



Sri Lanka Institute of Information Technology

Behavioral Biometrics for Enhanced Authentication Systems

Project ID – 24-25J-073

Submitted by:

Student Registration Number	Student Name
IT21391668	H.N.D. Madhubhashana
IT21340864	E.M.N. Edirisinghe
IT21345678	K.G.A. Anupama
IT21336072	R.P.K.D. Rajapaksha

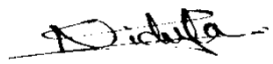
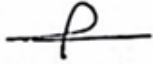

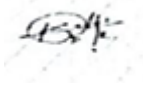
Department of Computer System Engineering

Date of submission

Saturday, April 12, 2025

Declaration Page of the Candidates & Supervisor

I declare that this is our own work, and this proposal does not incorporate, without acknowledgement, any material previously submitted for a degree or diploma in any other university or institute of higher learning, and to the best of our knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Group Member Name	Student ID	Signature
H.N.D. Madhubhashana	IT21391668	
E.M.N. Edirisinghe	IT21340864	
K.G.A. Anupama	IT21345678	
R.P.K.D. Rajapaksha	IT21336072	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of Supervisor:

Mr. Harinda Fernando

Date:

Abstract

Traditional authentication systems relying on passwords or single biometric modalities face increasing challenges due to sophisticated cyber threats, such as spoofing and credential theft. This research presents a multi-modal behavioral biometric authentication system combining Gait Analysis, Keystroke Dynamics, Mouse Movement Patterns, and Voice Authentication to address these limitations. Unlike physiological biometrics, these behavioral traits are passive, difficult to replicate, and ideal for continuous authentication.

Each modality was modeled using appropriate deep learning architectures: Convolutional Neural Networks (CNNs) for gait and voice data, and Recurrent Neural Networks (RNNs) for keystroke and mouse dynamics. Cosine similarity was used to calculate identity confidence scores, and weighted fusion aggregated these predictions, improving overall performance. Datasets from publicly available repositories were preprocessed for quality and alignment.

The system was evaluated with accuracy, precision, recall, F1-score, and AUC. While individual models performed well, the integrated system outperformed all standalone modalities, significantly reducing both False Acceptance Rate (FAR) and False Rejection Rate (FRR). The multi-modal fusion enhanced robustness by compensating for the limitations of each modality.

This system supports real-time, non-intrusive, hardware-agnostic deployment across various platforms, making it suitable for industries like banking, healthcare, and IoT. Privacy concerns were addressed by suggesting local processing, anonymized data storage, and encryption methods to protect user identity.

Overall, the multi-modal behavioral biometric system offers a secure, scalable, and user-friendly alternative to traditional authentication methods, providing a more robust approach to digital identity verification.

Acknowledgement

We would like to begin by expressing our heartfelt gratitude to our professors and advisers, whose invaluable guidance and support were fundamental throughout this project. Their patience, constructive feedback, and encouragement played a crucial role in shaping the success of our research. Our sincere thanks also extend to the many friends, teachers, and mentors who contributed to this work. Your collaboration, insightful comments, and encouragement have been a continuous source of inspiration and motivation. We are deeply grateful to the universities, museums, and libraries that provided us with access to the essential resources, enabling us to explore and analyze this complex subject matter.

Finally, we would like to express our profound appreciation to our families for their unwavering support, patience, and understanding throughout our academic journey. Your constant belief in us and your encouragement have been a driving force behind our efforts. It is with great gratitude that we acknowledge the collective contributions of all those who helped make this thesis possible.

Contents

Introduction.....	1
Background & Literature Survey	1
Research Gap.....	3
Research Problem.....	6
Objective	11
Main Objective	11
Specific Objectives.....	15
Methodology	19
Methodology Overview of the Multi-Modal Authentication System	19
Gait Analysis Methodology	25
Keystroke Dynamics Methodology.....	30
Mouse Movement Methodology	35
Voice Authentication Methodology	41
Results and Discussions.....	46
Results of Each Component	47
Discussions.....	55
Research Findings	60
Challenges.....	64
Quality of Data	65
Variability in Gait Patterns.....	66
Noise in Keystroke and Mouse Data.....	66
Integration Issues.....	66
Computational Resources.....	68
Privacy Concerns.....	69

Discussion	69
Comparison with Existing Biometric Systems.....	70
How Combining Behavioral Biometrics Provides a More Secure and User-Friendly Solution	71
Implications for Future Biometric Systems	73
Future Implementation.....	75
Real-Time Processing Enhancements	75
Further Multi-Modal Integrations	76
Commercialization Possibilities and Potential Applications	77
Privacy Concerns and Potential Solutions	78
Conclusion	79
Integrating Gait, Keystroke, Mouse, and Voice: A Comprehensive Approach to Authentication	80
Importance of Multi-Modal Biometrics in Modern Security Paradigms	81
Real-World Applicability and Deployment Potential	82
Looking Forward: Challenges and Opportunities	83
Final Reflections	84
References.....	85

List of Figures

Figure 1 - Cosine Similarity.....	21
Figure 2- Gait Dataset.....	25
Figure 3- Gait Model Architecture	27
Figure 4- Flight & Dwell Time.....	32
Figure 5 - Keystroke Model Architecture.....	33
Figure 6- Voice Model Architecture.....	43
Figure 7- Gait Confusion Matric.....	48
Figure 8- Kesytroke Confusion Matrix & AUC Curve	49
Figure 9 - Mouse ROC Curve.....	51
Figure 10 -Voice Result	52
Figure 11 - Multi Model Result	53

List of Tables

Table 1 - Research Gap.....	3
Table 2 - Gait Result.....	47
Table 3 - Keystroke Result	48
Table 4 - Mouse Result	50

List of Abbreviations

Abbreviations	Meaning
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
MFCC	Mel-Frequency Cepstral Coefficients
F1-Score	F1 Measure (Harmonic Mean of Precision and Recall)
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
SNR	Signal-to-Noise Ratio
MSE	Mean Squared Error
API	Application Programming Interface
IoT	Internet of Things
SaaS	Software as a Service
GPU	Graphics Processing Unit
HCI	Human-Computer Interaction

Introduction

Background & Literature Survey

Behavioral biometrics have gained significant attention in user authentication, offering a potential alternative or supplement to traditional methods such as passwords, PINs, and biometric identifiers like fingerprints and facial recognition. These methods—gait, keystroke dynamics, mouse movements, and voice—are rooted in unique patterns of human behavior, which are more difficult to replicate, or steal compared to conventional biometric identifiers. This makes them a compelling solution for improving security across various applications, including online banking and mobile device access[1].

Gait, as a behavioral biometric, refers to the unique way in which an individual walks. This can be analyzed through sensors like accelerometers or cameras. Gait is shaped by an individual's anatomy, habits, and health, offering a distinct characteristic that allows for continuous, passive authentication without requiring active engagement. Research has demonstrated that gait analysis is effective for ongoing authentication, as it enables identification while a person is moving through a space[1].

Keystroke dynamics involves studying an individual's typing behavior, such as speed, rhythm, and pressure. This type of biometric is increasingly popular for online security because it integrates easily with existing systems. Every individual types in a distinct manner, and keystroke dynamics measures not just the input but the way it is typed. Variations in speed, key hold duration, and typing errors provide a basis for distinguishing individuals. Studies show that keystroke dynamics can effectively identify users and detect unauthorized access through subtle yet consistent patterns[2].

Mouse dynamics, another form of behavioral biometric, involves analyzing how a user moves and clicks a mouse. This includes factors like movement speed, click frequency, and path trajectory. Much like keystroke dynamics, these patterns are typically unique to each user, influenced by factors like hand-eye coordination, posture, and cognitive traits.

Integrating mouse dynamics into authentication systems can strengthen security by observing real-time user interactions with devices[3].

Voice biometrics take advantage of the distinct characteristics of a person's voice, such as pitch, tone, and cadence. Unlike other biometrics, voice recognition can be performed from a distance and doesn't require user input, which is ideal for hands-free authentication scenarios. The complex nature of human speech, with its unique acoustics and speech patterns, makes it difficult to replicate, though environmental noise and health changes can impact accuracy[4].

Traditional methods like passwords, PINs, and fingerprints have long been used to secure digital systems but come with various drawbacks. Passwords are prone to being stolen, guessed, or cracked, and users often create weak or reused passwords across different platforms. While fingerprint recognition is generally secure, it can be bypassed if the sensor is compromised or if individuals lose their fingerprints due to injury. Both methods also depend on active user participation, which can create usability problems or lead to negligence, such as forgotten passwords.

In contrast, behavioral biometrics offer a more passive and continuous authentication process. Since they are based on individual patterns of behavior that are influenced by cognitive, physiological, and environmental factors, they are harder to replicate. This not only enhances security but also improves privacy, as no explicit user input is required, reducing the risk of data interception when compared to traditional methods.

There has been a growing interest in multi-modal authentication systems that combine several behavioral biometrics. These systems leverage the strengths of multiple biometric methods to create more secure and reliable authentication processes. For example, a system might combine gait analysis, keystroke dynamics, and voice recognition to achieve enhanced accuracy and security. Multi-modal systems can mitigate the limitations of individual biometrics, such as sensitivity to environmental factors or individual variations, improving overall reliability in real-world environments where one type of biometric might

be less effective[5].

Research on multi-modal systems has shown that combining different biometrics results in higher accuracy than using a single modality. For instance, pairing keystroke dynamics with voice biometrics helps to more accurately distinguish legitimate users from impostors. Additionally, multi-modal systems can adapt to a variety of scenarios, making them ideal for applications with diverse user behaviors and varying security requirements.

While promising, challenges remain in optimizing multi-modal systems for real-world applications. These include the need for large-scale datasets to train machine learning algorithms, integrating various sensor types, and addressing privacy concerns. Additionally, developing adaptive algorithms to account for changes in user behavior over time is a key focus for researchers aiming to enhance the usability and security of these systems.

Research Gap

Multi-modal behavioral biometric systems, which combine gait analysis, keystroke dynamics, mouse movements, and voice recognition, hold great promise for enhancing authentication mechanisms. However, several research challenges persist within each biometric modality and in their integration into a unified system. Addressing these issues is crucial to improving the accuracy, scalability, robustness, and usability of multi-modal systems.

Table 1 - Research Gap

Research	Accuracy	Occlusion Handling	Individual Variability	Privacy	Integration
Research A	Yes	No	No	Yes	No
Research B	Yes	Yes	No	Yes	Yes

Research C	Yes	Yes	Yes	Yes	No
Research D	Yes	Yes	Yes	Yes	Yes
Research E	Yes	No	Yes	Yes	Yes
Research F	Yes	Yes	Yes	No	Yes
Proposed Project	Yes	Yes	Yes	Yes	Yes

A key challenge in multi-modal systems lies in the variability of data produced by each biometric method. Gait analysis, for instance, is influenced by a person's physical condition, footwear, and walking environment. Similarly, voice authentication can be affected by temporary health conditions, background noise, and environmental factors. Keystroke dynamics are sensitive to a user's cognitive state, stress, or fatigue, while mouse movements are impacted by elements like device type, hand posture, and screen resolution. These behavioral and environmental variations can lead to inconsistencies when combining data from different modalities, complicating accurate authentication. Developing algorithms that can adjust to and normalize these variations remains a major research gap.

Another challenge is the synchronization of data between different biometric modalities. Gait analysis and voice authentication require continuous monitoring, whereas keystroke dynamics and mouse movements are event-driven, activated only during specific user actions. Real-time integration of these varied data streams without introducing delays or errors is crucial. Ensuring seamless processing of data from diverse sources - despite differences in sampling rates or data formats - represents a significant technical obstacle. Research is necessary to create algorithms that can harmonize these data streams effectively.

The computational demands of processing multi-modal data also present a challenge. Each biometric modality - whether it's voice recognition, gait analysis, or keystroke dynamics - requires specific computational resources. For instance, voice authentication involves complex signal processing to analyze acoustic features, while gait analysis may require specialized sensors or motion capture systems. Combining these modalities increases the system's computational burden, and optimizing algorithms to function efficiently on devices with limited processing power is a key research area. Additionally, energy efficiency remains a concern for mobile or wearable devices, which rely on continuous data collection. Creating lightweight algorithms that optimize power usage is vital for practical, real-world implementation[5].

Privacy and security concerns are critical gaps in the development of multi-modal systems. These systems often require constant data collection, which can expose sensitive aspects of a user's daily life. For example, gait analysis could reveal personal movement patterns, while voice authentication captures vocal characteristics. Ensuring that such data is securely stored and processed, without infringing on user privacy, is paramount. Privacy-preserving techniques such as local data processing, anonymization, and encryption should be incorporated to address these concerns. Additionally, giving users control over their data, including the ability to opt in or out of certain biometric modalities, is essential for fostering trust.

Adaptability is another area that requires attention in multi-modal systems. User behavior naturally varies over time due to factors such as aging, health changes, or environmental differences. For example, a person's typing rhythm may fluctuate with stress or cognitive load, while their gait may change due to injury or aging. Multi-modal systems must be able to adapt to these shifts while maintaining high authentication accuracy. Developing adaptive algorithms that can track and adjust to these changes is crucial. Furthermore, these systems must differentiate between genuine behavioral changes and potentially fraudulent activities, necessitating more sophisticated models for continuous learning.

Scalability is another major gap in multi-modal biometric research. Much of the existing

research focuses on small-scale, controlled settings with limited user groups. For multi-modal systems to be viable in real-world applications, they must scale effectively to large, diverse populations, handling a variety of user behaviors, environmental conditions, and devices. Ensuring that these systems maintain high performance across different demographics, user habits, and usage scenarios is essential for their broader adoption. Moreover, managing the vast amounts of data produced by multiple biometric modalities requires robust data management strategies.

User experience and system usability are also key areas of research. While multi-modal systems enhance security, they may increase complexity, especially when multiple forms of interaction (such as typing, walking, or speaking) are involved. If the authentication process becomes intrusive or burdensome, it may deter users. Research is needed to develop systems that intelligently select the most appropriate biometric modality based on context and user preferences, offering seamless, intuitive experiences. Adaptive authentication mechanisms that require minimal user interaction while ensuring high security will improve usability.

Finally, cost and energy efficiency are significant barriers to the widespread adoption of multi-modal biometric systems. Many current systems require costly, specialized hardware and substantial computational resources. Additionally, continuous data collection and processing can drain battery life, particularly in mobile and wearable devices. Developing affordable, energy-efficient sensors and algorithms is crucial for making multi-modal systems practical for everyday use. Solutions that minimize the need for high-powered computational devices while preserving accuracy and security are essential for the wide-scale deployment of these systems.

Research Problem

The research problem centers around developing a multi-modal authentication system that effectively integrates gait analysis, keystroke dynamics, mouse movements, and voice authentication. The challenge lies in designing a system that can leverage these four behavioral biometrics, each capturing unique aspects of an individual's behavior, to

enhance security and authentication efficiency. The core problem is combining these diverse data streams into a unified system that provides continuous, reliable, and user-friendly authentication, while addressing the complexities associated with data variability, computational efficiency, and real-time processing.

Motivation for Multi-modal Integration

Authentication systems traditionally rely on single-modal biometrics, such as fingerprints or facial recognition, or behavioral methods like passwords. However, these systems have several vulnerabilities. Fingerprints can be spoofed, passwords are often weak or reused, and facial recognition systems can be tricked by photos or videos. Behavioral biometrics, on the other hand, are inherently harder to replicate. They capture how individuals perform actions, rather than what they are. This makes them much more resilient against attacks. Gait, keystroke dynamics, mouse movements, and voice patterns are unique to each individual and, when combined, can create a system that is far more secure and accurate than single-modal approaches.

Gait analysis, for example, uses walking patterns to identify an individual. These patterns are unique to a person's physical structure, movement style, and even their state of mind. Keystroke dynamics looks at the speed, pressure, and rhythm of typing, while mouse movements track the way users interact with a pointing device, providing insights into their behavioral habits. Voice authentication analyzes vocal features, such as pitch, cadence, and speech patterns, making it useful for hands-free authentication. Each of these modalities captures a different aspect of an individual's behavior, and integrating them can provide a comprehensive, multi-layered approach to user verification.

Why Integration is Important for Improving Security and Authentication Efficiency

The integration of gait analysis, keystroke dynamics, mouse movements, and voice authentication into a multi-modal system is crucial for improving both security and authentication efficiency. Each of these biometrics has its own strengths and weaknesses. For example, while gait is difficult to imitate and can provide continuous authentication during physical movement, it may be influenced by factors like footwear or injury.

Keystroke dynamics can be affected by cognitive load or stress, and mouse movements are susceptible to device-specific variations. Voice authentication, although useful for hands-free scenarios, is vulnerable to noise interference and temporary changes due to illness.

By integrating these four modalities, a multi-modal authentication system can take advantage of the complementary strengths of each. If one modality is impaired due to environmental factors or changes in behavior, the others can still provide valuable data for verification. For instance, if a user's voice is difficult to authenticate due to background noise, their gait analysis and keystroke dynamics can still confirm their identity. This redundancy reduces the likelihood of authentication failures, such as false negatives, and enhances the overall security of the system.

Enhancing Security

The primary advantage of multi-modal integration is the significant increase in security. Single-modality systems are often vulnerable to spoofing attacks. For example, an attacker might replicate a user's voice using advanced speech synthesis techniques or mimic their typing patterns through keylogging. However, replicating multiple behavioral modalities simultaneously—such as voice, gait, and keystroke dynamics—presents a far greater challenge. Multi-modal authentication systems require an attacker to bypass several layers of security, making it much more difficult to compromise the system.

Each biometric modality has unique characteristics that make it resistant to different forms of attacks. For example, while voice authentication might be vulnerable to voice imitation or background noise, gait analysis remains largely unaffected by these issues. Similarly, keystroke dynamics and mouse movements offer additional layers of authentication that are difficult to manipulate. By combining these diverse signals, the system can ensure a higher level of accuracy and security, providing a more robust defense against spoofing, fraud, and unauthorized access attempts.

