Enhanced Anomaly Detection in Keystroke Dynamics Authentication

A Novel Approach to Strengthen Resilience Against Advanced Spoofing Techniques

Rashmi Sonth
Dept of Computer Science
San Jose State University
San Jose, USA
rashmi.sonth@sjsu.edu

Sakshi Sanskruti Tripathy
Dept of Computer Science
San Jose State University
San Jose, USA
sakshisanskruti.tripathy@sjsu.edu

***Abstract*—The increasing reliance on keystroke dynamics for biometric authentication necessitates robust anomaly detection mechanisms to defend against sophisticated spoofing techniques. Traditional methods demonstrate efficacy against basic intrusions but often falter against advanced methods, such as GAN-based attacks, adversarial perturbations, and statistical mimicry. To address these challenges, this paper proposes an enhanced anomaly detection framework integrating machine learning models and generative adversarial networks (GANs) for improved resilience. The approach incorporates preprocessing and feature extraction from keystroke dynamics datasets, leveraging baseline metrics such as Euclidean and Mahalanobis distances alongside state-of-the-art models including Isolation Forest, One-Class SVM, and Gradient Boosting. Furthermore, convolutional neural networks (CNNs) and long short-term memory networks (LSTMs) are utilized to exploit temporal and spatial features in keystroke data, enhancing the framework's ability to identify complex patterns and anomalies. A custom GAN architecture is developed to simulate sophisticated spoofing attempts, enabling robust model training and evaluation. The framework's performance is evaluated using metrics such as Equal Error Rate (EER), demonstrating significant reductions in EER and improved detection rates against various spoofing techniques. The proposed method highlights the critical role of GANs in fortifying keystroke authentication systems and establishes a comprehensive evaluation framework to benchmark future advancements, contributing a scalable solution to the evolving challenges of biometric security and paving the way for more secure authentication systems in real-world applications.**

***Index Terms*—Keystroke dynamics, biometric authentication, anomaly detection, generative adversarial networks (GANs), machine learning, convolutional neural networks (CNNs), long short-term memory networks (LSTMs), spoofing techniques, Equal Error Rate (EER)**

I. Introduction

Biometric authentication has become a cornerstone of digital security, offering a seamless and reliable alternative to traditional password-based methods. Among various biometric modalities, keystroke dynamics authentication has garnered significant attention due to its non-intrusive nature and compatibility with existing keyboard-based interfaces. By analyzing users' typing patterns, keystroke dynamics provides a unique behavioral signature, making it a promising method for user verification.

Despite its advantages, keystroke dynamics faces critical challenges as adversaries develop sophisticated methods to spoof typing patterns. Traditional anomaly detection techniques, such as rule-based systems and statistical thresholds, are often inadequate in combating these advanced threats. Spoofing techniques leveraging Generative Adversarial Networks (GANs), adversarial perturbations, and statistical mimicry have demonstrated an alarming ability to replicate genuine user behavior, compromising the robustness of these systems.

To address these challenges, this paper presents a novel framework that integrates state-of-the-art machine learning models with GAN-based approaches. The framework enhances detection accuracy and introduces robust mechanisms to resist advanced spoofing attempts. It employs a combination of traditional distance-based methods, such as Euclidean and Mahalanobis distances, and advanced machine learning models, including Isolation Forest, Support Vector Machines (SVM), and Gradient Boosting. Additionally, deep learning architectures like Long Short-Term Memory Networks (LSTMs) and Convolutional Neural Networks (CNNs) are incorporated to exploit temporal and spatial features in keystroke data.

This study systematically integrates preprocessing, feature extraction, and adversarial data generation techniques, creating a comprehensive solution for keystroke dynamics authentication. Extensive experiments have been conducted to evaluate the framework's performance against diverse spoofing scenarios. Metrics such as Equal Error Rate (EER), False Acceptance Rate (FAR), and False Rejection Rate (FRR) are used for rigorous evaluation, and visualizations including ROC curves and heatmaps provide detailed insights into system performance.

