Report On

AI/ML-Based Speaker Identification

Using
Custom Made

FraudGuard Website for Cybercrime and Fraud Detection

A Project For

IIITH Speech Analytics

By:-

Sayali Bambal
Rochan Awasthi
Ananya Rajurkar

Contents

1. Introduction

Speaker identification is a transformative application of Artificial Intelligence (AI) and Machine Learning (ML) that has gained increasing relevance in today's digital world. As society shifts toward voice-activated systems for both convenience and security, the need to accurately identify speakers based on their unique vocal characteristics becomes paramount. This technology offers a new, secure, and efficient way of authentication, making it an indispensable tool in cybersecurity. Voice biometrics, which focuses on individual vocal traits such as pitch, cadence, and accent, serves as a natural form of identification, akin to fingerprints but far more accessible and seamless. It's poised to become a cornerstone in secure systems, replacing traditional password and PIN-based authentication methods. The practical applications of speaker identification extend far beyond personal security; they are revolutionizing industries such as banking, law enforcement, and customer service. In the realm of **cybersecurity**, speaker identification provides a robust defense against fraud and identity theft. By analyzing voiceprints, it ensures that only authorized users can access sensitive services, such as banking transactions or confidential accounts. The **FraudGuard** website, developed as part of this project, takes speaker identification a step further by integrating it into a real-time fraud detection system. FraudGuard offers an effective solution to prevent cybercrimes such as voice phishing (vishing) and unauthorized access to secure services. This fusion of AI-driven voice recognition and fraud detection makes this project highly relevant and impactful in today's ever-evolving digital landscape, ensuring that organizations can combat fraud proactively and securely.

---

2. Dataset

The data set used in this project is a combination of two data set. First part is from Kaggle which has 5 Speakers/Classes. Second part is a custom data set which we collected on our own, it has 3 Speakers/Classes.We also trained on 1 hour of noise data so that the model can differentiate between background noise and a human The whole data set consists of 8 different Speakers or Classes this provides a wide range of vocal characteristics essential for training a highly effective speaker identification model.
Each class has 1500 total recordings out of which we used 1400 for training purpose and 100 for testing purpose.The recordings were made in a controlled environment to minimize external noise, ensuring clarity and accuracy of the speech data. The total duration of the dataset is approximately 260 minutes. As this is a prototype therfore we used only this much data but we can further expand it.

Dataset Statistics	Value
Number of Speakers	8
Number of Utterances	1,500+ per dataset
Total Training Utterances	1,400
Total Testing Utterances	100

Dataset Statistics	Value
Total Duration	260 minutes
Environment	Controlled, minimal noise

The dataset's design and quality are vital to the success of the speaker identification model. It includes voices from different speakers—representing various genders, ages, and vocal traits—ensuring that the model is capable of identifying speakers across a wide range of conditions. The controlled environment in which the data was recorded further enhances the reliability and precision of the model, minimizing noise and ensuring that the training data is of the highest quality.

---

3. Experimental Setup

The experimental setup for the speaker identification model is based on deep learning methodologies, particularly **Convolutional Neural Networks (CNNs)**, which are particularly well-suited for sequential data like audio. Feature extraction plays a critical role in this setup, where **Mel-Frequency Cepstral Coefficients (MFCCs)** are used to capture the most important spectral features of the speakers' voices. These features are then fed into the CNN model, allowing the system to learn and distinguish the unique characteristics of each speaker's voice.
The architecture of the model consists of several convolutional layers that progressively learn higher-level features from the audio spectrograms. The model is trained using the **Adam Optimizer** with an initial learning rate of 0.001 for **15 epochs(limited due to computer performance)**, using a **batch size of 32** to ensure stable learning while maintaining efficiency. The loss function chosen for this model is **cross-entropy loss**, which is optimal for multi-class classification tasks like speaker identification. To improve the model's generalization ability, **data augmentation techniques**, such as noise addition and speech speed variations, are used to expose the model to a range of possible real-world conditions.
In addition to traditional speaker identification, the FraudGuard platform enhances the model by incorporating **real-time fraud detection capabilities**, which ensures that voice-based transactions or interactions are legitimate. By integrating voice authentication with fraud monitoring systems, FraudGuard can proactively detect fraudulent attempts and prevent unauthorized access to sensitive accounts or services, making it a critical tool for industries at risk of cybercrime.

---

4. Results

The performance of the CNN-based model was evaluated using key metrics: **accuracy**, **precision**, **recall**, and **F1-score**. The results were compared to traditional methods such as **Gaussian Mixture Models (GMM)** and **Support Vector Machines (SVM)**. The CNN model demonstrated superior performance across all metrics, with a higher **accuracy**, **precision**, and **recall**, thus ensuring a more reliable and effective system for speaker identification.

Model	Accuracy
Version 4	76.8%

The superior performance of the CNN-based model can be attributed to its ability to learn complex and hierarchical features from raw audio spectrograms, compared to traditional models like GMM and SVM, which rely on handcrafted features. Furthermore, the model was robust to noise and variations in speech, thanks to the data augmentation techniques employed during training. FraudGuard utilizes this high-performing model to enhance its fraud detection capabilities, ensuring that only authorized individuals can access sensitive services.

---

5. FraudGuard Website: Revolutionizing Cybercrime and Fraud Detection

The **FraudGuard** website, a key component of this project, integrates the power of AI-driven speaker identification with a real-time fraud detection system, revolutionizing the way organizations handle voice-based authentication. FraudGuard enhances security by offering an intuitive, seamless authentication process based on voice biometrics, while also enabling real-time fraud monitoring. This makes the platform an ideal solution for financial institutions, government agencies, and any organization that requires secure, fraud-free access to sensitive data or services.
FraudGuard excels in combating **cybercrime** and **voice-based fraud** by identifying impersonators in voice-driven interactions. For example, in the banking sector, FraudGuard can detect unauthorized attempts to access accounts by analyzing the voiceprint of the individual making the request. It also prevents **vishing** (voice phishing), where fraudsters use deception to steal personal information over the phone. The website's real-time detection system continuously verifies the authenticity of voice transactions, enabling organizations to act immediately if a fraudulent voice attempt is detected.
By seamlessly integrating **voice authentication** with fraud detection, FraudGuard provides businesses with a powerful tool to mitigate the risks associated with voice-based cybercrimes. Its ability to instantly validate the identity of callers and authenticate voice transactions makes it an essential asset for enhancing digital security across various sectors, from banking to e-commerce, ensuring that fraud is identified and prevented before it can cause harm.

---

6. Conclusion

In conclusion, this project has demonstrated the transformative potential of **AI** and **ML** in addressing the critical challenge of speaker identification. The **CNN-based model** outperformed traditional methods, delivering exceptional accuracy and robustness. More significantly, the integration of speaker identification into the **FraudGuard** platform has created a powerful tool to combat **cybercrime** and **voice-based fraud**. FraudGuard not only provides a secure means of voice-based authentication but also strengthens the ability to detect and prevent fraud in real time, which is a game-changer for industries vulnerable to voice phishing and identity theft.

The success of this project emphasizes the crucial role that voice biometrics will play in the future of digital security. The **FraudGuard website** is an innovative solution that addresses a growing threat in the cybersecurity landscape, offering a seamless, effective way to safeguard sensitive services from fraud. Looking ahead, further advancements could involve refining the noise reduction algorithms and exploring **multi-modal biometric systems** that combine voice with other biometric features such as facial recognition. This would create even more secure and adaptable authentication systems capable of handling a wider range of fraud scenarios. With the rise of **AI-driven fraud detection**, the future of digital security looks brighter, and FraudGuard stands at the forefront of this change.