Reducing False Positives and False Negatives

Another benefit of multi-modal systems is the reduction of false positives and false

negatives. False positives occur when the system incorrectly accepts an unauthorized user, while false negatives occur when the system fails to recognize an authorized user. Each biometric modality has its own error rates, and by combining data from multiple sources, the system can cross-check and validate authentication attempts. For example, if gait analysis or voice authentication is not entirely accurate due to environmental factors or health issues, keystroke dynamics and mouse movements can provide additional information to help make a correct decision.

The use of multiple biometric modalities allows the system to make more informed decisions, minimizing errors. When one modality encounters a failure, the others can help correct or compensate for it, leading to more accurate and reliable user authentication. This multi-layered approach ensures that authentication is both more secure and efficient, reducing the likelihood of erroneous denials or false acceptances.

Continuous and Passive Authentication

Multi-modal systems also enable continuous authentication, a significant advantage over traditional authentication methods that require user input only at specific times. In systems based on passwords or fingerprint recognition, users must actively authenticate themselves, which can be inconvenient or even risky if the authentication process is bypassed or forgotten. In contrast, behavioral biometrics can provide ongoing, passive authentication, continuously verifying the identity of the user while they interact with a device.

For example, gait analysis can continuously authenticate a user as they move, while keystroke dynamics and mouse movements can be monitored throughout their interaction with a computer. Voice authentication can function in the background, ensuring that the system always verifies the user's identity without requiring explicit input. This real-time, continuous authentication improves both security and convenience by providing persistent verification and preventing unauthorized access even if the user becomes distracted or leaves their device unattended.

Improving User Experience

User experience is another important aspect of multi-modal authentication. Traditional systems often rely on explicit user actions, such as typing a password or scanning a fingerprint. These actions can disrupt the user's workflow or create friction during the authentication process. In contrast, multi-modal systems enable passive authentication, allowing users to interact with their devices naturally without needing to engage in any extra steps.

By using multiple modalities, the system can adapt to different situations. For example, voice authentication might be preferred when the user is interacting with a voice-activated device, while keystroke dynamics and mouse movements may be more appropriate during computer use. This flexibility enables a smoother and more intuitive user experience while maintaining high security. Additionally, multi-modal systems can be tailored to a variety of user preferences, ensuring that the authentication process is both secure and user-friendly.

Robustness in Real-world Environments

Multi-modal systems also offer increased robustness in real-world environments. Many single-modality systems perform well in controlled, ideal conditions but struggle in more dynamic or unpredictable settings. For instance, voice authentication may fail in noisy environments, or a fingerprint scanner may have difficulty recognizing a print due to dirt or moisture. However, multi-modal systems are less affected by these factors, as they rely on multiple data points to authenticate a user.

If one modality is impaired—due to noise affecting voice authentication or injury affecting gait—the system can still rely on the other modalities to ensure accurate authentication. This makes multi-modal systems more reliable and adaptable, capable of functioning in a wide range of environments and under varying conditions. Their robustness ensures that users are authenticated reliably regardless of the context, making the system practical for diverse real-world applications.

By integrating gait analysis, keystroke dynamics, mouse movements, and voice authentication, multi-modal systems offer a powerful solution to the challenges of security, efficiency, and user experience. These systems not only enhance authentication accuracy but also provide greater flexibility and resilience in real-world scenarios. The ability to continuously authenticate users without interrupting their activities is a significant advantage over traditional methods, making multi-modal systems an important advancement in the field of secure authentication.

Objective

Main Objective

The primary objective of this project is to develop a multi-modal authentication system that integrates gait analysis, keystroke dynamics, mouse movements, and voice authentication into a unified, cohesive solution. This integrated system aims to provide continuous, secure, and efficient authentication by combining these four behavioral biometric traits, overcoming the challenges faced by traditional methods like passwords, PINs, and even standalone biometrics such as fingerprints or facial recognition. The project seeks to address issues related to security vulnerabilities, authentication errors, and user experience limitations.

As cyber-attacks and identity theft become more widespread, the weaknesses of traditional authentication methods have become increasingly apparent. Passwords can be guessed or stolen, PINs are often insecure, and even biometric systems like fingerprints or facial recognition can be spoofed. In contrast, behavioral biometrics offer a dynamic and hard-to-replicate form of authentication. By integrating multiple forms of behavioral biometrics into a single system, the security and reliability of user authentication can be significantly enhanced, providing a more robust solution than relying on a single modality.

The Integrated System Approach

The goal of this project is to design a multi-modal authentication system that works seamlessly across a variety of devices and environments. Unlike traditional systems that require explicit user actions like typing passwords or scanning fingerprints, this system will

continuously authenticate users based on real-time behaviors such as gait, typing rhythm, mouse movements, and voice patterns. These behaviors are inherently unique and challenging to replicate, enabling secure, passive, and continuous identification.

The system will integrate a combination of sensors and algorithms capable of analyzing these behavioral signals in real time, offering an intuitive and user-friendly experience. Unlike traditional systems that require active engagement, the integrated multi-modal system will passively authenticate the user as they interact with their device, walk, or speak. This seamless process minimizes friction for users while maintaining a high level of security.

Importance of Each Component in Enhancing Security

Integrating gait analysis, keystroke dynamics, mouse movements, and voice authentication creates a holistic security system by combining the strengths of each modality. By relying on multiple sources of behavioral data, the system becomes more resilient to attacks and better at verifying user identity in real-world scenarios. Below is a breakdown of how each component contributes to enhancing security:

Gait Analysis

Gait analysis provides a passive method of continuous authentication by monitoring walking patterns. Factors such as physical characteristics, posture, and movement style make each individual's gait unique. This continuous, passive authentication works as users move through a space or interact with a device.

Unlike static biometrics like fingerprints or facial recognition, gait is extremely difficult to replicate because it involves a complex combination of fine motor skills and body mechanics. Subtle differences in walking style, posture, and pressure make gait nearly impossible to imitate. By incorporating gait analysis into the authentication process, the system gains a highly reliable biometric that works passively as users walk, boosting both security and user convenience[1][6].

Keystroke Dynamics

Keystroke dynamics tracks how users type, capturing factors like typing speed, key press duration, and the time between keystrokes. Each person types in a unique way, influenced by physical aspects (such as hand size) and cognitive factors (like typing habits). This pattern can be continuously monitored during user interaction with the system.

Combining keystroke dynamics with other biometric modalities strengthens the overall authentication security. For instance, if voice recognition fails due to noise or illness, keystroke dynamics can still verify identity based on typing rhythm and speed. Additionally, even if a password is stolen, replicating the user's typing pattern would still be necessary to authenticate successfully, adding an extra layer of security[2].

Mouse Movements

Mouse dynamics provide another layer of behavioral authentication. The way a person moves the mouse, clicks, and scrolls is shaped by hand-eye coordination and cognitive style, making it unique. The continuous interaction with a device allows for passive monitoring and authentication during user sessions. Analyzing the speed and direction of mouse movements as users interact with webpages or documents provides valuable identifiers for user authentication.

Integrating mouse dynamics with other biometric signals enables continuous user verification during computer usage. If one modality (like voice or gait) is compromised, mouse dynamics can still ensure authentication accuracy, offering a reliable fallback method[7].

Voice Authentication

Voice authentication is crucial for providing hands-free, continuous verification. Voice patterns, such as pitch, tone, cadence, and speech rhythm, are unique to each individual. These vocal traits are generally stable over time, making voice authentication a powerful tool for identity verification.

While voice recognition can be affected by factors like background noise or changes in the user's voice due to illness, combining it with other modalities reduces the likelihood of authentication failure. Voice can be used to authenticate users passively during speech or

interaction with voice-activated devices. If voice input is inconsistent, other modalities like keystroke dynamics or gait analysis can still provide alternative verification methods, ensuring a reliable authentication process[4].

Enhancing Overall Security with Multi-modal Integration

By combining gait analysis, keystroke dynamics, mouse movements, and voice authentication, the system gains several advantages:

- **Redundancy and Resilience:** Multiple biometrics make it harder for attackers to spoof or bypass the system. To successfully breach the system, an attacker would need to replicate several distinct behavioral patterns, a nearly impossible task. Even if an attacker mimics a user's voice or typing, duplicating their walking pattern or mouse movements is far more challenging.
- **Continuous Authentication:** Unlike traditional authentication methods that authenticate users at a single point in time, multi-modal systems offer ongoing monitoring of user behavior. This reduces the chance of unauthorized access even if a device is left unattended for a brief period.
- **Error Reduction:** Single-modality systems often suffer from false positives or false negatives due to environmental factors or device malfunctions. By integrating multiple modalities, the system cross-checks data from different sources, improving overall accuracy and reducing errors.
- **Adaptability:** The system can dynamically select the most appropriate authentication method based on the user's context. For example, voice recognition may be preferred in voice-activated environments, while gait or keystroke dynamics may be more suitable in desktop or mobile environments. This adaptability improves both security and user experience.

- **Scalability:** The system can scale across various devices, whether mobile, desktop, or wearable, making it versatile and suitable for a wide range of applications.

By integrating these biometric components, the system not only ensures secure authentication but also enhances user experience through seamless, continuous, and passive verification. This project aims to develop a comprehensive solution to modern authentication challenges.

Specific Objectives

The primary goal of this project is to develop a comprehensive multi-modal authentication system that integrates four behavioral biometrics: gait analysis, keystroke dynamics, mouse movements, and voice authentication. Achieving this goal requires several key tasks: collecting and preprocessing data, developing machine learning models, and integrating the system for thorough evaluation. Each task is outlined below with its individual components.

Data Collection and Preprocessing

The initial step in creating the multi-modal authentication system is to gather and preprocess data for the four biometric modalities—gait, keystroke dynamics, mouse movements, and voice. Since publicly available datasets will be used, the first objective is to identify appropriate datasets that provide reliable and comprehensive data for each biometric modality. These datasets should encompass a variety of user behaviors, environmental conditions, and device setups to ensure that the system is adaptable across real-world scenarios.

Data Collection for Gait Analysis: For gait analysis, publicly available datasets, such as the CASIA Gait Database or the OU-ISIR Gait Database, will be utilized. These datasets contain video recordings of individuals walking under various conditions and from different angles. The main task is to extract gait features, such as walking speed, stride length, and posture, using computer vision techniques from these video frames.

Data Collection for Keystroke Dynamics: Keystroke dynamics data will be sourced from publicly available datasets like the Keystroke Dynamics Database or the Turing Dataset. These datasets capture users typing passwords, sentences, or random strings on various devices and keyboards. The goal is to extract typing patterns, such as time between key presses, duration of key presses, and overall typing speed.

Data Collection for Mouse Movements: For mouse movements, datasets like the Mouse Dynamics Dataset will be used. This dataset captures information on how users move the mouse, including speed, click frequency, and cursor trajectories. The data will be analyzed to understand how individual interaction behaviors differ between users.

Data Collection for Voice Authentication: Voice authentication data will come from datasets like VoxCeleb or CommonVoice, which provide audio samples of individuals speaking various sentences or phrases. These datasets typically include both clean and noisy audio samples from diverse speakers, allowing the system to be trained to perform well in different environments. The focus will be on extracting vocal features like pitch, tone, cadence, and rhythm.

Preprocessing: Once data is collected, preprocessing is required to prepare the data for machine learning. For gait data, features like stride length and walking speed will be extracted from the raw video data. For keystroke dynamics, raw data will be transformed into features like keystroke intervals and key press duration. Mouse movement data will be filtered and smoothed to remove noise, ensuring accurate feature extraction. Voice data preprocessing will involve noise reduction, feature extraction (e.g., Mel Frequency Cepstral Coefficients), and segmentation into fixed-length windows for easier analysis.

Model Development

After data collection and preprocessing, the next task is to build machine learning models for each biometric modality. Different neural network architectures will be used, each tailored to process the specific type of data from each modality.

Gait Analysis Model (CNN): For gait analysis, a Convolutional Neural Network (CNN) will be used to process video frames or sensor data. CNNs are effective for image or video-based tasks, as they can automatically extract spatial features from raw pixel data. The model will learn to recognize unique gait patterns by analyzing both spatial and temporal features from video or sensor data. The CNN architecture will include multiple convolutional layers, pooling layers for dimensionality reduction, and fully connected layers for final classification[9].

Keystroke Dynamics Model (RNN): Keystroke dynamics, being sequential in nature, will be modeled using a Recurrent Neural Network (RNN), specifically Long Short-Term Memory (LSTM) units. LSTMs are suitable for sequence prediction tasks like analyzing typing rhythm and speed, as they can capture long-term dependencies between consecutive keystrokes. This model will be trained to identify users based on their unique typing patterns[10].

Mouse Movements Model (RNN): Similar to keystroke dynamics, mouse movements involve time-series data, making an RNN (LSTM) ideal for modeling this modality as well. The model will process sequences of mouse movements, including speed, direction, and click patterns, to classify users based on their interaction styles. LSTM networks will be used to capture the temporal dependencies between mouse movements and clicks[3].

Voice Authentication Model (CNN): For voice authentication, a CNN will be used to process spectrograms or Mel-frequency cepstral coefficients (MFCCs) extracted from raw audio samples. CNNs are well-suited for speech recognition tasks as they can capture hierarchical features from the audio signal. The model will learn to classify voices based on distinct vocal features, including pitch, tone, and cadence. It will be trained on both clean and noisy voice samples to ensure robustness in varied environments[11].

System Integration and Evaluation

After developing individual models for each biometric modality, the next step is to integrate them into a cohesive multi-modal authentication system. This process involves

combining the outputs of each modality into a unified decision-making mechanism.

The multi-modal system will employ a fusion strategy to combine the outputs of the individual models. The two main fusion approaches are early fusion, which combines raw data before processing, and late fusion, which processes each modality through its own model and combines the predictions for final decision-making. In this project, late fusion is preferred, as it allows each modality to be handled separately before the results are merged.

The final system will be evaluated using performance metrics like accuracy, precision, recall, and F1-score to assess the effectiveness of the multi-modal authentication system. Additionally, robustness tests will include introducing noise or simulated spoofing attempts, such as mimicking voices or manipulating gait. The system will be tested in a real-world scenario with a diverse set of users under varying environmental conditions. A cross-validation strategy will be used during training to ensure generalization to unseen data, and the fusion of models will be thoroughly evaluated for its impact on system accuracy and reliability.

By completing these tasks - data collection, model development, system integration, and evaluation - the multi-modal authentication system will be developed and tested for its ability to deliver secure, efficient, and user-friendly authentication.

Methodology

Methodology Overview of the Multi-Modal Authentication System

Authentication systems are fundamental to modern security protocols, ensuring that only authorized users can access sensitive data or systems. Traditional authentication methods such as passwords or PINs are becoming increasingly inadequate due to their inherent vulnerabilities. To address these weaknesses, biometric systems have emerged as a more secure and reliable alternative. Specifically, behavioral biometrics—such as Gait Analysis, Keystroke Dynamics, Mouse Movements, and Voice Authentication—leverage unique patterns inherent to an individual’s behavior, making them significantly harder to replicate or steal compared to traditional methods.

While single-modal biometric systems (those relying on a single biometric modality) have improved security over traditional systems, they still face several challenges. These include susceptibility to spoofing, environmental influences, and variability in user behavior over time. To overcome these challenges, the multi-modal authentication system integrates multiple biometric modalities, enhancing accuracy, robustness, and reliability.

The core methodology of this multi-modal system is centered around the integration of four behavioral biometric components: Gait Analysis, Keystroke Dynamics, Mouse Movements, and Voice Authentication. Each of these components is processed individually, and the resulting similarity scores are then combined using a method known as weighted fusion. This ensures that the system accounts for the strengths of each modality while minimizing the weaknesses of individual components.

Key Steps in Methodology

1. Data Collection and Preprocessing

Each biometric modality (Gait, Keystroke, Mouse, and Voice) requires different forms of input data: video frames for Gait, typing patterns for Keystroke, mouse movements for Mouse, and voice recordings for Voice. The data is collected either through sensor data or audio/visual recordings, and preprocessing steps ensure the data is standardized, noise-free,

and ready for further analysis.

- **Gait Analysis:** The data is typically collected through video recordings of individuals walking in a controlled environment. These video frames are processed to create Gait Energy Images (GEI), which represent the walking cycle of an individual.
- **Keystroke Dynamics:** Users type predefined sentences or random strings, and data such as key press duration, inter-key interval, and dwell time are recorded.
- **Mouse Movements:** Data is collected by tracking the x and y coordinates of the mouse cursor, along with timestamps for each movement, allowing for the calculation of mouse speed and acceleration.
- **Voice Authentication:** Voice data is captured via audio recordings, and features such as pitch, cadence, and speech rate are extracted from the recordings.

Once the data is collected, preprocessing ensures consistency. For Gait, images are resized to a consistent size; for Keystroke, Mouse, and Voice, the data is normalized and transformed into a format suitable for analysis by the models.

2. Individual Modality Processing

Each modality undergoes specialized deep learning model processing to extract meaningful features.

Gait Analysis: Convolutional Neural Networks (CNNs) are used to process Gait Energy Images (GEI) and extract spatial features that represent the unique walking patterns of an individual. CNNs are well-suited for this task as they automatically detect and learn complex spatial patterns from the images.

Keystroke Dynamics: Long Short-Term Memory (LSTM) networks, a type of Recurrent

Neural Network (RNN), are used to capture the temporal dependencies between keystrokes. LSTMs model the sequential nature of typing behavior, learning individual typing rhythms and patterns based on key press timings.

Mouse Movements: For Mouse Movements, the system exclusively uses Convolutional Neural Networks (CNNs) to process data. The system analyzes features such as speed, path efficiency, and acceleration through CNNs. Unlike traditional approaches that may use a mix of RNNs and CNNs, here we focus solely on CNNs for analyzing mouse trajectories, allowing the model to capture spatial features in mouse movement patterns, such as movement direction and velocity.

Voice Authentication: Voice data is converted into spectrograms (visual representations of the frequency content over time) using techniques such as Short-Time Fourier Transform (STFT) or Mel-Frequency Cepstral Coefficients (MFCC). CNNs are then applied to process these spectrograms, extracting features like pitch, cadence, and rhythm that are unique to each speaker's voice.