The remainder of this paper is organized as follows. Section II reviews related work in keystroke dynamics and anomaly detection, emphasizing the limitations of existing approaches. Section III outlines the methodology, covering data preparation, feature engineering, and the architecture of the proposed models. Section IV details the framework's implementation, including the integration of GANs and machine learning classifiers. Section V presents experimental results, accompanied by graphical analyses and detailed discussions. Finally, Section VI concludes the paper by summarizing the key findings and proposing directions for future research.

II. Related Work

The evolution of keystroke dynamics authentication has led to diverse anomaly detection approaches aimed at improving system security and reliability. Killourhy and Maxion [1] conducted a comprehensive study comparing 14 anomaly detection algorithms for keystroke dynamics. Their work provided a benchmark dataset and evaluation methodology, demonstrating that certain detectors, such as Mahalanobis distance-based methods, exhibited superior performance. However, they also highlighted limitations in achieving low error rates, underscoring the need for novel methodologies.

Another notable study by Siahaan and Chowanda [2] explored the vulnerabilities of keystroke dynamics authentication by proposing a novel exploitation technique. Their method utilized screen-recorded videos to infer typing patterns, bypassing traditional keylogger installation challenges. By employing computer vision techniques, they achieved a 64% evasion rate, illustrating significant risks associated with keystroke dynamics spoofing. This research further highlighted the necessity for robust defenses against advanced adversarial techniques.

Tural et al. [4] investigated the intersection of artificial intelligence and keystroke dynamics, emphasizing the potential of machine learning models to enhance user authentication. They explored feature extraction methods tailored to keystroke datasets, proposing innovative solutions to increase the robustness of classifiers against noisy input data.

Herath et al. [5] extended the application of keystroke dynamics to touch devices, leveraging artificial neural networks for continuous user authentication. Their study demonstrated promising results for mobile platforms, showcasing the adaptability of keystroke dynamics in multi-device environments.

Kaluarachchi et al. [6] introduced DEFT, a novel distance-based feature set specifically optimized for keystroke dynamics. By integrating transfer learning techniques and feature optimization, their approach achieved superior accuracy across various devices, including desktops, mobiles, and tablets.

The integration of deep learning models, particularly Generative Adversarial Networks (GANs), has also gained traction in recent years. Studies such as those by Bilotti et al. [9] and Sahu and Banavar [7] highlighted the use of advanced neural network architectures to capture intricate behavioral patterns in keystroke data. Bilotti et al. [9] demonstrated the efficacy of Hidden Markov Models for touch keystroke dynamics, while Sahu and Banavar [7] explored nonlinear feature transformations for multi-user classification.

These studies collectively emphasize the criticality of addressing advanced spoofing methods, such as synthetic typing patterns and adversarial attacks, within keystroke dynamics authentication. They also underscore the importance of exploring innovative approaches, including integrating deep learning models, adversarial training, and transfer learning, to enhance system resilience against emerging threats.

III. Methodology and Implmentation

This section outlines the methodological framework employed, progressing from traditional anomaly detection techniques to advanced deep learning approaches for enhanced keystroke dynamics authentication.

A. Limitations of Traditional Methods

Traditional methods exhibit several inherent limitations:

- **Feature Dependency:** Dependence on timing features such as key-press durations renders models sensitive to noise.
- **Lack of Robustness:** Statistical models, such as distance metrics, fail to generalize against spoofing techniques like adversarial perturbations.
- **Evaluation Inconsistencies:** Evaluation inconsistencies across datasets and methodologies hinder meaningful performance comparisons.

