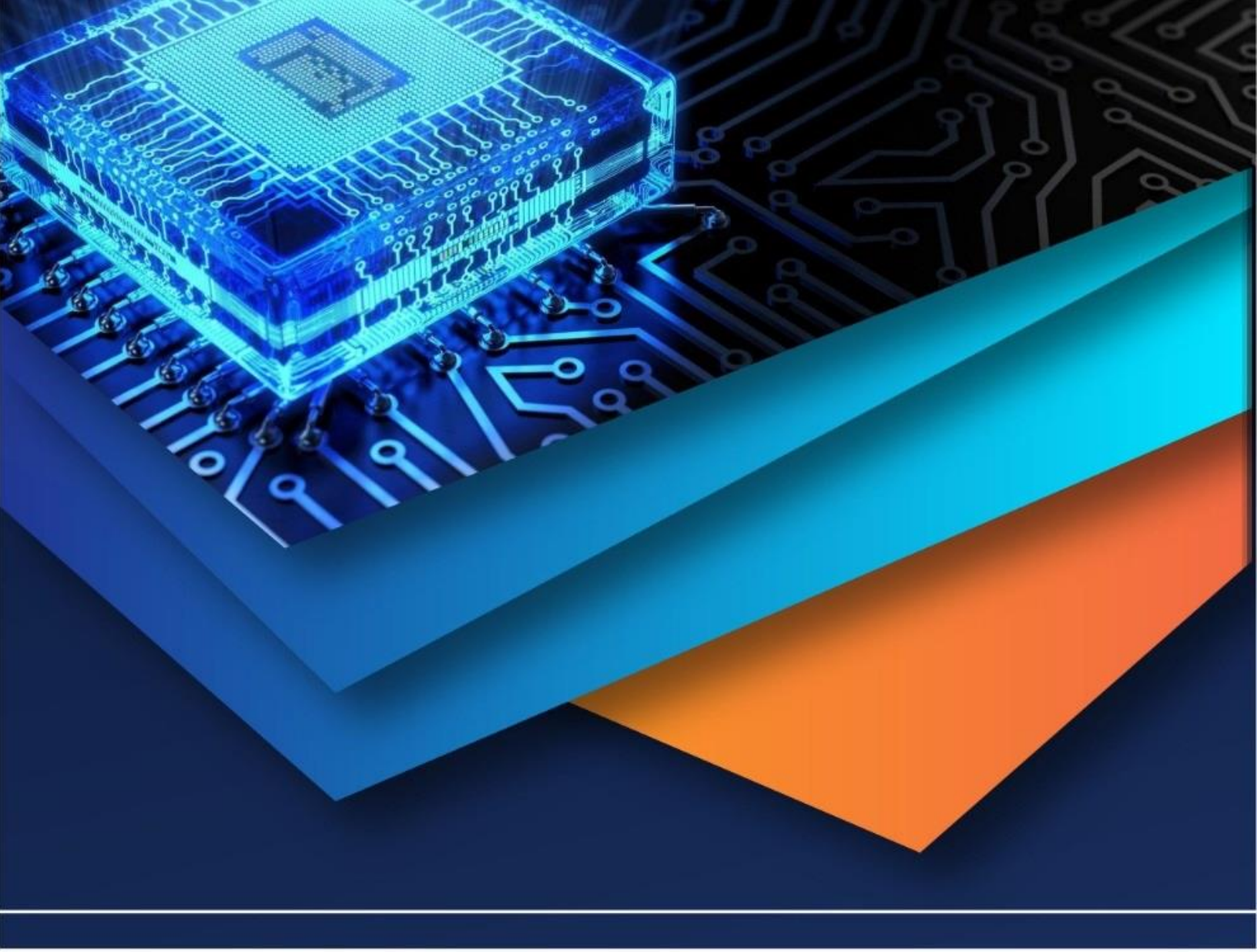
This paper presented an AI-based Personal Data Leak Detection System designed to identify and analyze potential data exposure associated with individual users. The system integrates techniques such as biometric recognition, data pattern analysis, and media verification to provide a comprehensive approach to personal data security. The results demonstrate that the system is capable of detecting user-related data leaks using available data sources while maintaining secure and user-specific data handling.

Future work can focus on improving detection accuracy, expanding data sources, and enabling real-time monitoring capabilities. Additional enhancements such as advanced user consent mechanisms and performance optimization can further strengthen the system for large-scale deployment

REFERENCES

[1] J. Deng et al., “ArcFace: Additive Angular Margin Loss for Deep Face Recognition,” in Proceedings of CVPR, 2019.
[2] H. Nguyen et al., “Deep Learning for Deepfakes Creation and Detection,” arXiv preprint arXiv:1909.11573.
[3] M. Abadi et al., “Deep Learning with Applications in Security and Privacy,” Communications of the ACM.
[4] A. Narayanan et al., “Machine Learning Approaches for Data Leak Detection,” IEEE Security & Privacy.
[5] Have I Been Pwned, “Data Breach Search API,” [Online]. Available: https://haveibeenpwned.com/API.
[6] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” Nature, 2015.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)