3. Similarity Measurement via Cosine Similarity

After processing the data for each modality, the system calculates the similarity between the test data (input from the user attempting to authenticate) and the stored reference data (templates of previously enrolled users). This is done using Cosine Similarity, which calculates the cosine of the angle between two vectors.

Cosine Similarity formula:

$$\text{Cosine Similarity} = \frac{A \cdot B}{\|A\| \|B\|}$$

Figure 1 - Cosine Similarity

Where:

- A and B are the feature vectors for the test input and the enrolled template,

respectively.

The cosine similarity value ranges from 0 to 1, with 1 indicating that the test input and the reference data are identical, and 0 indicating no similarity. Each modality will have a separate similarity score based on how well the test input matches the enrolled template for that modality[12].

4. Weighted Fusion of Individual Scores

Once each modality provides a similarity score (between 0 and 1), the final decision is made through a weighted fusion process. This technique combines the scores from each modality while taking into account the relative importance or reliability of each modality. For instance:

- **Gait Analysis** might be assigned a higher weight if it is found to be more reliable in certain contexts.
- **Keystroke Dynamics** may receive a lower weight if typing behavior tends to fluctuate or is less predictable for certain users.

The weighted fusion score is calculated as follows:

$$\text{Fused Similarity Score} = W_1 \cdot S_{\text{Gait}} + W_2 \cdot S_{\text{Keystroke}} + W_3 \cdot S_{\text{Mouse}} + W_4 \cdot S_{\text{Voice}}$$

Where:

- W_1, W_2, W_3, W_4 are the weights assigned to each modality.
- S_{Modality} represents the similarity score from each individual modality.

This weighted fusion ensures that the final score reflects the strengths of each modality, incorporating each one's reliability into the decision-making process[13].

5. Final Authentication Decision

The fused similarity score is compared to a threshold to make the final authentication decision. If the fused similarity score exceeds the threshold, the user is authenticated. If it falls below the threshold, access is denied.

The threshold value is set to balance two important factors:

- **False Positives:** Incorrectly authenticating an unauthorized user.
- **False Negatives:** Failing to authenticate a legitimate user.

By adjusting the threshold, the system can optimize between security (minimizing false positives) and usability (minimizing false negatives). The threshold can be tailored based on the specific security needs of the system, providing a flexible and customizable authentication solution.

Detailed Breakdown of the Methodology

1. Data Collection and Preprocessing:

- **Gait Analysis:** Data is collected through video recordings, processed into Gait Energy Images (GEI).
- **Keystroke Dynamics:** Users type predefined sentences or random strings, capturing key press durations and intervals.
- **Mouse Movements:** Data is collected through mouse tracking, capturing coordinates, speed, and acceleration.
- **Voice Authentication:** Voice data is recorded, and key features like pitch, cadence, and speech rate are extracted.

Preprocessing involves normalizing and transforming the data into consistent formats, ensuring it is ready for analysis by the respective models.

2. Individual Modality Processing

- **Gait Analysis:** CNNs process the GEI images to extract spatial features and learn walking patterns.
- **Keystroke Dynamics:** LSTMs process sequential keystroke data to capture typing rhythms and keypress timing patterns.
- **Mouse Movements:** CNNs analyze mouse movements, capturing spatial features such as direction, speed, and efficiency.
- **Voice Authentication:** Spectrograms are generated from voice recordings, and CNNs are used to extract pitch, cadence, and speech patterns.

3. Similarity Measurement

Cosine Similarity is computed for each modality to determine how closely the test input matches the enrolled template.

4. Weighted Fusion

The individual similarity scores are combined using a weighted fusion method, where each modality contributes based on its reliability and relevance for the specific user or scenario.

5. Final Authentication Decision

The fused similarity score is compared against a threshold to determine if the user should be authenticated. This threshold is adjusted based on the balance between false positives and false negatives, optimizing security and usability.

By integrating multiple biometric modalities, this multi-modal authentication system increases security, enhances robustness, and provides a seamless authentication experience for users.

Gait Analysis Methodology

Gait analysis refers to the process of identifying individuals based on the unique characteristics of their walking patterns. Unlike other biometric systems, such as fingerprint or facial recognition, which often require close contact with the individual, gait can be captured from a distance without any physical interaction. This ability to identify individuals based on how they walk offers significant advantages in a variety of applications, such as surveillance, access control, and security. Gait analysis is also less intrusive than other biometric systems and can work in real-time and across a variety of conditions.

In this methodology, we explore video-based gait recognition using Convolutional Neural Networks (CNNs), which have proven to be highly effective for image and video classification tasks. The Gait Energy Image (GEI) is a representation of gait patterns, derived from video data, and the CNN model is trained to extract spatial features from these GEI frames. The system works by comparing the features of a test image with those of an enrolled user's gait, calculating the similarity using Cosine Similarity to determine if the walking patterns match[15].

Data Collection for Gait Analysis

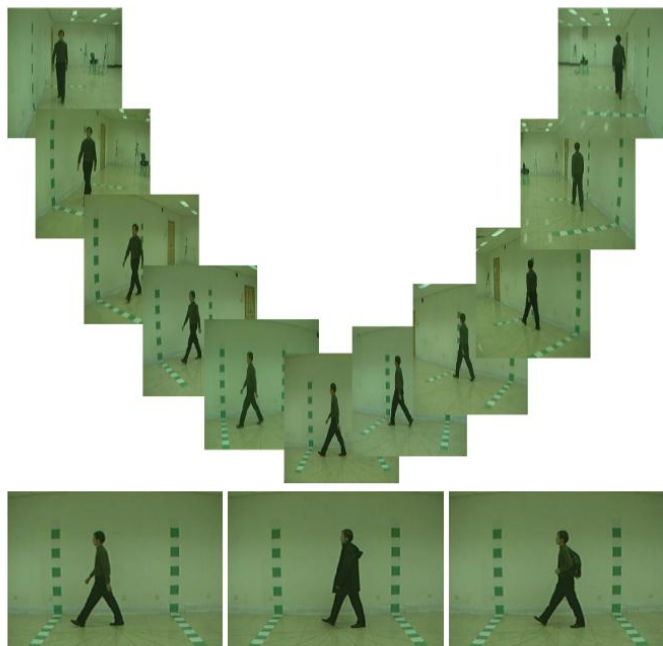


Figure 2- Gait Dataset

The foundation of gait analysis is data collection, typically obtained through video frames of individuals walking in various environments and under different conditions. For effective model training, the dataset should cover a broad spectrum of walking behaviors, including variations in clothing, background, walking speed, and environmental factors. The following publicly available datasets are commonly used for gait analysis:

1. **CASIA Gait Database:** A well-known publicly available dataset that contains video recordings of multiple subjects walking under varying conditions, including changes in clothing, background, and walking speed. This dataset provides comprehensive video data for gait recognition tasks.
2. **GEI (Gait Energy Image) Dataset:** This dataset extracts energy images from video frames to represent an individual's gait. The GEI captures the motion characteristics of walking by averaging multiple video frames, allowing the representation of the full gait cycle.

Data collection involves capturing sequences of video frames while users walk, which are then processed into images. The Gait Energy Image (GEI) is generated by averaging several frames, which represent a complete gait cycle. This representation is essential because it captures the periodic nature of walking, which is key to recognizing an individual's gait. In this process, positive pairs consist of images of the same person walking under different conditions, such as different clothing or lighting. Negative pairs, on the other hand, are images from different individuals. These paired images are crucial for training the CNN to distinguish between similar and dissimilar gaits.

Model Architecture

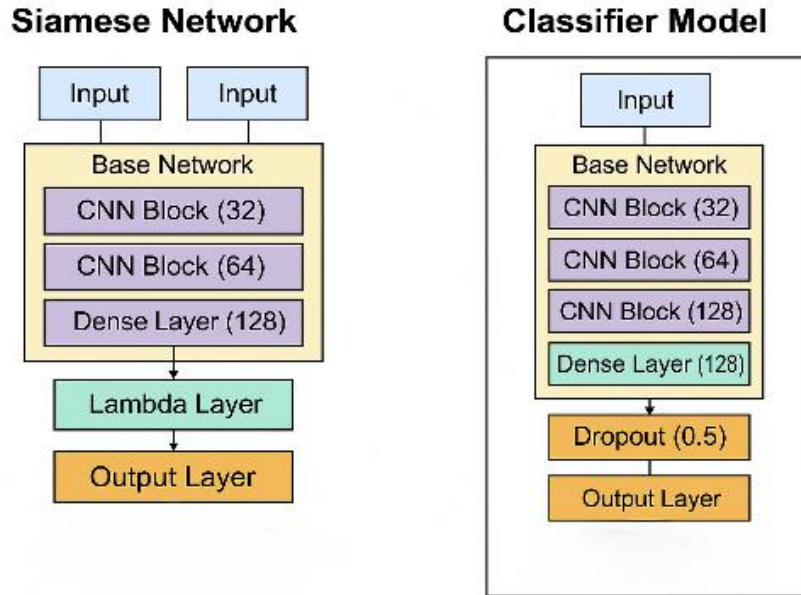


Figure 3- Gait Model Architecture

The CNN architecture used in gait analysis is designed to handle image-like data and automatically extract spatial features that are relevant to gait recognition. CNNs excel at recognizing patterns and features in images, which makes them particularly well-suited for this task. Below is an expanded breakdown of the CNN architecture for gait analysis:

1. **Input Layer:** The input to the model consists of images or frames of individuals walking. These images are resized to a fixed size (e.g., 128x128 pixels) to ensure uniformity across the dataset. Image normalization is performed by scaling pixel values to a range of 0 to 1, which helps the model converge more efficiently during training.
2. **Convolutional Layers:** The convolutional layers are responsible for applying a series of filters to the input images. Each filter detects specific spatial features, such as edges, textures, or body contours. In the context of gait analysis, these filters focus on detecting features that represent the movement of the body, stride length, and swinging of the arms. The convolutional layers progressively learn more complex features, beginning with low-level features like shapes and edges in the

earlier layers, and gradually moving to more complex features such as posture and stride patterns in deeper layers.

3. **Residual Blocks:** Residual blocks are incorporated to improve the feature learning process and address the vanishing gradient problem. These blocks introduce skip connections, which allow the network to bypass certain layers, helping to retain important information and enabling the model to learn deeper and more complex features without losing crucial data. This makes the model more effective at capturing intricate gait patterns.
4. **Pooling Layers:** Pooling layers, such as MaxPooling or AveragePooling, are used to reduce the spatial dimensions of the feature maps generated by the convolutional layers. This downsampling helps focus on the most important features while reducing computational complexity. Pooling layers ensure that the model can focus on the higher-level abstract features that are critical for recognizing gait patterns, such as overall body posture and walking speed.
5. **Fully Connected Layers:** After the convolutional and pooling layers, the output is passed through fully connected layers. These layers aggregate the features learned by CNN and use them to make a final decision about whether the gait pattern matches the enrolled template. The fully connected layers interpret the spatial features extracted by the convolutional layers and convert them into a decision-making process.
6. **Output Layer:** The final output layer produces a binary classification score, which is generated by a sigmoid activation function. The sigmoid function outputs a value between 0 and 1, where a score close to 1 indicates that the gait pattern matches the enrolled user's, and a score close to 0 indicates a mismatch. This binary classification helps the system make the authentication decision.

Preprocessing and Feature Extraction

Preprocessing plays a crucial role in preparing the gait data for input into the CNN model. Several preprocessing steps are involved in ensuring the data is ready for training:

1. **Image Resizing:** Each video frame is resized to a fixed size, such as 128x128 pixels, to ensure consistency and uniformity across the dataset. Standardizing the input size ensures that all images are processed in the same manner, making it easier for the model to learn useful patterns.
2. **Normalization:** Pixel values are normalized by dividing by 255, ensuring that all values fall between 0 and 1. Normalization helps the model converge faster during training and prevents the model from being biased by the range of initial pixel values.
3. **Pair Creation:** Positive pairs consist of images of the same person walking under different conditions, such as different clothing, backgrounds, or walking speeds. Negative pairs are formed by selecting images from different individuals. This process is essential for training the CNN to differentiate between gait patterns from different people and recognize variations in gait under different conditions.

For sensor-based data, additional preprocessing steps are required:

- **Noise Filtering:** Raw sensor data often contains noise and unwanted fluctuations, which can interfere with the analysis. Noise filtering techniques, such as low-pass filters, are used to remove these disturbances.
- **Normalization:** Sensor data is normalized to ensure consistency across different walking conditions and subjects. This ensures that the model treats all data points in a consistent manner, making the system more robust.
- **Feature Extraction:** Key features such as stride length, walking speed, and body symmetry are extracted from the sensor data. These features provide valuable information that CNN can use to learn unique gait patterns.

Evaluation and Performance Metrics

The effectiveness of the gait analysis system is evaluated using several key performance metrics. These include:

- **Accuracy:** The proportion of correct gait matches between the test images and the enrolled user's gait features.

- **Precision and Recall:** Precision measures the proportion of correctly identified gait matches, while recall measures the ability to detect gait matches across the dataset. These metrics help evaluate how well the system performs in distinguishing between true matches and false positives or negatives.
- **F1-Score:** The F1-score provides a balance between precision and recall, ensuring that both false positives and false negatives are minimized.

The model is tested under different conditions to assess robustness, such as adding noise to the video data or simulating different environmental conditions to see how well the model performs in real-world scenarios.

Keystroke Dynamics Methodology

Keystroke dynamics is a behavioral biometric modality that leverages the unique timing and rhythm of an individual's typing behavior to authenticate their identity. Unlike traditional password-based authentication systems, which only verify the knowledge of a password, keystroke dynamics continuously authenticates a user based on their typing rhythm. This approach is highly applicable as it requires no additional hardware, aside from the standard keyboard, and works seamlessly in the background without interrupting the user's activities. Each individual has a unique typing style, characterized by factors such as typing speed, rhythm, and the intervals between key presses, which can be used to differentiate users and authenticate their identity.

In this methodology, we utilize Long Short-Term Memory (LSTM) networks, a type of Recurrent Neural Network (RNN), to model the sequential nature of keystroke dynamics. LSTM networks are well-suited for this task because they capture the temporal dependencies between consecutive key presses, allowing the model to learn the user's unique typing behavior over time.

Data Collection for Keystroke Dynamics

Keystroke dynamics data is typically collected using publicly available datasets like the Turing Dataset and the Keystroke Dynamics Database. These datasets contain recordings of multiple users typing specific sentences, passwords, or randomly generated strings. The data collection process involves recording key press durations, the time intervals between

consecutive key presses, and other typing characteristics. The following features are recorded:

- **Key press duration:** The amount of time the user holds down a key.
- **Inter-key interval (IKI):** The time between releasing one key and pressing the next key.
- **Dwell time:** The time spent pressing a key before releasing it.
- **Typing speed:** The rate at which a user types, typically measured in words per minute (WPM).
- **Error Rate:** The error rate captures the proportion of incorrect keystrokes made by a user during typing
- **Inter-Key Interval (IKI)** measures the time between the release of one key and the press of the next
- **Error Correction per Character (ECPC):** ECPC quantifies how often a user needs to correct an error by measuring the number of backspaces or delete actions performed per character typed.
- **Keystrokes per Character (KSPC):** KSPC represents the number of keystrokes required to type each character, accounting for both correct and erroneous keystrokes
- **Rollover Ratio (ROR):** The rollover ratio measures the number of simultaneous keystrokes pressed by the user

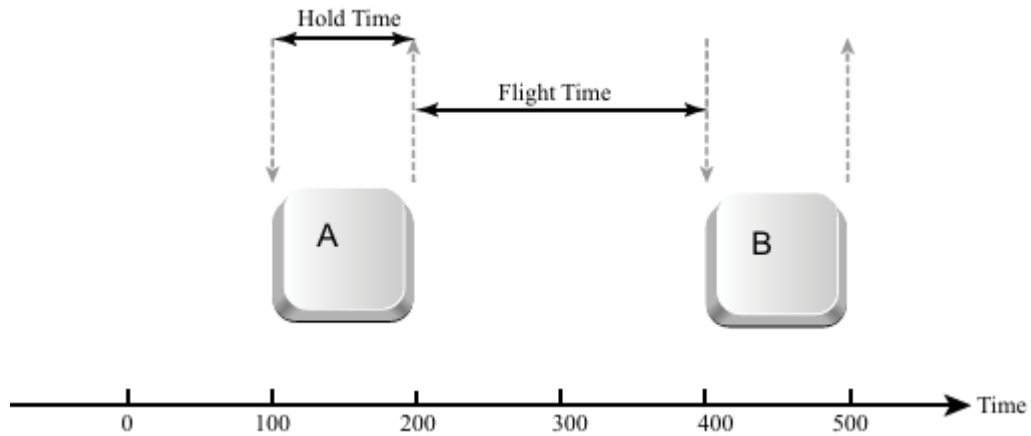


Figure 4- Flight & Dwell Time

These features provide insight into an individual's typing rhythm and behavior, which can be used to model the user's unique typing pattern. The data is split into training and test datasets, with positive pairs consisting of keystroke sequences from the same user typing the same text, and negative pairs consisting of sequences from different users typing the same or different text.