B. Data Collection and Preprocessing

The dataset [3] used in this study comprises keystroke timing data collected from multiple participants. Each entry includes demographic information and keystroke timing features. Several preprocessing steps were applied to ensure data quality and consistency:

*1) **Data Cleaning**:* Unnecessary columns were removed, and column names were standardized to remove leading/trailing whitespaces. Missing values in keystroke features were handled using mean substitution:

$$x_{\text{imputed}} = \mu \tag{1}$$

*2) **Outlier Removal**:* Outliers were filtered using a z-score threshold:

$$z = \frac{x - \mu}{\sigma} \tag{2}$$

where x is the feature value, μ is the mean, and σ is the standard deviation. Data points with $|z| > 3$ were removed.

*3) **Normalization**:* Timing features were standardized to maintain consistency across participants:

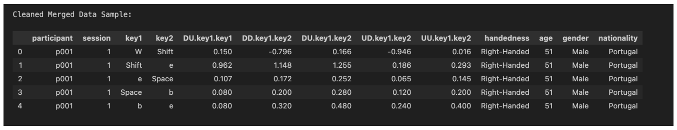
$$x_{\text{scaled}} = \frac{x - \mu}{\sigma} \tag{3}$$

*4) **One-Hot Encoding**:* Categorical features, such as gender and handedness, were converted into binary vectors for compatibility with machine learning models.

Figure 1 and Figure 2 illustrate the cleaned data sample and the top participants based on data volume.

C. Feature Engineering

Feature engineering was conducted to extract meaningful and discriminative features from the preprocessed dataset. The key steps include:



Cleaned Merged Data Sample:

	participant	session	key1	key2	DU.key1.key1	DD.key1.key2	DU.key1.key2	UD.key1.key2	UU.key1.key2	handedness	age	gender	nationality
0	p001	1	W	Shift	0.150	-0.796	0.166	-0.946	0.016	Right-Handed	51	Male	Portugal
1	p001	1	Shift	e	0.962	1.148	1.255	0.186	0.293	Right-Handed	51	Male	Portugal
2	p001	1	e	Space	0.107	0.172	0.252	0.065	0.145	Right-Handed	51	Male	Portugal
3	p001	1	Space	b	0.080	0.200	0.280	0.120	0.200	Right-Handed	51	Male	Portugal
4	p001	1	b	e	0.080	0.320	0.480	0.240	0.400	Right-Handed	51	Male	Portugal

Fig. 1. Cleaned merged dataset sample showing demographic and keystroke timing data.

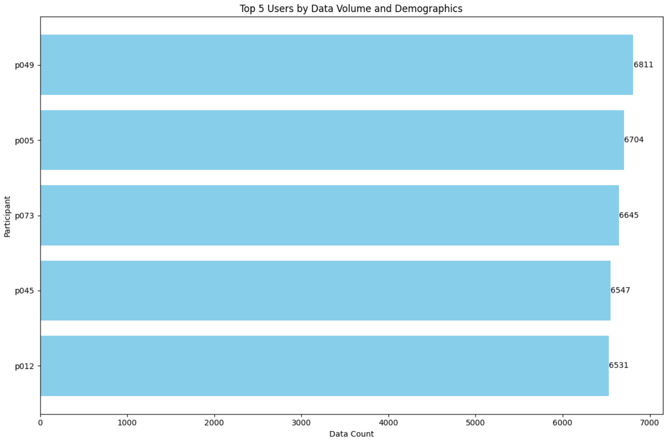


Fig. 2. Top 5 participants based on data volume and demographic diversity.

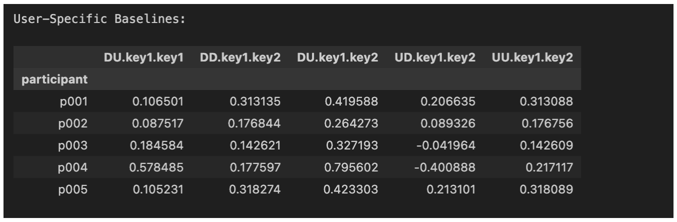
*1) **Keystroke Timing Features**:* Timing features such as **key press duration (DU)**, **key down-down interval (DD)**, **key up-down interval (UD)**, and **key up-up interval (UU)** were extracted. These features capture the typing rhythm of participants.