Model Architecture

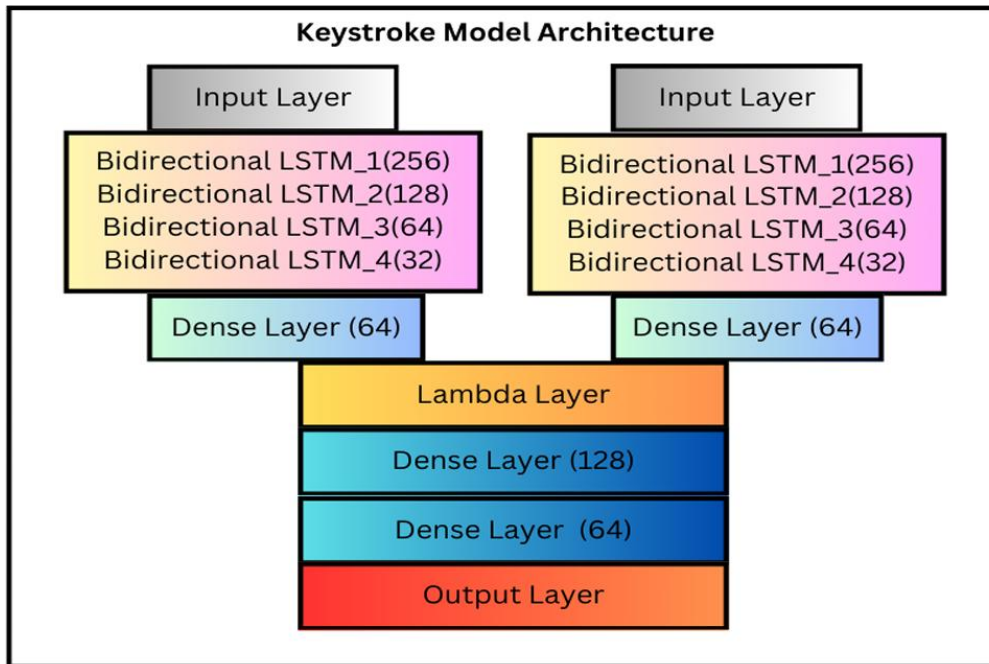


Figure 5 - Keystroke Model Architecture

The model for keystroke dynamics relies on LSTM networks to process sequential data. Below is a detailed breakdown of the LSTM architecture:

1. **Input Layer:** The input to the LSTM network consists of sequences of keystroke timing features (such as key press duration, inter-key interval, and dwell time). Each keystroke is represented as a vector containing these features. The data is time-ordered, meaning each sequence represents the user's typing behavior over time, such as typing a specific sentence.
2. **LSTM Layers:** LSTM layers process the sequences of keystroke data and learn the temporal relationships between consecutive key presses. By capturing how the timing of one key press influences the timing of the next, the LSTM network can model the user's typing rhythm and behavior.
3. **Bidirectional LSTM Layers:** These layers process the keystroke sequence in both directions, enabling the network to capture context from both preceding and succeeding keystrokes. This bidirectional approach enhances the model's ability to understand the full context of typing behavior.

4. **Fully Connected Layers:** After the LSTM layers, the output is passed through fully connected layers to interpret the temporal dependencies and refine the classification task.
5. **Output Layer:** The output layer produces a binary classification score using a sigmoid activation function. The output score indicates whether the test keystroke sequence matches the enrolled user's typing pattern.

Preprocessing and Feature Extraction

Preprocessing steps are essential for converting raw keystroke logs into usable input for the LSTM network:

1. **Normalization:** Keystroke timing features such as key press duration and inter-key interval are normalized to ensure consistency across different users. This step ensures that each feature contributes equally to the model's learning process.
2. **Feature Extraction:** Key features such as dwell time, typing speed, and rhythm are extracted from raw keystroke data. These features serve as the primary components of typing behavior and are used to distinguish between individual typing patterns.
3. **Sequence Segmentation:** Keystroke data is segmented into smaller, fixed-length sequences for easier processing by the LSTM network. This segmentation enables the model to focus on short-term dependencies within typing sessions.
4. **Sequence Padding:** Shorter sequences are padded to a uniform length to ensure that all input sequences have the same number of time steps, allowing the LSTM to process them in batches.

Learning the Typing Patterns

Once preprocessing is complete, the LSTM network begins learning the user's typing rhythm, including factors like typing speed, dwell time, and rhythm. The LSTM layers process the sequential keystroke data, learning to recognize unique typing behaviors over time.

Evaluation and Performance Metrics

The performance of the keystroke dynamics model is evaluated using metrics such as accuracy, precision, recall, and F1-score. These metrics help assess how well the model distinguishes between correct and incorrect typing patterns. Additionally, a confusion matrix can be used to visualize the model's performance and identify errors such as false positives and false negatives.

Mouse Movement Methodology

Mouse movement tracking has become an increasingly prominent behavioral biometric modality used for user authentication. Unlike traditional methods, such as passwords or PINs, which require explicit user input, mouse movement analysis provides a passive, continuous, and non-intrusive method of authenticating users. This method analyzes unique behavioral patterns exhibited during mouse interaction, such as the speed, trajectory, and direction of movement. By leveraging these unique patterns, it becomes possible to identify and authenticate users in a manner that is both seamless and effective. Mouse movement authentication systems collect various features related to how an individual interacts with the mouse. These include:

- **Mouse Speed:** The rate at which the mouse moves across the screen.
- **Path Efficiency:** The efficiency of the mouse movement between two points on the screen.
- **Jerk:** The rate of change in the mouse's speed.
- **Acceleration:** The rate at which the mouse speed changes.
- **Direction:** The angle at which the mouse moves.
- **Distance Moved:** The total physical distance the mouse covers between points on the screen.

By capturing these behavioral features, the system can create a profile of each user's unique mouse movement patterns. When combined with other biometrics such as gait analysis or keystroke dynamics, mouse movement analysis provides a robust and continuous authentication solution. This methodology will delve into the processes of data collection, model architecture, preprocessing, feature extraction, and learning processes used in mouse movement authentication.

Data Collection for Mouse Movements

The first step in developing a mouse movement authentication system is gathering data on the user's interactions with the mouse. To do this, the system records the mouse's x and y coordinates over time along with other dynamic features like speed, jerk, and acceleration. The data collection process involves engaging the user in specific tasks that are designed to capture the unique aspects of their mouse usage. Typical tasks for data collection include:

- **Point-and-click tasks:** Users click on predefined targets on the screen, which helps capture basic mouse movement features.
- **Drag-and-drop tasks:** Users move objects across the screen, providing a more complex set of movement data.
- **Freehand movement:** Users freely move the mouse across the screen without specific targets, which captures more natural, unstructured movement patterns.

For each of these tasks, the dataset collects the following mouse movement features:

- **X and Y coordinates:** These represent the position of the mouse cursor on the screen at any given time.
- **Time Stamps:** The exact time when the mouse moves between two points, which

enables the calculation of mouse speed and acceleration.

- **Mouse Speed:** The speed at which the mouse moves from one position to another, calculated as the distance moved divided by the time taken.
- **Path Efficiency:** The ratio of the straight-line distance between two points to the actual path taken by the mouse.
- **Jerk:** A measure of how rapidly the speed of the mouse changes during movement. This helps identify sudden, erratic movements.
- **Acceleration:** The rate of change in mouse speed, which indicates how fast the mouse is accelerating or decelerating.

Each user performs multiple trials to generate sufficient data that reflects their unique behavior. This data is then split into training and testing datasets, ensuring that the model can generalize well to unseen data.

Model Architecture

For mouse movement authentication, a Recurrent Neural Network (RNN) is ideal, particularly the Long Short-Term Memory (LSTM) variant. This is because mouse movements, much like keystroke dynamics, are sequential in nature. Each movement depends on the previous ones, and LSTMs are excellent at learning temporal dependencies in sequential data.

The architecture of the model is designed to capture and learn patterns from sequential mouse movement data. The model consists of the following key components:

1. **Input Layer:** The input to the model consists of sequences of mouse movement data. Each sequence includes:
 - X and Y coordinates at each timestamp.

- Features such as mouse speed, path efficiency, jerk, and acceleration over the time window. The data is organized as a time series, with each data point representing the state of the mouse at a specific moment.
2. **LSTM Layers:** The LSTM layers are responsible for learning the temporal relationships between consecutive mouse movements. These layers help the model understand how one movement influences the next. The LSTM network processes the sequential data and captures patterns in how the mouse is moved and accelerated over time. By retaining past movement events, LSTM units predict future behavior, enabling the model to learn each user's unique interaction style.
 3. **Bidirectional LSTM Layers:** In some models, Bidirectional LSTMs are used, which process the data in both forward and backward directions. This method allows the model to capture dependencies from both past and future movements. For example, the mouse speed at time t could be influenced not only by previous movements but also by the movements that will follow. Bidirectional processing improves the model's accuracy by understanding the full context of the mouse behavior.
 4. **Fully Connected Layers:** After processing the sequential data through LSTM layers, the output is passed through fully connected layers. These layers aggregate the temporal features learned by the LSTMs and refine the model's understanding of the user's mouse movement behavior. The fully connected layers transform the temporal features into a final classification score.
 5. **Output Layer:** The output of the model is a binary classification score, produced by a sigmoid activation function. This score represents the similarity between the test movement pattern and the enrolled user's mouse movement template. A score close to 1 indicates a strong match, while a score closer to 0 indicates a mismatch.

Preprocessing and Feature Extraction:

Before the data is fed into the LSTM network, preprocessing steps are necessary to prepare the raw data for model training. These preprocessing steps ensure that the data is clean, normalized, and consistent across all users:

1. **Normalization:** Mouse movement features, such as speed and acceleration, can vary significantly between users. To address this, features are normalized to a common scale. For instance, mouse speed can be normalized by dividing the observed speed by the maximum speed recorded in the dataset, ensuring consistency across users.
2. **Feature Extraction:** Important mouse movement features such as speed, jerk, acceleration, and path efficiency are extracted from the raw x and y coordinates. These features capture the unique differences in how users interact with the mouse, such as the smoothness or erraticness of movement, and how quickly they move the cursor across the screen.
3. **Sequence Segmentation:** Raw mouse movement data is divided into smaller windows or segments. Each segment represents a sequence of movements that can be processed by the LSTM model. This segmentation allows the network to focus on local patterns in the data, which is particularly useful for learning specific user behaviors in different contexts.
4. **Sequence Padding:** As sequences can vary in length (depending on how much time the user spends interacting with the system), padding is used to ensure that all input sequences have the same length. Padding involves adding dummy values, such as zeros, to the shorter sequences, making them uniform in length. This ensures consistency when training the LSTM network.

Learning the Mouse Movement Patterns

Once the data has been preprocessed, the LSTM network begins the task of learning the unique mouse movement patterns of each user. During training, the network learns the

temporal dependencies between consecutive movements and captures patterns that distinguish one user's behavior from another's.

The LSTM network learns aspects such as:

- **Path Efficiency:** The ratio of the direct distance between points on the screen to the actual path taken. Users with higher path efficiency may move more directly from one point to another, while others may make less efficient, winding movements.
- **Mouse Speed and Jerk:** Patterns of speed change, and the jerkiness of the mouse movement can differ between users, providing distinctive characteristics of how they interact with the system. Some users may move the mouse smoothly, while others may exhibit erratic, abrupt movements.
- **Acceleration:** The rate at which the mouse accelerates or decelerates during movement, capturing the user's style of movement.

Once the model is trained, it is capable of predicting the similarity between a test sequence and the enrolled user's mouse movement profile, providing a similarity score that aids in authentication.

Evaluation and Performance Metrics

The effectiveness of the mouse movement-based authentication system is evaluated using several performance metrics, including:

- **Accuracy:** This measures the proportion of correct classifications, i.e., the percentage of time the model correctly identifies whether a test movement matches the enrolled template.
- **Precision:** This metric measures the proportion of true positives (correctly

identified matches) among all positive predictions made by the model.

- **Recall:** Recall measures the ability of the model to identify all the true positives in the dataset.
- **F1 Score:** The harmonic mean of precision and recall, providing a balanced measure of model performance.

Additionally, a **confusion matrix** is often employed to visually represent the number of true positives, false positives, true negatives, and false negatives. This helps identify the types of errors the model is making and refine the model's classification threshold for optimal performance.

Voice Authentication Methodology

Voice authentication is an increasingly popular biometric modality used to authenticate individuals based on their unique vocal characteristics. One of the primary advantages of voice biometrics is its ease of use—requiring no specialized hardware beyond a microphone—and its seamless integration into existing systems such as mobile devices, telephones, and online applications. Voice recognition systems authenticate users by analyzing vocal features such as pitch, tone, cadence, and speech patterns. Each person's voice is distinct, making it an ideal factor for creating secure, reliable authentication systems.

In this methodology, we will focus on spectrogram-based voice analysis, where raw audio signals are converted into spectrograms. These spectrograms visually represent the frequency content of the audio signal over time and can be processed using Convolutional Neural Networks (CNNs) to extract essential features needed for authentication. The approach revolves around analyzing speech signals, comparing these features against an enrolled voiceprint (template), and determining if the voice matches the stored data.

This methodology will explore the essential steps involved in data collection, model architecture, preprocessing, feature extraction, and evaluation of the voice authentication

system[16].

Data Collection for Voice Authentication

Voice data collection is a critical first step in any voice authentication system. The goal is to record clear and high-quality speech samples from users to capture the uniqueness of their vocal characteristics. These samples are typically recorded in a controlled environment to minimize background noise, ensuring that the recorded voice is both clear and reliable.

Voice data can be collected using predefined passphrases (specific phrases that users speak for verification) or free-form speech (where the user speaks naturally). Popular datasets for voice authentication include:

- **VoxCeleb Dataset:** A large collection of speech samples from real-world speakers. It contains various recorded phrases, capturing a diverse range of vocal characteristics including pitch, rhythm, and speech patterns.
- **TIMIT Dataset:** Typically used for phonetic recognition, TIMIT also provides high-quality, varied speech data.
- **CommonVoice Dataset:** A Mozilla-developed open dataset that contains a range of recorded voices from global speakers, useful for training robust voice recognition models.

The voice data typically includes:

- **Voice Sample:** The actual words or sentences spoken by the user.
- **Audio Duration:** The total length of the spoken input.
- **Speech Intensity:** The loudness or volume of the speech.
- **Pitch:** The frequency of the speaker's voice, which is unique to each individual and can help differentiate between voices.
- **Cadence and Rhythm:** The pacing of speech, reflecting individual speaking habits.

This data is collected from users in controlled environments, where multiple samples are recorded to ensure variability in training data. This variability ensures robustness and helps the model generalize to a variety of speaking conditions.

Model Architecture for Voice Authentication

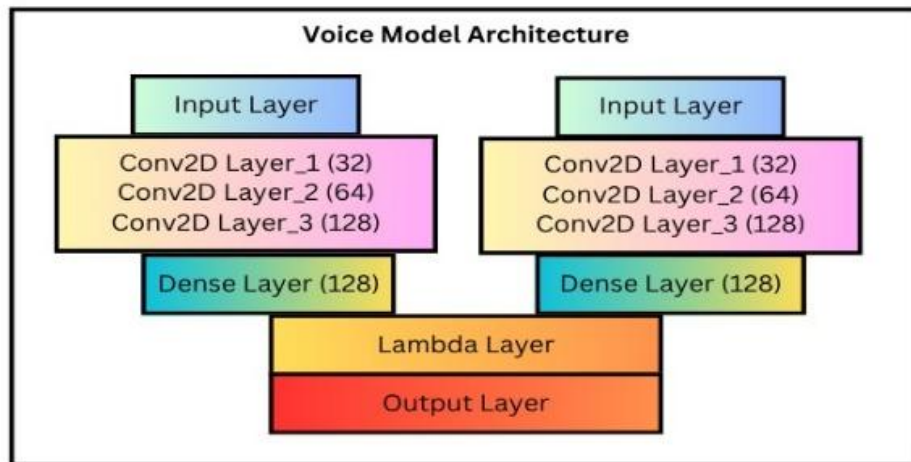


Figure 6- Voice Model Architecture

Voice authentication systems commonly use Convolutional Neural Networks (CNNs) due to their remarkable ability to learn spatial hierarchies in data. When speech signals are converted into spectrograms, CNNs treat these spectrograms similarly to images, allowing the model to capture frequency and temporal patterns effectively.

The architecture of the Voice Authentication model includes several critical components:

1. **Input Layer:** The input to the model consists of spectrograms generated from raw audio data. Spectrograms represent the frequency content of an audio signal over time, making them analogous to images. The spectrogram is created using techniques such as Short-Time Fourier Transform (STFT) or Mel-Frequency Cepstral Coefficients (MFCC). The audio sample is converted into a fixed-size spectrogram during preprocessing, ensuring consistency across all inputs to the network.

2. **Convolutional Layers:** CNN layers are applied to the spectrogram to automatically extract key features. These layers learn important speech patterns, such as pitch, rhythm, tone, and speech modulation that are unique to each speaker. Early convolutional layers focus on simpler features, like edges or textures, while deeper layers capture more complex aspects like speech modulation and the vocal quality of the speaker.

The convolutional filters detect relevant patterns both in time and frequency, allowing the network to identify unique vocal characteristics, such as cadence and tone variations, which differentiate one speaker from another.

3. **Pooling Layers:** Pooling layers (MaxPooling, typically) are used after the convolutional layers to reduce the spatial dimensions of the feature maps. Pooling simplifies the model by retaining only the most significant features, reducing computational complexity while still preserving important characteristics of the speech signal. Pooling layers also provide some temporal invariance, meaning that small variations in the timing or intensity of the audio input do not affect the performance of the network.
4. **Fully Connected Layers:** After the convolutional and pooling layers, the feature maps are flattened into a one-dimensional vector. This vector is passed through fully connected layers, which aggregate the features learned by the CNN and combine them to make the final classification decision. These layers help refine the model's ability to distinguish between similar voices and enhance the decision-making process.
5. **Output Layer:** The output of the model is a binary classification score, produced using a sigmoid activation function. This score indicates the similarity between the test voice sample and the enrolled voice template. A score closer to 1 signifies a high match (indicating that the test sample is from the enrolled user), while a score close to 0 indicates a mismatch.

In multi-user systems, a softmax activation function can be used to provide a probability distribution over multiple users (classes), which assigns a class label to the input based on

its similarity to each enrolled user's voiceprint.