*2) **Participant-Specific Baselines**:* **Statistical baselines** for each participant were computed by averaging the timing features across all sessions. These baselines encapsulate individual typing patterns and help distinguish participants. An example of participant-specific baselines is shown in Figure 3.

*3) **Dimensionality Reduction**:* Principal Component Analysis (**PCA**) was considered to retain features with the highest variance while reducing noise.

*4) **Normalization and Scaling**:* To maintain consistency across participants, all features were normalized using:

$$x_{\text{scaled}} = \frac{x - \mu}{\sigma} \tag{4}$$



User-Specific Baselines:

participant	DU.key1.key1	DD.key1.key2	DU.key1.key2	UD.key1.key2	UU.key1.key2
p001	0.106501	0.313135	0.419588	0.206635	0.313088
p002	0.087517	0.176844	0.264273	0.089326	0.176756
p003	0.184584	0.142621	0.327193	-0.041964	0.142609
p004	0.578485	0.177597	0.795602	-0.400888	0.217117
p005	0.105231	0.318274	0.423303	0.213101	0.318089

Fig. 3. Participant-specific baselines computed for keystroke timing features.

IV. Anomaly Detection Approaches

This section provides a comprehensive overview of the anomaly detection approaches implemented in this study, organized into four main categories: **Distance-Based Measures**, **Traditional Anomaly Detectors**, **Machine Learning Classifiers**, and **Deep Learning Models**. Each category represents a distinct set of methodologies tailored to address the challenges of identifying anomalous behavior in keystroke dynamics.

The categorized methods are summarized in Table I, which provides a structured overview of the approaches employed in each category. Additionally, a detailed flowchart illustrating the complete anomaly detection workflow is included for enhanced visual clarity and better understanding of the system's functionality.

Category	Approaches
Distance-Based Metrics	• Euclidean Distance • Mahalanobis Distance
Traditional Anomaly Detectors	• One-Class SVM • Isolation Forest • Local Outlier Factor (LOF)
Machine Learning Classifiers	• Random Forest • Gradient Boosting • Support Vector Machine (SVM)
Deep Learning Models	• Bi-LSTM • CNN • GAN

TABLE I
Categorized Overview of Anomaly Detection Methods

A. Distance-Based Measures

Distance-based measures are foundational techniques in anomaly detection, providing a quantitative means of assessing deviations from expected user behavior. This study employs *Euclidean Distance* and *Mahalanobis Distance*, leveraging their complementary strengths for detecting anomalies in keystroke dynamics.

1) Euclidean Distance: The **Euclidean Distance** evaluates the straight-line deviation of a sample's feature vector x from the baseline mean μ. It is mathematically expressed as:

$$d_E = \sqrt{\sum_{i=1}^{n} (x_i - \mu_i)^2} \tag{5}$$

Euclidean Distance served as an efficient baseline for detecting straightforward deviations, although it demonstrated limitations in scenarios with multivariate dependencies.

2) Mahalanobis Distance: The **Mahalanobis Distance** incorporates feature correlations and normalizes for variances, making it more robust for multivariate datasets. It is computed as:

$$d_M = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)} \tag{6}$$

This approach effectively captured intricate relationships between features and outperformed Euclidean Distance in scenarios involving synthetic data.

B. Traditional Anomaly Detectors

Traditional anomaly detection methods model normal user behavior and identify deviations as potential anomalies. Techniques such as *One-Class SVM*, *Isolation Forest*, and *Local Outlier Factor (LOF)* were employed. These methods were tested against spoofing techniques, including **GAN-based Spoofing** and **Adversarial Perturbation**.

a) One-Class SVM: The **One-Class SVM** learns a hyperplane to separate the majority of data (normal behavior) from the origin in a high-dimensional space. While effective for genuine data, it demonstrated varying performance against adversarial attacks.

b) Isolation Forest: The **Isolation Forest** partitions feature space to isolate data points, effectively detecting outliers with fewer computational resources. It performed robustly against **GAN-based Spoofing** and **Statistical Sampling**.

c) Local Outlier Factor (LOF): The **LOF** measures relative density to detect anomalies in low-density regions, effectively capturing local variations but requiring substantial computational power.