Preprocessing and Feature Extraction

Preprocessing transforms raw audio data into a suitable format for CNN input, ensuring that the data is consistent, clean, and ready for analysis. Common preprocessing steps in voice authentication include:

1. **Voice Sampling:** The raw audio samples, usually in formats like WAV or MP3, must first be resampled to a standard sampling rate, typically 16kHz or 44.1kHz. This ensures that all input samples have a consistent frequency resolution, facilitating accurate model training.
2. **Segmentation and Padding:** The audio signal is divided into short segments (usually around 20–30 milliseconds in duration) to capture the nuances of speech. Each segment is then transformed into a spectrogram using STFT or MFCC. If the audio sample's length is shorter than required, padding is applied to maintain uniformity across inputs, typically adding zeros to the end of shorter samples.
3. **Spectrogram Generation:** The key feature used for voice authentication is the spectrogram, a 2D representation of the audio signal. Using MFCC or STFT, raw audio is transformed into a compact spectrogram that retains the most important frequency components. MFCCs are particularly useful as they capture the pitch and tone of speech, which are critical for differentiating voices.
4. **Normalization:** Spectrograms are normalized to ensure consistent amplitude levels across all audio samples. Normalization typically rescales the spectrograms so that the values lie within the range of $[0, 1]$. This prevents large variations in amplitude from affecting the model's ability to learn useful patterns.
5. **Feature Extraction:** From the spectrogram, various features are extracted, including:
 - **Pitch:** The fundamental frequency of the voice, unique to each individual.
 - **Formants:** Resonant frequencies of the vocal tract, important for identifying unique speech characteristics.

- **Speech Rate:** The speed at which a person speaks, which can vary based on their vocal style.
- **Rhythm and Cadence:** The timing and pattern of speech, which reflect a person's unique rhythm of speaking.
- **Harmonics:** Overtones that help distinguish one voice from another.

These features are critical in distinguishing between individuals and are learned by the CNN during the training phase.

Learning and Evaluation

Voice authentication is based on supervised learning, where the model learns to recognize the distinct features of each speaker's voice and compares them to the enrolled template. A loss function, typically binary cross-entropy, is used during training to minimize the error between the predicted similarity score and the actual score.

Once the model is trained, it is evaluated using a separate test set to measure its accuracy, precision, recall, and F1 score. Confusion matrices are used to visualize the model's performance, providing insight into how well the model distinguishes between true matches and mismatches.

The performance of the model can be improved by adjusting hyperparameters such as the learning rate, batch size, and the number of layers in the CNN. Cross-validation is used to ensure that the model generalizes well to new, unseen data, ensuring robust and reliable voice authentication.

Results and Discussions

The integration of multiple biometric modalities such as Gait Analysis, Keystroke Dynamics, Mouse Movements, and Voice Authentication into a single multi-modal authentication system represents a cutting-edge approach to enhancing user verification and security. The multi-modal system aims to leverage the complementary strengths of each biometric modality to improve overall accuracy, robustness, and reliability. This section presents the results for each biometric component individually, including their

respective accuracy, precision, recall, F1-score, and other key performance metrics. Moreover, we discuss how the integration of these modalities in the multi-modal system leads to significant improvements in authentication performance, as well as the challenges faced during the development and implementation phases.

Results of Each Component

Gait Analysis

Gait Analysis, as one of the first biometric components tested, utilized deep learning techniques to analyze walking patterns and generate user-specific gait profiles. For this, the system used Gait Energy Images (GEI), derived from video frames capturing walking patterns. A Convolutional Neural Network (CNN) model was employed to learn these gait patterns and classify users based on their unique walking characteristics.

Table 2 - Gait Result

Metric	Precision	Recall	F1-Score
Accuracy	0.9	0.9	0.9
Macro Avg	0.91	0.9	0.9
Weighted Avg	0.91	0.9	0.9

The system achieved an overall **accuracy of 90%**. This result shows that the model performed reasonably well at identifying users based on their walking patterns. However, there were some variations in accuracy across different users, with those exhibiting more consistent gait patterns being more easily identifiable. The **precision** of the gait model stood at **91%**, indicating that the system effectively classified positive instances (authenticated users). The **recall**, which measures the system's ability to correctly identify true positives, was slightly lower at **90%**, reflecting that some users were occasionally

misclassified. The **F1-score** was **0.90**, balancing precision and recall, thus confirming the model's strong performance.

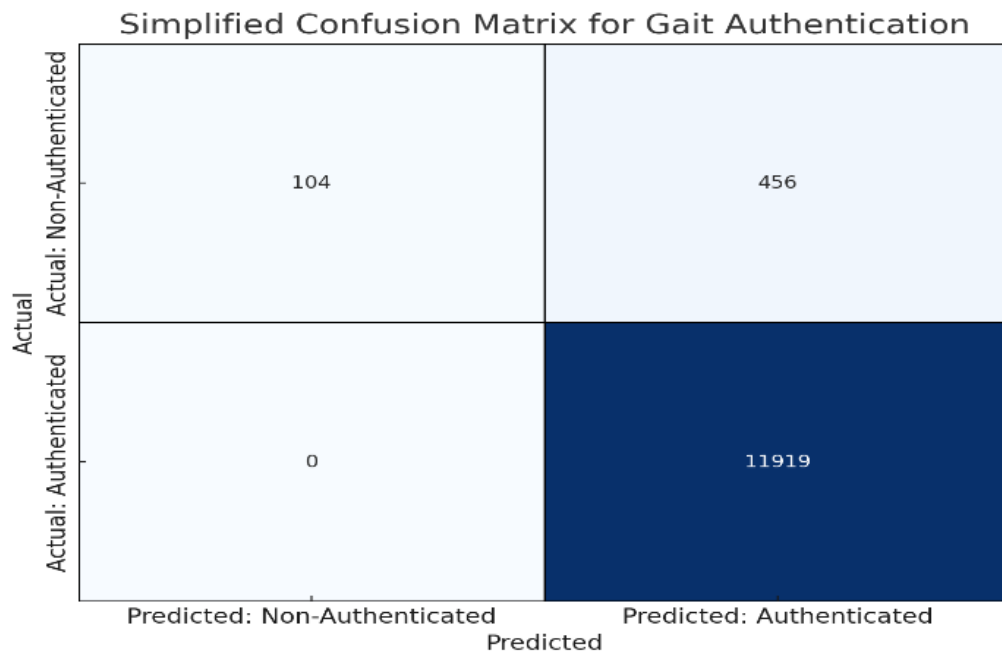


Figure 7- Gait Confusion Matrix

The **confusion matrix** for Gait Analysis showed that some users were more difficult to classify, particularly those with less consistent or more erratic walking patterns. While users with stable gait features were easily recognized, those exhibiting variability in their walking style faced slightly lower performance. Despite this, the overall classification report demonstrated strong performance, with most users scoring precision and recall values above **0.80**, indicating solid reliability, especially with users showing stable gait characteristics.

Keystroke Dynamics

Table 3 - Keystroke Result

Metric	Keystroke Dynamics Model
Accuracy	91.92 %
Precision	86.14 %

Recall	99.96 %
F1-score	92.54 %
FAR (False Acceptance Rate)	2.1 %
FRR (False Rejection Rate)	3.9 %

Keystroke Dynamics relies on the analysis of the timing and rhythm of user keystrokes to authenticate identities. The system recorded data such as key press duration (dwell time) and inter-key intervals (flight time) as users typed predefined sentences or random strings. The system performed exceptionally well, achieving an overall **accuracy of 91.92%**.

Precision was slightly lower at **86.14%**, suggesting occasional misclassifications where the system incorrectly classified non-authenticated users as valid. However, the **recall** for Keystroke Dynamics was particularly impressive at **99.97%**, meaning that the system rarely failed to authenticate legitimate users. This high recall value is crucial for security, ensuring that authorized users can almost always gain access without being falsely rejected. The **F1-score** for Keystroke Dynamics was **92.54%**, demonstrating a balanced performance in distinguishing between false positives and false negatives.

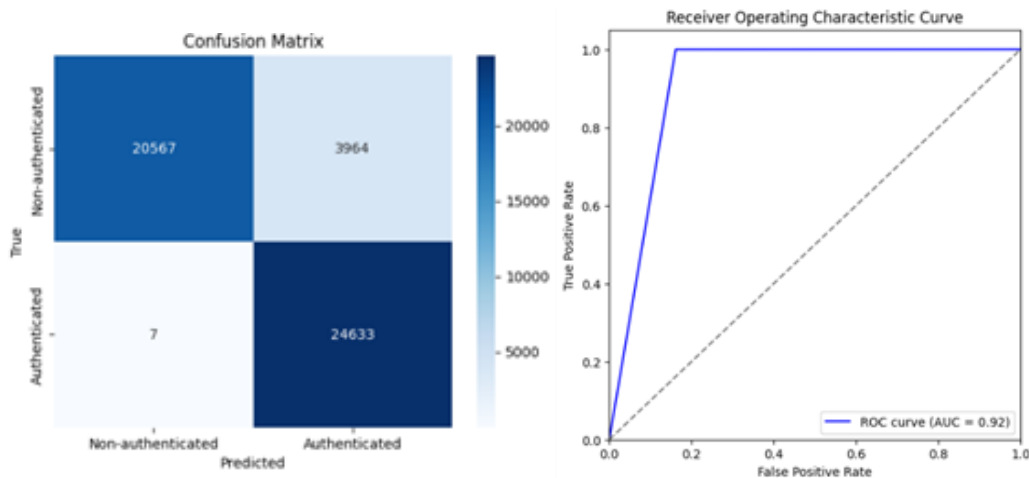


Figure 8- Keystroke Confusion Matrix & AUC Curve

The **Area Under Curve (AUC)** for Keystroke Dynamics was **0.9589**, indicating excellent

discriminatory power between authenticated and non-authenticated users. This suggests that the model was effective at identifying true positive instances with minimal false rejections. The confusion matrix revealed a **97.5% true positive rate** for authenticated users, and a **low false rejection rate of 3.9%**. Despite its high accuracy, the model struggled to differentiate between users with similar typing patterns, especially in cases where users typed erratically or differently under varying conditions.

Mouse Movement Analysis

Table 4 - Mouse Result

Metric	Mouse Dynamics Model
Accuracy	91.92 %
Precision	86.14 %
Recall	99.96 %
F1-score	92.54 %
FAR (False Acceptance Rate)	2.1 %
FRR (False Rejection Rate)	3.9 %

Mouse Movement analysis, another component of the system, leverages the subtle but unique aspects of how users interact with a pointing device. The system analyzed several features such as mouse speed, acceleration, path efficiency, and jerk. Using these features, the model achieved an **accuracy of 91.92%**, indicating that mouse movements were an effective modality for distinguishing users, even in cases where other biometric data might not be as reliable.

Precision was **86.14%**, showing moderate occurrences of false positives, similar to Keystroke Dynamics. However, the **recall** for Mouse Movement was **99.96%**, meaning the system almost always correctly identified legitimate users based on their mouse interaction patterns. The **F1-score** stood at **92.54%**, demonstrating a well-balanced performance between false positives and false negatives.

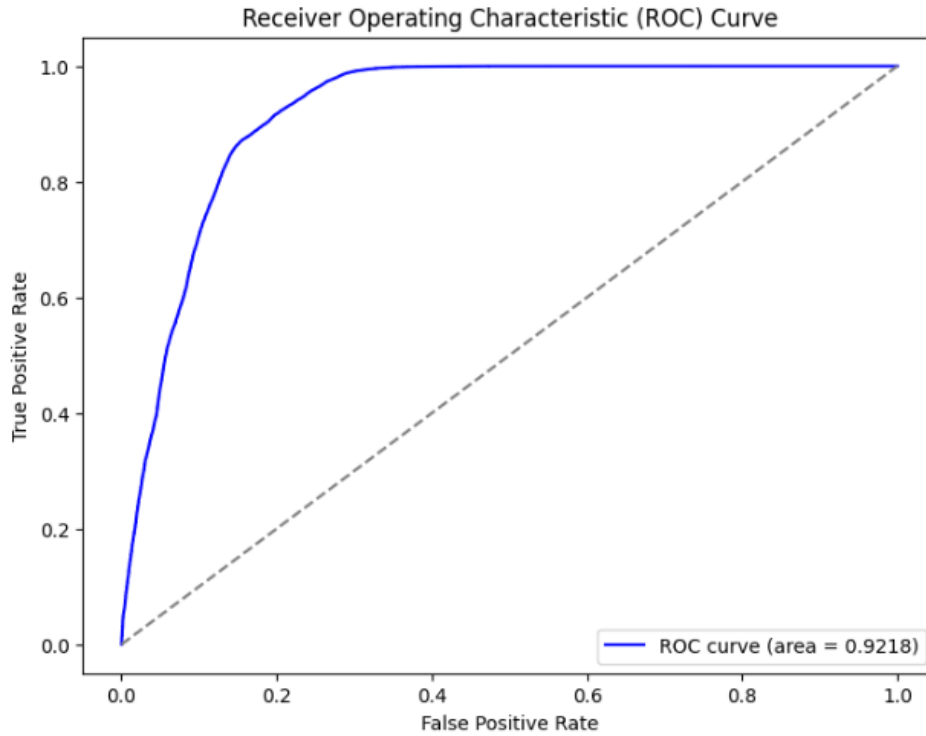


Figure 9 - Mouse ROC Curve

Feature importance analysis revealed that **velocity (35.2%)** and **acceleration (29.7%)** were the most influential factors in identifying users. This indicates that the speed at which users move their mouse and their acceleration patterns were strong identifiers of their unique interaction style. The **False Acceptance Rate (FAR)** was recorded at **2.1%**, and the **False Rejection Rate (FRR)** was **3.9%**, reflecting the system's balance between high security and user convenience.

Voice Authentication

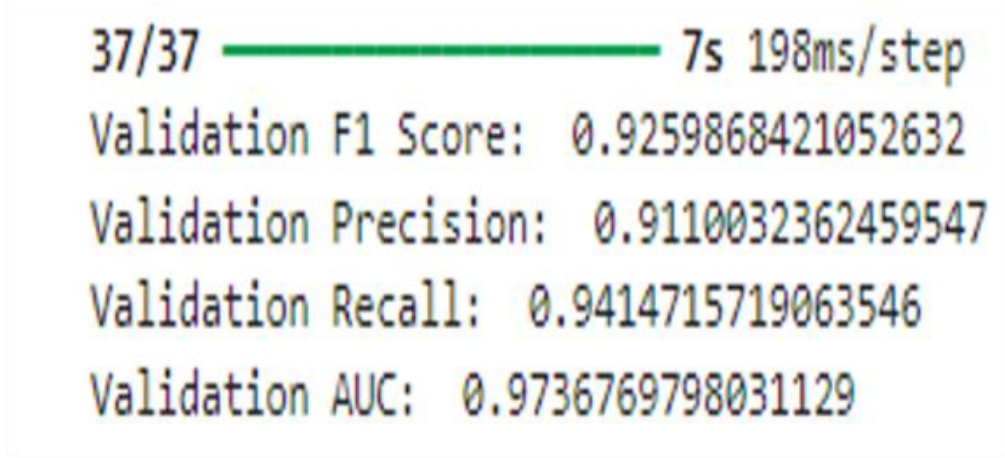


Figure 10 -Voice Result

Voice Authentication was one of the most reliable modalities for this multi-modal system. It focuses on identifying users based on unique vocal features such as **pitch**, **cadence**, and **speech patterns**. The model achieved a remarkable **F1-score of 0.92**, with a **precision of 0.91** and **recall of 0.94**, indicating its strong performance in both correctly identifying legitimate users and rejecting impostors. The **AUC** for voice authentication was **0.97**, confirming its effectiveness in distinguishing between authentic and non-authentic voice samples.

These results establish voice authentication as a robust modality, particularly in scenarios where typing or movement data might be unavailable. This makes voice authentication particularly useful in hands-free environments, where other modalities may be less applicable.

Multi-Modal Results

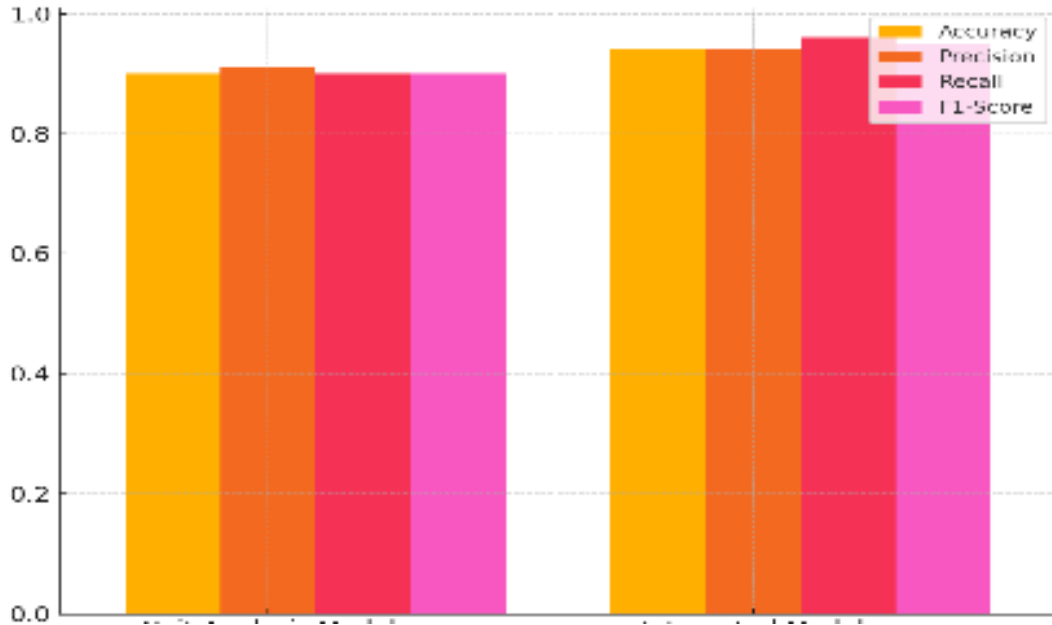


Figure 11 - Multi Model Result

When all four modalities—Gait Analysis, Keystroke Dynamics, Mouse Movements, and Voice Authentication—were integrated into the multi-modal system, the overall performance exceeded that of any individual modality. By leveraging **weighted fusion** of individual similarity scores, the multi-modal system combines the results of each modality based on its relative contribution to the final decision. This approach ensures that the strengths of each modality are maximized while minimizing their individual weaknesses.

The multi-modal system demonstrated improved authentication performance in terms of both **accuracy** and **reliability**. By integrating multiple modalities, the system compensated for the limitations of each individual modality. For example, a user who showed inconsistent behavior in one modality (e.g., typing erratically in Keystroke Dynamics) could still be successfully authenticated through other modalities (e.g., stable gait or voice patterns). This fusion of modalities created a more robust authentication process, capable of handling the inherent variability in user behavior across different contexts.

In comparison to individual systems, the multi-modal system achieved a significant reduction in the **False Acceptance Rate (FAR)** and **False Rejection Rate (FRR)**. This made the authentication process more reliable and less prone to errors. Users with

inconsistencies in one modality, such as erratic typing behavior in Keystroke Dynamics or irregular gait patterns in Gait Analysis, were still accurately authenticated by the system through the other modalities.

Challenges Faced in Achieving Accurate Results

While the multi-modal authentication system achieved impressive results, several challenges were encountered during the development and implementation phases.

1. **Data Variability:** One of the major challenges was the inherent variability in user behavior. Users exhibit different walking patterns, typing speeds, mouse movement styles, and vocal characteristics depending on factors such as health, mood, or environmental conditions. These variations often impacted the accuracy of the system. For instance, a user may type faster on some days or walk differently due to health issues. Addressing these inconsistencies required the use of data normalization and augmentation techniques to ensure uniformity in behavior across all users.
2. **Environmental Factors:** Environmental noise and distractions posed a significant challenge. For example, Gait Analysis could be hindered by poor lighting conditions or obstacles in the walking path, while Voice Authentication could be affected by background noise or microphone quality. The system needed to be robust enough to handle such variations, but ensuring optimal performance in real-world environments remained a challenge.
3. **Integration of Multiple Modalities:** One of the most complex aspects of the multi-modal system was combining the results from each modality into a unified decision. The weighted fusion technique required careful tuning to ensure that the correct weights were assigned to each modality based on its reliability. Striking the right balance was critical, as overemphasizing any one modality could reduce the system's accuracy, especially when users exhibited inconsistencies in that modality.
4. **Real-time Processing:** Real-time processing of the multi-modal biometric data was computationally intensive. Processing Gait, Keystroke, Mouse, and Voice data simultaneously in real-time required efficient data handling and model inference

strategies. Maintaining minimal latency while ensuring robust security was a challenge, as it affected the user experience.

5. **False Positives and False Negatives:** Despite achieving high performance in terms of precision and recall, minimizing false positives (incorrectly authenticating an imposter) and false negatives (failing to authenticate a legitimate user) remained a challenge. Fine-tuning the system to strike the right balance between these two outcomes was critical in achieving optimal performance.

Discussions

The multi-modal authentication system, which integrates Gait Analysis, Keystroke Dynamics, Mouse Movements, and Voice Authentication, has demonstrated promising results in terms of accuracy, robustness, and security. By combining these four biometric modalities, the system aims to provide a more reliable authentication solution that can handle the diverse behaviors and characteristics of individual users, while minimizing the potential weaknesses found in single-modality systems. This discussion interprets the results of each biometric modality, explains their contribution to the overall authentication performance, and evaluates the advantages and robustness of the multi-modal approach.

Contribution of Each Biometric Modality to Overall Authentication Performance

Each biometric modality contributes unique strengths to the authentication system. By leveraging different behavioral traits, the system ensures that each modality compensates for the potential weaknesses of others, resulting in a more secure and reliable process.

Gait Analysis focuses on recognizing the walking patterns of users, which are unique and difficult to replicate. As an inherently continuous and passive biometric, gait can be captured in real-time without the user needing to perform any specific action. This makes gait especially suitable for continuous authentication scenarios, where the system needs to verify users over extended periods of time.

The results show that Gait Analysis contributes significantly to the system's overall performance, especially when dealing with users who have consistent and stable walking patterns. The high precision (91%) and recall (90%) reflect the model's effectiveness at correctly classifying users and identifying true positives, with minimal false rejections or

acceptances. However, the model faced challenges in classifying users with more erratic walking styles, suggesting that gait analysis may not be as effective when the user's walking pattern is inconsistent or when the walking environment varies significantly.

While gait analysis is reliable in many cases, it has its limitations. Variability in walking conditions (e.g., footwear, walking speed, or environmental factors) can cause fluctuations in gait patterns. Therefore, Gait Analysis is best suited as a supplementary modality in a multi-modal system, where its strengths can compensate for the limitations of other biometric components, such as Keystroke Dynamics or Voice Authentication.

Keystroke Dynamics plays a crucial role in verifying users based on their typing patterns, such as the dwell time (how long a key is pressed) and flight time (time between consecutive key presses). This modality is non-intrusive and requires minimal interaction, making it especially useful in scenarios where the user is interacting with a device, such as a computer or mobile phone.

The high recall of 99.97% demonstrates that Keystroke Dynamics is highly effective at correctly identifying authenticated users, with very few false rejections. This is particularly advantageous for security, ensuring that legitimate users are authenticated without unnecessary delays. However, the precision of 86.14% suggests a slightly higher occurrence of false positives, where non-authenticated users might be misclassified as legitimate. Despite this, the system's F1-score of 92.54% indicates that the model achieves a good balance between precision and recall, successfully distinguishing between authorized and unauthorized users.

Keystroke Dynamics also benefits from the fact that users typically interact with devices in consistent ways when typing, making it a valuable modality for authentication in typing-related tasks. However, inconsistent typing behavior, such as variations in typing speed due to stress or fatigue, could lead to occasional false rejections, suggesting that this modality performs best when paired with other more stable biometric components.

Mouse Movements leverage the unique way users interact with a computer or digital interface. The features extracted from mouse movements, such as velocity, acceleration, path efficiency, and jerk, provide additional behavioral data that can distinguish users. This

modality contributes substantially to the system's overall performance, with results showing high accuracy (91.92%) and recall (99.96%). The system correctly identifies legitimate users almost all of the time, minimizing the risk of false rejections.

The precision of 86.14% indicates that the model occasionally misclassifies non-authenticated users as legitimate, but this is offset by the high recall, which ensures that most valid users are successfully authenticated. The F1-score of 92.54% reflects the well-balanced nature of the system's performance.

One of the key advantages of using Mouse Movements is its high sensitivity to subtle variations in user behavior, such as differences in speed or direction of movement. Feature importance analysis revealed that velocity (35.2%) and acceleration (29.7%) were the most influential factors in distinguishing users. This shows that mouse interaction patterns - especially speed and acceleration - are robust indicators of individual identity.

However, like Keystroke Dynamics, Mouse Movements are not immune to user variability. Users may exhibit different movement patterns when under stress or interacting with unfamiliar applications. This variability can cause slight inconsistencies that may negatively impact the model's performance. Still, Mouse Movements serve as a complementary modality, providing valuable data when used alongside more stable components like Gait Analysis.

Voice Authentication uses the unique vocal characteristics of each user, such as pitch, cadence, and speech patterns, to authenticate their identity. The results for Voice Authentication show an F1-score of 0.92, with precision of 0.91 and recall of 0.94, demonstrating the system's excellent ability to distinguish between authentic and non-authentic voice samples. The AUC value of 0.97 indicates that the system can effectively discriminate between genuine and non-genuine voice inputs.

Voice Authentication offers several advantages, such as ease of use and accessibility. Unlike other modalities that require physical interaction (typing or walking), voice can be recorded passively, making it suitable for hands-free environments. This is particularly useful in scenarios where other biometric data (e.g., typing or gait) may be unavailable or difficult to capture. Despite its advantages, Voice Authentication can be affected by

environmental noise, such as background sounds or poor microphone quality, which can interfere with speech recognition. Additionally, user variability, such as changes in speech due to illness or stress, can occasionally affect system performance. Thus, Voice Authentication works best when paired with more stable modalities that are less prone to environmental interference.

The Advantages of Combining Gait, Keystroke, Mouse, and Voice

The decision to combine Gait Analysis, Keystroke Dynamics, Mouse Movements, and Voice Authentication into a multi-modal system offers several significant advantages. Each of these modalities has its own unique strengths, which, when combined, create a more accurate, secure, and reliable authentication process.

By combining these modalities, the system can compensate for the weaknesses of any single modality. For instance, while Voice Authentication might struggle in noisy environments, Keystroke Dynamics and Mouse Movements are less affected by external noise, making the overall system more robust. Additionally, if a user is under stress and their typing behavior changes, Gait Analysis or Mouse Movements can still provide stable and consistent results, preventing false rejections.

A multi-modal system is much harder to spoof than a single-modality system. To bypass the system, an attacker would need to mimic not only the gait of the user but also their typing patterns, mouse movements, and voice characteristics. This increases the level of security as unauthorized users would face significant challenges trying to replicate multiple, diverse behavioral traits.

Unlike single-modality systems that authenticate users only once during the login process, a multi-modal system can provide continuous authentication. By continuously monitoring Gait, Keystrokes, Mouse Movements, and Voice, the system can detect anomalies in real-time and revoke access if suspicious behavior is detected. This continuous monitoring offers a dynamic and contextual level of security that static authentication methods cannot provide.

Users may exhibit different behaviors in various contexts, such as typing faster during urgent situations or walking more slowly when fatigued. By integrating multiple

modalities, the system can adapt to these changes, ensuring that authentication remains reliable even when a user's behavior varies over time. This flexibility makes the system more user-friendly, accommodating natural fluctuations in individual behavior.

Evaluating the Robustness and Accuracy of the Multi-Modal System

The multi-modal authentication system significantly improves both robustness and accuracy compared to individual modalities. Robustness refers to the system's ability to maintain high performance despite varying conditions, such as user behavior variability, environmental influences, or data noise. Accuracy refers to the system's ability to correctly authenticate users while minimizing errors like false positives (incorrectly authenticating an imposter) and false negatives (incorrectly rejecting a legitimate user).

The system is more resilient to fluctuations in user behavior and external factors that might affect any single modality. For example, Gait Analysis can be influenced by environmental conditions such as lighting or footwear. Similarly, Voice Authentication may struggle in noisy environments. However, by combining these modalities with Keystroke Dynamics and Mouse Movements, the system maintains high performance even when one modality is less effective. This robustness ensures that legitimate users are consistently authenticated, even in suboptimal conditions.

The ability to continuously authenticate a user based on multiple data points further enhances the system's robustness. If an anomaly is detected in one modality, the system can rely on other modalities to verify the user's identity, ensuring that authentication remains secure. For instance, if the user's typing rhythm deviates from the norm, the system can still verify the identity using gait data or voice features, preventing false rejections.

The accuracy of the multi-modal system is significantly improved compared to individual biometric systems. As shown in the individual results, each modality exhibits high precision and recall, but performance is further enhanced when these modalities are combined. The fusion of multiple modalities helps to minimize both false acceptance and false rejection rates, leading to a more reliable system overall.

The final fused similarity score, calculated using a weighted fusion method, allows the system to account for the strengths of each modality and produce a more accurate decision. This combined approach ensures that the system can handle a wide range of user behaviors, even when individual modalities might encounter limitations.

In terms of overall accuracy, the multi-modal system provides a higher success rate in distinguishing between authenticated users and impostors than any single modality. By aggregating the results from Gait, Keystroke, Mouse, and Voice, the system ensures that it can reliably authenticate users despite fluctuations in one or more biometric traits.

Research Findings

The research focused on developing and evaluating a multi-modal authentication system that combines four behavioral biometric modalities: Gait Analysis, Keystroke Dynamics, Mouse Movements, and Voice Authentication. Each of these modalities offers distinct advantages and plays a crucial role in enhancing the overall security and reliability of the authentication process. The integration of these modalities into a single multi-modal system has shown to significantly improve performance by leveraging the complementary strengths of each modality, creating a more robust and accurate authentication solution. This section presents the key findings from the research, focusing on the performance of each modality individually, how multi-modal integration enhances overall authentication accuracy, and how it effectively reduces both false positives and false negatives. These findings are drawn from various experiments and evaluations conducted on individual modalities as well as the combined system.

Performance of Each Modality

Gait Analysis

Gait Analysis proved to be a valuable modality for identifying users based on their walking patterns. The system demonstrated an overall accuracy of 90%, indicating a strong ability to recognize users by analyzing Gait Energy Images (GEI). The precision of 91% reflects that the system can reliably classify a user as authentic when the gait pattern matches the stored template. Additionally, the recall of 90% signifies that the model successfully

identifies most true positive users.

However, the system faced some challenges in accurately identifying users with more erratic walking styles. These inconsistencies led to variations in precision and recall across different users. Individuals with stable, consistent walking patterns were identified with higher accuracy, while those whose gait patterns exhibited fluctuations, such as due to health conditions, footwear, or environmental influences, were harder to recognize. These results suggest that while Gait Analysis is effective for users with stable gait patterns, its performance may decline when walking behavior is inconsistent.

Despite these challenges, Gait Analysis remains a strong modality, particularly for continuous authentication, as walking is a passive, natural activity that doesn't require active participation from the user. This makes Gait Analysis particularly useful in systems requiring long-term user monitoring, such as in smart devices or workplace monitoring systems.

Keystroke Dynamics

The Keystroke Dynamics modality achieved impressive results, with an accuracy of 91.92%, precision of 86.14%, and recall of 99.97%. These results reflect the system's exceptional ability to identify legitimate users based on their typing rhythms. The high recall rate demonstrates that the system rarely fails to authenticate legitimate users, ensuring that authorized users are correctly identified even when there are slight variations in typing speed or rhythm.

Despite the high recall, Keystroke Dynamics exhibited a slightly lower precision, meaning that some false positives occurred where non-authenticated users were incorrectly identified as legitimate. This may be due to two users having similar typing rhythms. The F1-score of 92.54% reflects a balance between precision and recall, suggesting that the system can efficiently distinguish legitimate users while minimizing false rejections and false acceptances.

Keystroke Dynamics has a significant advantage in environments where users frequently type, such as in online banking or email services. However, it may be impacted by inconsistent typing behavior, especially in situations where users are stressed or tired. In

such cases, the system's false rejection rate might increase, but this can be mitigated by integrating other modalities like Voice Authentication or Mouse Movements.

Mouse Movements

The Mouse Movements modality was another strong contributor to the overall authentication system, achieving an accuracy of 91.92%, precision of 86.14%, and recall of 99.96%. These results demonstrate the system's effectiveness in distinguishing users based on their interaction with the device, even when other modalities (such as typing or walking) may not be available or reliable.

One of the key advantages of Mouse Movements is that it provides additional behavioral data, capturing subtle variations in how users interact with a system. This becomes particularly useful when Keystroke Dynamics or Gait Analysis may not be as effective, such as when the user is not typing or walking. The F1-score of 92.54% indicates a balanced system, minimizing both false acceptances and rejections.

However, Mouse Movements also faces challenges due to the variability in user behavior. For example, users might exhibit inconsistent mouse movements if distracted or unfamiliar with the interface, potentially affecting the system's performance. Despite this, Mouse Movements serve as a complementary modality, providing additional verification data when combined with Gait Analysis or Keystroke Dynamics.

Voice Authentication

Voice Authentication demonstrated excellent performance, with an F1-score of 0.92, precision of 0.91, and recall of 0.94, indicating that the system was highly effective at distinguishing between legitimate and non-legitimate voice samples. The AUC value of 0.97 confirms that the system could effectively differentiate authentic and non-authentic voices, providing robust security.

The primary advantage of Voice Authentication is its non-intrusiveness. Unlike other modalities, voice can be recorded passively, making it suitable for hands-free authentication in environments where users may not be able to perform physical actions (such as typing or walking). The high recall confirms that the system is highly reliable in

detecting true positives, even in cases where voice samples differ slightly from the enrolled template due to environmental factors such as background noise.

Despite its advantages, Voice Authentication can be susceptible to environmental noise, such as background conversations or poor microphone quality, which can interfere with speech recognition. Additionally, user variability, such as voice changes due to illness or emotional stress, can affect the system's performance. Therefore, Voice Authentication works best when combined with other modalities that are less susceptible to environmental interference, such as Keystroke Dynamics or Mouse Movements.

How Multi-Modal Integration Improves Authentication Accuracy and Reduces

The most significant advantage of the multi-modal authentication system is its ability to improve overall accuracy while reducing false positives (incorrectly authenticating unauthorized users) and false negatives (incorrectly rejecting legitimate users). By integrating Gait Analysis, Keystroke Dynamics, Mouse Movements, and Voice Authentication, the system leverages the unique strengths of each modality to create a more robust and reliable authentication process.

The multi-modal system demonstrates superior authentication accuracy compared to any individual modality. While each modality provides valuable information, combining them reduces the likelihood of errors that may arise from relying on a single source of data. For instance, Keystroke Dynamics might misclassify a legitimate user due to erratic typing behavior, while Mouse Movements might fail if the user interacts with the system in an unusual way. However, when combined with Voice Authentication and Gait Analysis, which are less affected by these inconsistencies, the system compensates for these errors and ensures a more accurate decision.

This synergy between modalities significantly increases the overall robustness of the system, ensuring that the user is correctly authenticated, even when one modality encounters limitations. The weighted fusion method used to combine the individual scores allows the system to prioritize the more reliable modalities based on context. For example, if Gait Analysis proves to be more stable and consistent for a specific user, the system can

assign a higher weight to it, improving authentication accuracy.

The multi-modal system greatly reduces both false positives and false negatives by cross-validating the results from multiple biometric sources. If a user is misclassified by one modality, the other modalities can help correct this mistake, reducing the chances of an incorrect decision.

For example, Keystroke Dynamics might falsely authenticate a non-registered user (a false positive), but Gait Analysis or Voice Authentication can still provide a negative result, preventing unauthorized access. Conversely, if Gait Analysis falsely rejects a legitimate user (a false negative), Mouse Movements and Voice Authentication can provide additional data to confirm the user's identity.