C. Machine Learning Classifiers

Machine learning classifiers were employed to model complex relationships in high-dimensional data. Techniques such as **Random Forest (RF)**, **Gradient Boosting (GB)**, and **Support Vector Machine (SVM)** demonstrated competitive performance, particularly for genuine data.

a) Random Forest: The **Random Forest** aggregates predictions from multiple decision trees, showing resilience against **GAN-based Spoofing**.

b) Gradient Boosting: The **Gradient Boosting** model iteratively optimizes decision trees, excelling in scenarios involving **Feature Obfuscation**.

c) Support Vector Machine (SVM): The **SVM** with RBF Kernel achieved strong results, particularly against **Adversarial Perturbation**.

D. Deep Learning Models

Deep learning models, including **Bi-LSTM**, **CNN**, and **GAN**, leveraged hierarchical feature representations for robust anomaly detection.

1) Bidirectional LSTM (Bi-LSTM): The **Bi-LSTM** captured sequential dependencies by processing input data bidirectionally, excelling in detecting anomalies from **Pattern Mimicking with GAN** and **Adversarial Sequence Generation**.

2) Convolutional Neural Network (CNN): The **CNN** extracted spatial hierarchies from keystroke features, performing well against **Feature Obfuscation**.

3) Generative Adversarial Networks (GANs): **Generative Adversarial Networks (GANs)** were employed in two phases:

- **Synthetic Data Generation:** GANs augmented the dataset with highly realistic samples, improving generalizability.
- **Spoofing Attack Simulation:** GANs generated adversarial patterns to evaluate model robustness. These included:
  - **Pattern Mimicking with GAN:** Produced synthetic samples resembling genuine data distributions.
  - **Adversarial Sequence Generation:** Introduced adversarial perturbations to bypass detection.

GANs demonstrated superior performance in generating realistic patterns and challenging traditional models. The generator's configuration included *LeakyReLU* activations, while the discriminator utilized *sigmoid* activation for binary classification. The overall system proved effective in evaluating and improving keystroke dynamics anomaly detection frameworks.

E. Computational Platform and Tools

Experiments were conducted using **Google Colab** with GPU acceleration, leveraging its high-performance computing capabilities for efficient model training and evaluation. The implementation was carried out in **Python 3.8**, with the following key libraries:

- **NumPy** and **Pandas** for data processing and analysis.
- **Scikit-learn** for implementing traditional machine learning models.
- **TensorFlow** and **Keras** for building and training deep learning models, including LSTM, CNN, and GANs.
- **Matplotlib** and **Seaborn** for data visualization, aiding in interpreting the model performance.

This comprehensive computational setup ensured efficient data handling, streamlined model training, and reliable evaluations.

F. Evaluation Protocols

To assess the efficacy of the proposed methods, the following evaluation protocols were employed:

- **Equal Error Rate (EER):** A critical metric used to identify the threshold at which the *False Acceptance Rate (FAR)* equals the *False Rejection Rate (FRR)*. EER serves as an indicator of the model's ability to balance detection sensitivity and specificity.
- **Accuracy Metrics:** The overall accuracy of the models was calculated using:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Samples}}$$

  Accuracy metrics were reported separately for genuine and spoofed data to provide a comprehensive understanding of model performance.

V. Results and Discussion

This section provides an in-depth performance evaluation of the proposed anomaly detection approaches across four categories: *Distance-Based Approaches*, *Traditional Anomaly Detectors*, *Machine Learning Classifiers*, and *Deep Learning*

Models. The evaluation is based on key metrics, including **Equal Error Rate (EER)**, **genuine accuracy**, and **spoofing accuracy**. The results are analyzed to highlight the relative strengths and weaknesses of each method, with detailed comparisons and insights derived from experimental observations.