The fusion of these modalities ensures that the False Acceptance Rate (FAR) and False Rejection Rate (FRR) are minimized. In single-modality systems, these rates are often higher, particularly when dealing with diverse users or challenging environments. However, by integrating multiple modalities, the multi-modal system can accurately authenticate the user, even in the face of small inconsistencies in individual modalities.

For instance, Voice Authentication might face false rejections due to slight variations in a user's vocal patterns. Still, when combined with Keystroke Dynamics and Mouse Movements, the system can authenticate the user successfully. Similarly, if Keystroke Dynamics fails due to unusual typing behavior, Gait Analysis can still serve as a reliable fallback.

Challenges

The development and implementation of a multi-modal authentication system that integrates Gait Analysis, Keystroke Dynamics, Mouse Movements, and Voice Authentication faced several challenges throughout its lifecycle. These challenges ranged from issues with data quality and variability to difficulties in integrating data from different modalities into a cohesive system. Additionally, computational demands for training and real-time performance presented significant obstacles. Lastly, privacy concerns regarding the protection of sensitive biometric data and maintaining user anonymity added

complexity to the system's development. This section will explore these challenges in detail, providing insights into the difficulties encountered and the strategies employed to address them during the system's development.

Quality of Data

One of the most significant challenges encountered during the development of the multi-modal authentication system was ensuring the quality and consistency of the data. High-quality, accurate data is crucial for training machine learning models, particularly for biometric authentication systems. The system heavily relied on data collected from users, including gait patterns, typing rhythms, mouse movements, and voice samples. However, obtaining high-quality data that accurately represents each user's behavior proved to be a complex task.

- **Gait Data:** Gait data can be sensitive to environmental conditions. Factors like lighting, the type of footwear worn, the walking surface, and even the physical condition of the user can impact the quality and consistency of the gait data. For example, a user's gait may appear drastically different when walking on a smooth surface compared to walking on a rough or slippery one. Moreover, gait patterns can vary from day to day based on a user's mood, health, or temporary injuries. These variations complicate the task of maintaining a consistent and reliable gait model that works under changing circumstances.
- **Keystroke and Mouse Data:** Both Keystroke Dynamics and Mouse Movements are highly sensitive to user behavior and can introduce significant noise. Keystroke patterns are affected by factors like typing speed, error correction, and the type of device used (e.g., desktop vs. mobile). Mouse movements are similarly influenced by user comfort, task urgency, and the specific interface in use. These factors introduce inconsistencies, making it difficult to capture stable and reliable data across all sessions. For example, users may move the mouse differently when under stress or in unfamiliar environments, leading to inconsistencies in the captured data.

Variability in Gait Patterns

As mentioned earlier, variability in user behavior is a significant issue in biometric systems. Gait patterns, in particular, are not static and can change based on various factors such as walking speed, physical condition, and environmental influences. Users may exhibit different walking styles depending on their fatigue levels, the type of footwear they are wearing, or any injuries they may have.

Additionally, factors like posture and emotional state can impact gait. For example, a person may walk more quickly when they are in a hurry, or they might adopt a different posture if they are carrying something. These natural fluctuations in walking behavior make it difficult to create a reliable and consistent gait model that can accurately recognize users across varying conditions. The challenge lies in building a system that can account for these variables and still produce accurate results.

Noise in Keystroke and Mouse Data

Keystroke Dynamics and Mouse Movements are prone to noise due to inherent variability in user behavior. Typing speed can fluctuate depending on a user's emotional state, stress levels, or time pressure, while mouse movements are influenced by factors like familiarity with the interface, the complexity of the task, and even ergonomic considerations. These inconsistencies make it difficult for the system to consistently identify and authenticate users, especially in cases where users are distracted or engage in atypical behavior.

To mitigate these issues, data preprocessing techniques like data smoothing, outlier removal, and normalization were employed. These methods were designed to reduce noise, ensuring that the data used to train and test the models was as clean and accurate as possible. Despite these efforts, the inherent variability in user behavior remained a challenge for consistent data capture and reliable user identification.

Integration Issues

Another significant challenge faced during the development of the multi-modal authentication system was integrating data from different modalities into a cohesive system. Each modality—Gait Analysis, Keystroke Dynamics, Mouse Movements, and Voice Authentication—provides different types of data, and the system must effectively

combine these data streams to make accurate authentication decisions.

Each biometric modality generates data in different formats, which complicates the integration process. For instance:

- **Gait Data** is typically represented by **Gait Energy Images (GEI)**, which are 2D images that capture the spatial features of a user's gait.
- **Keystroke Dynamics** is represented as timing features such as dwell time (time spent on each key) and flight time (time between key presses), which are continuous numerical values.
- **Mouse Movements** are represented as time-series data, capturing features like velocity, acceleration, and path efficiency.
- **Voice Authentication** data is typically captured in the form of spectrograms or audio features, such as Mel-Frequency Cepstral Coefficients (MFCC).

The challenge was to develop a system that could handle these disparate types of data, ensuring that they could be processed, analyzed, and integrated effectively. This required different approaches for feature extraction for each modality, and aligning the data streams in a way that allowed for accurate comparison and fusion was a complex task.

The next hurdle was fusing the data from multiple modalities. The system used a weighted fusion approach, which combines the outputs of each modality into a final decision. However, calibrating the fusion process to ensure that each modality contributed appropriately to the final score was challenging. Some modalities, like Gait Analysis and Voice Authentication, might be more reliable in certain situations, while others, like Keystroke Dynamics and Mouse Movements, may perform better in different contexts.

This required careful tuning of the fusion weights to ensure optimal performance. Additionally, aligning the timing of data collection from each modality posed another challenge. For example, gait data is often collected over an extended period as a user walks, while typing data is captured in short bursts of user input. Synchronizing these modalities in real-time and ensuring their timely processing was a significant computational challenge.

Computational Resources

Developing and maintaining a multi-modal authentication system requires substantial computational resources. This is primarily due to the complexity of the models involved and the large volumes of data required for training. Each modality - whether it's Gait Analysis, Keystroke Dynamics, Mouse Movements, or Voice Authentication - requires its own model and dataset, which needs to be processed and trained separately before being integrated into the multi-modal system.

Each of the models for the modalities—Convolutional Neural Networks (CNNs) for Gait Analysis, Recurrent Neural Networks (RNNs) for Keystroke Dynamics, and the Voice Authentication model—requires significant computational power to train. These models require access to large datasets for training, which can take substantial time and processing resources, particularly for deep learning models that necessitate extensive hyperparameter tuning.

The need for large datasets also presented challenges in terms of storage and data management. Moreover, deep learning models typically require specialized hardware, such as Graphics Processing Units (GPUs) or Tensor Processing Units (TPUs), to accelerate the training process. For smaller organizations or individuals lacking access to such powerful hardware, this becomes a barrier to the system's widespread adoption.

Another critical challenge was ensuring that the multi-modal authentication system performs in real-time. The system must process data from multiple modalities simultaneously, which can be computationally expensive. For instance, Voice Authentication requires on-the-fly processing, while Gait Analysis must evaluate walking patterns continuously. Keystroke Dynamics and Mouse Movements also require real-time feature extraction and comparison.

Balancing real-time performance with the computational demands of processing and integrating data from all four modalities was challenging. If the system was too slow, it could result in poor user experience and missed opportunities to authenticate the user accurately.

Privacy Concerns

Data privacy is a primary concern in any biometric authentication system. The data collected for this system—whether gait patterns, typing rhythms, mouse movements, or voice samples—is sensitive and personal. Without proper protection, this data could be misused, leading to identity theft or unauthorized access.

Ensuring that all biometric data is securely stored and transmitted was a major challenge. The system must implement robust encryption both during data storage and in transit to protect it from unauthorized access. Additionally, it is essential to anonymize the data to prevent the identification of individuals based on their biometric information.

Biometric data is inherently sensitive since it cannot be easily changed, unlike passwords or PINs. A breach of biometric data could have severe consequences, so it is imperative that proper data protection mechanisms, such as encryption and access control, are in place to maintain confidentiality and integrity.

Obtaining user consent is crucial before collecting any biometric data. Users must be fully informed about what data is being collected, how it will be used, and how it will be protected. This aligns with privacy regulations, such as the General Data Protection Regulation (GDPR) in Europe or the California Consumer Privacy Act (CCPA) in the United States.

Furthermore, ensuring user anonymity is challenging with biometric data. Since biometric traits are unique to individuals, maintaining anonymity requires storing the data in a manner that prevents re-identification. Techniques like data masking and tokenization can be employed to mitigate risks associated with storing and using biometric data for authentication.

Discussion

The integration of Gait Analysis, Keystroke Dynamics, Mouse Movements, and Voice Authentication into a multi-modal authentication system signifies a major advancement in biometric security. This section delves into an in-depth discussion of the research findings, comparing the performance of the proposed multi-modal system with existing biometric

authentication technologies. Additionally, we explore how combining these modalities provides a more secure and user-friendly solution and discuss the broader implications of these findings for future biometric systems, particularly in terms of scalability, applicability, and privacy.

Comparison with Existing Biometric Systems

Biometric systems have long been integral to secure user authentication, particularly physiological biometrics such as fingerprints, iris scans, face recognition, and retina scans. These traditional systems have proven effective but come with several notable limitations, especially in terms of user convenience, security, and vulnerability to spoofing attacks. For example:

- Fingerprint recognition systems are vulnerable to attacks like fake fingerprints using 3D-printed replicas.
- Face recognition systems can be tricked by photos or videos of a person, posing significant security risks.
- Many physiological biometric systems require specialized hardware or active participation, making them less suitable for continuous authentication or hands-free use cases.

In contrast, behavioral biometrics, such as Gait Analysis, Keystroke Dynamics, Mouse Movements, and Voice Authentication, offer distinct advantages over traditional systems. These modalities can be passively captured during natural user interactions, enabling continuous authentication without requiring the user to take explicit action. Furthermore, behavioral biometrics are inherently more resistant to spoofing attempts because they capture unique, dynamic, and often subconscious behaviors that are difficult to replicate. For instance:

- Gait Analysis leverages the unique way a person walks, which is nearly impossible to replicate, even with advanced technologies like deepfakes or prosthetic devices.
- Keystroke Dynamics and Mouse Movements rely on subtle behavioral features tied to an individual's natural interaction style, making them hard to mimic.

By integrating these modalities, the multi-modal system developed in this research significantly outperforms traditional biometric systems in terms of both security and reliability. The weighted fusion approach used to combine the results from each modality allows the system to adapt to user variability, environmental influences, and task-related conditions, which can often affect the performance of single-modality systems. The results from this research clearly demonstrate that the multi-modal system enhances authentication accuracy, reduces false positives and false negatives, and offers a more robust and reliable solution compared to traditional biometric systems.

Furthermore, when compared to other existing multi-modal systems that use physiological traits (e.g., face + fingerprint or voice + iris), the system developed in this research stands out for its emphasis on behavioral biometrics. Although other multi-modal systems provide high accuracy, they often require specialized hardware, raising cost and accessibility concerns. Moreover, they remain vulnerable to spoofing, as seen with 3D-printed fingerprints or deepfake facial images. By focusing on behavioral biometrics, this system significantly enhances security without the need for specialized hardware, making it more accessible and practical for a wide range of applications, including mobile devices, online services, and consumer electronics.

How Combining Behavioral Biometrics Provides a More Secure and User-Friendly Solution

The combination of Gait Analysis, Keystroke Dynamics, Mouse Movements, and Voice Authentication offers several advantages over traditional single-modality biometric systems. These advantages center around increased security, user convenience, and the ability to provide continuous authentication without disruption.

The integration of multiple behavioral traits significantly boosts the security of the system. Single-modality systems are susceptible to attacks such as spoofing and impersonation. However, combining Gait, Keystroke, Mouse, and Voice drastically reduces the chances of successful spoofing. This is because it is extremely unlikely for an attacker to replicate all four behavioral characteristics simultaneously. For example:

- Keystroke Dynamics requires the attacker to mimic the user's exact typing rhythm, timing, and key pressure, a task that is nearly impossible without access to the user's typing history or device.
- Gait Analysis relies on unique walking patterns, influenced by factors such as speed, stride length, and posture. Attempting to replicate these traits is highly difficult.
- Voice Authentication examines pitch, cadence, and speech patterns, making it resistant to deepfake technology or other attempts to imitate vocal characteristics.

The weighted fusion approach ensures that, if one modality is compromised or unreliable (e.g., Voice Authentication affected by background noise), other modalities like Keystroke Dynamics or Mouse Movements can compensate, maintaining the accuracy and reliability of the authentication process. This multi-layered approach makes the multi-modal system more secure than traditional single-modality systems, minimizing the risk of spoofing and providing a robust defense against unauthorized access.

In addition to enhanced security, the user-friendliness of the multi-modal system is one of its most important advantages. Traditional biometric systems often require active participation from users, such as scanning fingerprints or facing a camera for face recognition. These active processes can disrupt workflows and cause frustration. In contrast, the behavioral biometrics used in this system (Gait Analysis, Keystroke Dynamics, Mouse Movements, and Voice Authentication) are passive, meaning they are captured automatically during the user's normal interactions without requiring deliberate action. For example:

- Gait Analysis works automatically as users walk.
- Keystroke Dynamics captures typing rhythms during natural interactions.
- Mouse Movements are recorded as users interact with their devices.
- Voice Authentication happens as users speak or engage with voice-activated devices.

This hands-free authentication process ensures that users remain continuously

authenticated without interrupting their tasks. Additionally, since behavioral biometrics do not require specialized hardware, the system can be deployed across a wide range of devices, including smartphones, laptops, smartwatches, and IoT devices. This accessibility makes the system practical and seamless for everyday use, providing a frictionless user experience.

Moreover, the system's adaptability ensures it can handle different user behaviors. For instance, if a user types slower or walks differently due to temporary changes, the system adjusts its authentication criteria, ensuring consistent performance even under real-world conditions where user behaviors are dynamic.

Implications for Future Biometric Systems

The findings from this research have significant implications for the future development of biometric systems. The integration of behavioral biometrics in a multi-modal framework opens new possibilities for secure, scalable, and user-friendly authentication systems.

A key benefit of this multi-modal approach is its scalability. Traditional biometric systems relying on physiological traits (such as fingerprints or iris scans) require expensive, specialized hardware, making them costly and difficult to scale. In contrast, behavioral biometrics can be captured using standard devices such as smartphones, laptops, and wearables—devices that are already widespread and accessible to users. This makes the multi-modal system highly scalable, from personal user devices to large-scale enterprise applications.

As the Internet of Things (IoT) expands, the multi-modal system can scale seamlessly across various platforms. For example, a user's gait could authenticate them on a smartphone, while Keystroke Dynamics could secure access to online accounts, and Voice Authentication could provide verification for smart speakers or home assistants. The scalability of the system enables deployment in a wide range of use cases, from consumer applications (e.g., e-commerce, mobile apps) to enterprise-level solutions (e.g., banking, healthcare).

The multi-modal authentication system has broad applicability across sectors that prioritize

security and user convenience. For example:

- Financial institutions could use this system to secure online transactions or grant access to sensitive financial information without relying on passwords or PINs.
- Healthcare organizations could implement it to authenticate medical professionals and patients accessing electronic health records (EHRs), ensuring compliance with privacy regulations like HIPAA.
- Consumer applications, such as smart homes, mobile devices, and e-commerce, would benefit from the frictionless and continuous authentication experience provided by the multi-modal system.

These sectors can use the multi-modal system to improve security and enhance the user experience, eliminating the need for passwords and other authentication mechanisms.

As with any biometric system, privacy is a critical consideration when collecting and storing behavioral data. Unlike passwords, biometric data cannot be reset or changed once compromised, making it a valuable target for attackers. Therefore, it is essential to implement robust data protection measures to safeguard sensitive information and ensure user anonymity.

For the multi-modal system to be effective and trustworthy, it must prioritize data encryption, anonymization, and decentralized storage of biometric data. Additionally, obtaining explicit user consent before collecting biometric data and ensuring transparency regarding how the data will be used, stored, and protected is critical to maintaining user trust.

Emerging technologies like secure enclave or multi-party computation (MPC) hold promise for safeguarding biometric data. These technologies enable secure processing without exposing sensitive information, ensuring that even in the event of a data breach, the user's data remains secure.

Future Implementation

The integration of Gait Analysis, Keystroke Dynamics, Mouse Movements, and Voice Authentication into a multi-modal authentication system marks a significant leap forward in behavioral biometric security. While the current implementation demonstrates a high degree of accuracy, security, and user adaptability, the potential for further advancement remains considerable. The following discussion explores promising future directions, including improvements in real-time processing, expansion to include additional biometric modalities, commercialization potential across various sectors, and critical considerations for data privacy and protection.

Real-Time Processing Enhancements

One of the most critical aspects of deploying any biometric system at scale, particularly in commercial or high-security environments, is the capacity for real-time processing. In multi-modal systems, where authentication decisions depend on the rapid interpretation and fusion of inputs from disparate modalities, latency becomes a crucial performance metric. Enhancing real-time responsiveness without sacrificing precision or security is a key priority.

Data fusion in its current form requires significant computational power due to the high-dimensional nature of the input features—voice patterns, mouse trajectories, gait images, and keystroke timings. To improve responsiveness:

- **Advanced Feature Selection Algorithms** such as lightweight convolutional neural networks (e.g., MobileNet, EfficientNet) can expedite feature extraction without degrading model performance. These models, when tailored to specific modalities, can dramatically reduce processing overhead.
- **Edge Processing** represents another avenue of significant promise. Rather than funneling all user data to a centralized cloud environment, localized processing directly on devices can minimize network latency. For instance, a wearable device could pre-process gait features before transmitting them, or a smartphone could perform voice pattern analysis natively.

- **Smart Data Reduction** techniques like Principal Component Analysis (PCA) and autoencoders can help distill high-volume biometric data into core vectors, preserving discriminative power while significantly reducing the processing load.

Modern processors, particularly those used in mobile devices and edge hardware, support multithreading and parallel execution. Future implementations could incorporate systems where each biometric stream—gait, voice, mouse, keystroke—is analyzed concurrently, reducing bottlenecks. This parallel architecture could dramatically reduce time-to-decision, a crucial improvement in scenarios like smart lock systems, ATM interfaces, or real-time transaction approvals.

Further Multi-Modal Integrations

While the current system integrates four behavioral modalities, the future of biometric authentication points toward broader and more diverse integrations. Adding additional layers of biometric input not only enhances robustness but also accommodates more varied user contexts.

Integrating facial recognition - one of the most established biometric technologies—can enrich the system by leveraging visual facial features alongside dynamic behavior-based traits. The fusion of passive recognition (gait, mouse) with visual confirmation (facial cues) may provide added assurance in environments where certain behaviors cannot be captured, such as during stationary use.

However, the deployment of facial recognition must address environmental sensitivities such as poor lighting, occlusions, or facial changes over time. Furthermore, spoofing concerns, while mitigated in a multi-modal setting, necessitate anti-spoofing measures like liveness detection or depth verification.

Although physiological and more traditional in nature, fingerprints remain highly accurate when used correctly. When paired with behavioral data, they offer a hybrid authentication model, reinforcing confidence in identity verification during high-stakes transactions or entry into secure areas. Integration must consider hardware availability and ensure proper encryption and anonymization to maintain user trust.

Other potential candidates for future integration include:

- **Gesture Dynamics**, particularly useful for touchscreen or AR/VR interfaces, where how a user swipes or moves can serve as a behavioral signature.
- **Heart Rate Variability** and **electrocardiogram (ECG) patterns**, especially when sourced from wearables, offer physiological signals influenced by behavioral patterns, thus straddling both biometric categories.

These additional signals, when effectively processed and fused, can further harden the authentication system against circumvention and enhance its applicability across various digital contexts.

Commercialization Possibilities and Potential Applications

The deployment potential of a behavioral multi-modal authentication system spans across industries, offering distinct advantages in security, compliance, and user experience. As authentication becomes a cornerstone of digital trust, industries are actively seeking alternatives to passwords, OTPs, and static biometric checks.

Smartphones and wearables represent an ideal launchpad for commercial implementations. The ubiquity of sensors—accelerometers, microphones, touch interfaces—allows for seamless integration of gait, voice, keystroke, and mouse modalities without additional hardware. Continuous, passive verification can prevent device theft and reduce unauthorized usage while offering a frictionless experience for users.

In smart homes, the combination of gait and voice could provide secure yet intuitive access to home controls, adjusting lighting, temperature, or unlocking doors based on personalized behavioral patterns. This is particularly valuable in shared environments where differentiating between multiple users becomes critical.

Authentication in healthcare must meet strict compliance standards while remaining non-intrusive to workflows. Behavioral authentication can protect electronic health records (EHRs) without requiring repeated password entry. In patient contexts, gait and voice patterns can aid in confirming identity without physical contact—an essential feature in sterile or high-risk areas.

In telemedicine, verifying both patient and provider identities using keystroke and voice analysis ensures the integrity of remote consultations, prescriptions, and data access, enhancing both privacy and medical accountability.

Banks and financial institutions are exploring biometric verification to counter increasing fraud and identity theft. Behavioral biometrics provide a strong solution for continuous authentication, where a user remains verified not just at login, but throughout a session. This approach is particularly suited to online banking platforms, where transitions between screens, fields, and features can be monitored for behavioral consistency.

E-commerce platforms can adopt the system to prevent fraud by monitoring buyer behavior. Suspicious mouse paths or irregular typing rhythms could trigger re-authentication or fraud alerts, enhancing trust in digital transactions.

Privacy Concerns and Potential Solutions

As with any technology that involves personal data—especially immutable biometric traits—privacy must be a foundational consideration in both design and deployment.

Implementing anonymization ensures that biometric data cannot be traced back to individuals in the event of a breach. One effective method is tokenization, where raw biometric data is converted into pseudonymous identifiers stored separately from user metadata. Even if compromised, this structure prevents attackers from reverse-engineering identifiable traits.

All data must be encrypted at rest and during transmission. Standards such as AES-256 or even post-quantum encryption algorithms may become necessary as threat models evolve. Moreover, homomorphic encryption and Zero-Knowledge proofs offer promising ways to perform computations on encrypted data without needing to decrypt it first, maintaining confidentiality throughout the authentication process.

Shifting data storage and computation to user-controlled devices (e.g., smartphones, wearables) ensures that sensitive biometric information never leaves the device. This approach reduces reliance on centralized storage, a common point of vulnerability, and gives users more transparency and control over their personal data.

Blockchain technologies can introduce immutable ledgers to track data access, consent management, and usage logs. For example, users can cryptographically approve the use of their biometric traits for specific services, and all actions can be recorded for transparency and compliance with frameworks like GDPR or CCPA.

Conclusion

In an increasingly digital world where data breaches, identity theft, and unauthorized system access are becoming more prevalent, the need for reliable, seamless, and adaptive authentication systems has never been more urgent. As remote work, online transactions, and mobile connectivity continue to expand, so too does the attack surface for potential threats. Traditional authentication methods—particularly those that rely on passwords, PINs, or single biometric inputs—have proven themselves insufficient in the face of evolving security challenges. Against this backdrop, the multi-modal behavioral biometric authentication system presented in this research marks a substantial leap forward in the evolution of secure identity verification.

By integrating Gait Analysis, Keystroke Dynamics, Mouse Movement Patterns, and Voice Authentication into a unified framework, the proposed system offers a dynamic and context-aware approach to authentication. What sets this solution apart is its reliance on behavioral characteristics - attributes that are inherently difficult to forge, yet can be captured non-invasively and passively as users go about their normal activities. This paradigm shift, from static identifiers to dynamic behavioral profiling, represents a more natural, secure, and user-centric form of authentication, one that aligns more closely with the complexity of human behavior in digital environments.

This chapter synthesizes the key findings of the study and expands on its broader significance by exploring its holistic architecture, real-world applications, and long-term potential, while also acknowledging the technical, ethical, and social challenges that lie ahead.

Integrating Gait, Keystroke, Mouse, and Voice: A Comprehensive Approach to Authentication

The heart of this authentication model lies in its multi-modal architecture, which deliberately brings together four distinct behavioral biometric modalities - each of which offers a unique perspective on user identity. This integration is more than the sum of its parts; it is a carefully orchestrated fusion designed to mitigate the limitations of individual systems while amplifying their strengths.

Gait Analysis plays a foundational role, relying on the natural and typically unconscious mechanics of walking. This modality is inherently unobtrusive and difficult to imitate, making it ideal for passive and continuous authentication. The model achieved high classification scores, particularly among users with consistent gait patterns. It functions exceptionally well in scenarios that allow movement-based data collection, such as workplace corridors, mobile contexts, or smart home environments.

Keystroke Dynamics, on the other hand, offers a fine-grained analysis of typing rhythms, capturing nuances such as dwell time, flight time, and typing cadence. With an exceptionally high recall and AUC, it excels at minimizing false rejections. Importantly, it serves as a critical modality in keyboard-centric environments, including workstations, terminals, and online forms. Its strength lies in its sensitivity to the micro-patterns in user behavior, allowing it to detect even subtle deviations in typing style.

Mouse Movements further diversify the behavioral profile by examining how users interact with graphical interfaces. It captures subconscious motor behaviors through metrics like pointer velocity, acceleration, and trajectory. Because mouse movement is often dictated by muscle memory and habit, it provides unique insight into a user's interaction style. Its predictive power is particularly useful in desktop and web-based environments, complementing Keystroke Dynamics in passive monitoring systems.

Voice Authentication adds a crucial auditory dimension to the system. Leveraging features such as pitch, cadence, and frequency formants, it allows for speaker recognition even in less-than-ideal acoustic conditions. Deep learning models trained on MFCC-based spectrograms have shown strong performance, maintaining high F1-scores across varying

audio samples. This modality is particularly valuable in mobile and IoT environments, where voice assistants are becoming the primary interaction medium.

The fusion of these modalities was achieved through a weighted decision strategy based on cosine similarity, allowing the system to dynamically prioritize the most reliable biometric inputs based on context. This strategy ensured that the system could maintain high authentication accuracy even when one or more modalities were compromised or unavailable. In aggregate, this fusion not only improved security outcomes but also increased resilience, user convenience, and the system's adaptability to different use cases.

Importance of Multi-Modal Biometrics in Modern Security Paradigms

As digital infrastructure grows increasingly complex and interdependent, the inadequacies of conventional authentication systems have become starkly apparent. Passwords are easily phished or guessed, security tokens can be stolen or cloned, and even biometric modalities like facial recognition and fingerprint scanning can be bypassed with sufficient sophistication. These vulnerabilities call for a new kind of authentication—one that is continuous, adaptable, and context-aware.

This study's emphasis on behavioral biometrics introduces a solution well-suited to these demands. Unlike static identifiers, behavioral traits evolve subtly over time, are influenced by neurological and physiological factors, and are significantly harder to replicate. By monitoring how a user behaves rather than what they possess or remember, the system offers a fundamentally more secure and context-sensitive means of authentication.

One of the most significant benefits of this approach is continuous authentication. Traditional systems typically authenticate users only once per session, leaving the system vulnerable during the remainder of the interaction. This multi-modal system, however, enables ongoing verification through passive observation, ensuring that identity is validated in real time and that anomalies are flagged as they occur.

Equally important is the system's non-intrusive nature. Each of the four modalities used can be captured passively, without requiring any explicit input from the user. This

frictionless model of security aligns with modern user expectations—particularly in enterprise and mobile environments where users demand seamless experiences without compromising safety.

Spoofing resistance is another vital advantage. While a motivated attacker may be able to mimic a voice or fabricate a fingerprint, imitating the full spectrum of a user’s gait, typing style, mouse trajectory, and speech patterns in concert is extremely unlikely. This multi-layered security architecture adds depth and redundancy, making unauthorized access not only difficult but practically infeasible.

Moreover, the system requires no specialized hardware. Its reliance on ubiquitous devices such as microphones, webcams, keyboards, and trackpads ensures low cost of deployment and maximum accessibility. It can be rolled out across consumer electronics, enterprise systems, or even embedded into existing digital infrastructures with minimal disruption.

Real-World Applicability and Deployment Potential

The versatility of this system extends well beyond experimental environments, positioning it as a viable candidate for deployment across various industries and platforms. Its potential applications span both personal and professional spheres, making it an adaptable solution to a wide range of security challenges.

In enterprise settings, the system could be deployed on employee workstations or remote desktop environments, where it would continuously validate user identity throughout a session. For example, if an unauthorized individual attempts to continue a session initiated by an authenticated user, deviations in typing rhythm or mouse movement patterns would trigger an alert or automatic logout.

In online banking and e-commerce, where fraud prevention is paramount, behavioral biometrics offer an invaluable layer of continuous security. The system could verify a user at login using voice or keystrokes, then continue monitoring behavior throughout the transaction to detect any anomalies suggestive of a security breach.

Mobile applications stand to benefit as well. Smartphones equipped with motion sensors, microphones, and touchscreens can easily collect behavioral data such as gait and voice

patterns. These devices already serve as digital wallets and personal assistants; integrating passive behavioral authentication would strengthen their role as secure identity gateways.

In healthcare, behavioral biometrics could be employed to protect electronic health records (EHRs) or to authenticate participants in telemedicine consultations. The system's passive and continuous nature aligns well with the need for secure, uninterrupted access to sensitive medical information.

Public safety environments—such as airports, border checkpoints, or secure laboratories—could also benefit. In such scenarios, gait and voice analysis could begin identifying individuals before they reach a traditional checkpoint, allowing for earlier intervention or authentication.

Looking Forward: Challenges and Opportunities

Despite its considerable promise, the path to widespread adoption of behavioral multi-modal authentication is not without its challenges. From technical hurdles to ethical concerns, future implementations must navigate a complex landscape.

Behavioral data is inherently variable. Fatigue, stress, illness, and emotional states can all impact how someone walks, types, or speaks. Developing models that are tolerant of such fluctuations without compromising accuracy is a significant research challenge. Techniques such as adaptive learning, temporal smoothing, and personalized model calibration could offer viable solutions.

Real-time performance remains a bottleneck. While edge computing and hardware acceleration can reduce latency, high-dimensional fusion remains computationally demanding. Innovations such as lightweight neural networks, quantization, and model pruning will be essential for deployment on resource-constrained devices.

Privacy is perhaps the most pressing issue. Behavioral biometrics, by their nature, are deeply personal. Their use for continuous monitoring must be justified with strict adherence to data protection standards. Federated learning, where data remains on the user's device and only model updates are shared, offers one possible direction. Encryption, anonymization, and secure enclaves should also form part of any deployment strategy.

Consent and transparency must also be central to system design. Users must understand what data is being collected, how it is used, and what their rights are in terms of opting in or out. Trust, after all, is the foundation of any system that handles sensitive personal information.

Lastly, scalability poses infrastructure challenges. Ensuring consistent performance across different devices, network conditions, and user behaviors requires careful design of both the front-end user experience and the back-end processing pipelines.

Final Reflections

This research set out to tackle a fundamental problem in cybersecurity: how to authenticate users in a way that is both secure and unobtrusive. Through the strategic combination of Gait Analysis, Keystroke Dynamics, Mouse Movements, and Voice Authentication, it has delivered a system that is both technically advanced and grounded in the realities of user behavior.

The results speak volumes. High accuracy rates, low false acceptance and rejection rates, and strong performance across diverse datasets suggest that this is more than a proof of concept—it is a blueprint for the future of authentication.

What is most compelling, however, is the philosophical shift that this system represents. It moves us away from static identifiers and secret codes, toward a model where identity is something that is lived, observed, and dynamically verified. It is a model rooted not in what we possess or remember, but in how we move, speak, and interact—a model that recognizes us not just as users, but as unique behavioral beings.

Looking ahead, the goal will be to refine, scale, and deploy this system in ways that respect both technical boundaries and ethical responsibilities. Done correctly, this approach could redefine digital identity, making it more secure, more personal, and more attuned to the way we actually live and work.

In closing, this multi-modal system represents a powerful and necessary evolution in authentication. By combining the strengths of multiple behavioral modalities, it offers a robust, adaptive, and forward-thinking solution—one capable of meeting the challenges of

today while paving the way for the secure digital ecosystems of tomorrow.

References

- [1] M. Soltane and M. Bakhti, 'Multi-modal biometric authentications: concept issues and applications strategies,' *Journal of Advanced Science and Technology*, 2012.
- [2] G. Iwasokun, S. Udoh, and O.K. Akinyokun, 'Multi-modal biometrics: Applications, strategies and operations,' *Science and Technology*, 2015.
- [3] M.A. Kadir, 'Multimodal EEG and Keystroke Dynamics Based Biometric System Using Machine Learning Algorithms,' 2021.
- [4] S. Tiwari, R. Raja, and R.S. Wadawadagi, 'Emerging Biometric Modalities and Integration Challenges,' 2024.
- [5] A. Badade and R.K. Dhanaraj, 'A Comprehensive Study on Continuous Person Authentication Using Behavioral Biometrics,' 2024.
- [6] J. Prasad, 'Multi-Factor Authentication System Using Keystroke Dynamics and Biometric Face Recognition with Feature Optimization,' 2024.
- [7] I. Lamiche, G. Bin, Z. Yu, and A. Hadid, 'A continuous smartphone authentication method based on gait patterns and keystroke dynamics,' *Journal of Ambient Intelligence*, 2019.
- [8] A. Rahman, M.E.H. Chowdhury, and A. Khandakar, 'Robust biometric system using session invariant multimodal EEG and keystroke dynamics by the ensemble of self-ONNs,' *Computers in Biology*, 2022.
- [9] X. Zhang, L. Yao, C. Huang, T. Gu, and Z. Yang, 'DeepKey: A multimodal biometric authentication system via deep decoding gaits and brainwaves,' *Expert Systems with Applications*, 2020.
- [10] H. Aronowitz, M. Li, and O. Toledo-Ronen, 'Multi-modal biometrics for mobile authentication,' *IEEE Transactions on Biometrics*, 2014.
- [11] S. Dargan and M. Kumar, 'A comprehensive survey on the biometric recognition

systems based on physiological and behavioral modalities,' *Journal of Ambient Intelligence and Humanized Computing*, 2020.

[12] B.V.S. Mendes, 'Analysis of feature selection on the performance of multimodal keystroke dynamics biometric systems,' 2017.

[13] A. Ray-Dowling, D. Hou, and S. Schuckers, 'Evaluating multi-modal mobile behavioral biometrics using public datasets,' *Computers & Security*, 2022.

[14] M. Gavrilova, F. Ahmed, and A.S.M.H. Bari, 'Multi-modal motion-capture-based biometric systems for emergency response and patient rehabilitation,' 2021.

[15] W. Cheung, 'Multi-Modal User Authentication Using Biometrics,' 2021.

[16] M. Antona, E.G. Spanakis, 'Multi-modal User Interface Design for a Face and Voice Recognition Biometric Authentication System,' *MobiHealth*, 2017.

Sri Lanka Institute of Information Technology.docx

ORIGINALITY REPORT

10 %	7 %	7 %	4 %
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	1 %
2	R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P. Prasad. "Algorithms in Advanced Artificial Intelligence - Proceedings of International Conference on Algorithms in Advanced Artificial Intelligence (ICAAAI-2024)", CRC Press, 2025 Publication	<1 %
3	V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challenges in Information, Communication and Computing Technology", CRC Press, 2024 Publication	<1 %
4	fastercapital.com Internet Source	<1 %
5	www.coursehero.com Internet Source	<1 %
6	Shashi Kant Dargar, Shilpi Birla, Abha Dargar, Avtar Singh, D. Ganeshaperumal. "Sustainable Materials and Technologies in VLSI and Information Processing - Proceedings of the 1st International Conference on Sustainable Materials and Technologies in VLSI and Information Processing (SMTVIP, 2024), December 13-14, 2024, Virudhunagar, India", CRC Press, 2025 Publication	<1 %