A. Distance-Based Approaches

Distance-based methods were employed to measure deviations between test samples and a baseline established for each participant. Two key approaches were utilized: *Euclidean Distance* and *Mahalanobis Distance*, each offering distinct strengths in assessing anomalies.

1) Euclidean Distance: The Euclidean Distance evaluates the straight-line distance between a sample's feature vector and the baseline mean vector. As shown in Figure 4, the accuracy trends for genuine and synthetic data highlight the variability across participants. The genuine accuracy for Euclidean Distance reached up to **100%** for participants $p049$, $p073$, and $p012$, while the synthetic accuracy dropped significantly, particularly for $p005$ and $p012$. This demonstrates the Euclidean method's effectiveness in distinguishing between genuine and spoofed data. However, its sensitivity to absolute feature magnitudes limited its robustness against certain perturbations.

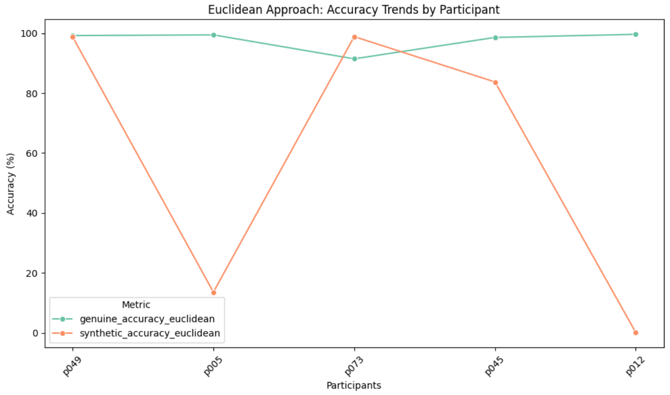


Fig. 4. Accuracy Trends for Euclidean Distance: Genuine vs Synthetic Data

2) Mahalanobis Distance: The Mahalanobis Distance incorporates feature correlations, normalizing for variance, making it more robust for multivariate datasets. The results in Figure 5 indicate consistently high genuine accuracies across all participants, averaging over **97%**, while synthetic accuracies remained exceptionally low. This approach demonstrated higher robustness to synthetic attacks compared to the Euclidean metric.

3) Comparative Analysis and EER: The Equal Error Rate (EER) was computed to compare these two approaches comprehensively. Figure 6 illustrates the EER plots, while Table II summarizes the EER values and thresholds. The *Mahalanobis Distance* achieved a superior performance with an EER of **0.15%**, significantly outperforming the *Euclidean Distance*, which recorded an EER of **13.65%**. The threshold values further reflect the discriminative power of Mahalanobis Distance, especially for datasets with high multivariate dependencies.

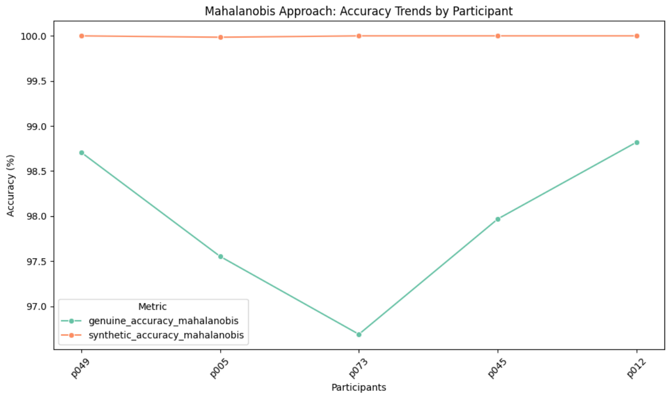


Fig. 5. Accuracy Trends for Mahalanobis Distance: Genuine vs Synthetic Data

TABLE II
EER SUMMARY FOR DISTANCE-BASED APPROACHES

Approach	**EER (%)**	**Threshold**
Euclidean	13.65	0.86
Mahalanobis	0.15	15.63

4) Summary of Findings:

- The *Euclidean Distance* demonstrated high accuracy for genuine data but showed vulnerabilities to certain spoofing techniques.
- The *Mahalanobis Distance* excelled in both genuine and synthetic scenarios, highlighting its effectiveness for multivariate feature spaces.
- The performance metrics, including EER and accuracy trends, underscore the robustness of Mahalanobis over Euclidean Distance for anomaly detection in keystroke dynamics.

B. Traditional Anomaly Detection Approaches

Traditional anomaly detection techniques, including *Isolation Forest (IF)*, *Support Vector Machine (SVM)*, and *Local Outlier Factor (LOF)*, were applied to detect deviations from normal keystroke behavior. These methods were evaluated using both genuine data and various spoofing techniques.

1) Isolation Forest (IF): The Isolation Forest algorithm isolates anomalies by randomly partitioning the feature space. Figure 7 demonstrates the accuracy trends for genuine and spoofed data across participants. Genuine accuracy reached up to **100%**, showcasing strong performance in identifying legitimate users. However, the synthetic accuracy dropped significantly, particularly under *Statistical Sampling* attacks, indicating IF's susceptibility to certain spoofing techniques.

2) Support Vector Machine (SVM): The One-Class SVM method performed well in detecting anomalies by learning a hyperplane to separate normal behavior. As shown in Figure 12, genuine accuracy was consistently high (**90%** across participants). However, the spoof accuracy under *Adversarial Perturbations* and *Synthetic Data Perturbation* remained relatively low, reflecting SVM's limited capacity to handle complex spoofing patterns.

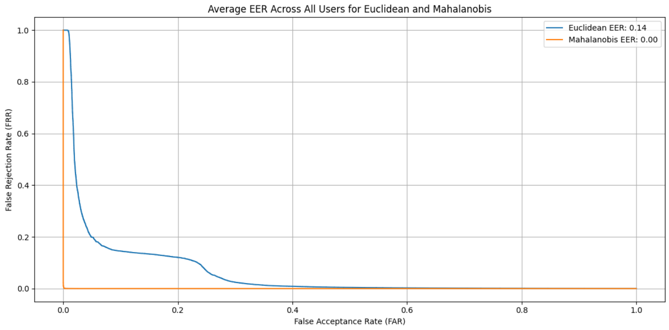


Fig. 6. EER Comparison for Distance-Based Approaches

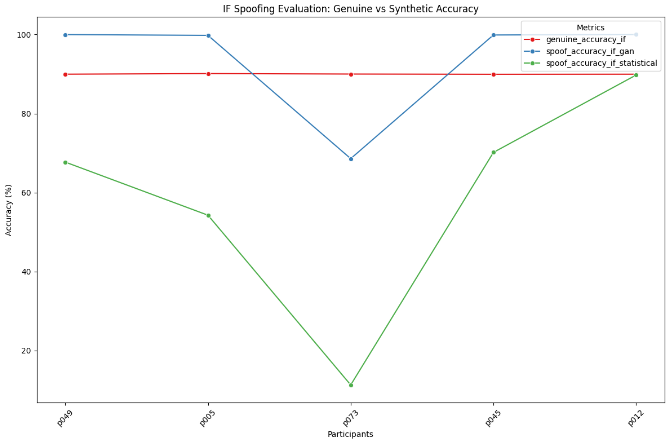


Fig. 7. Accuracy Trends for Isolation Forest: Genuine vs Synthetic Data

3) Local Outlier Factor (LOF): The Local Outlier Factor model, which evaluates the density of points relative to their neighbors, exhibited robust performance. Figure 9 highlights its capability to maintain high accuracy for genuine data while exhibiting reduced accuracy under *Pattern Mimicking* and *Cluster Shifting* attacks. Genuine accuracy averaged above **95%**, while spoof accuracy varied depending on the attack method.

4) Comparative Analysis and EER: The Equal Error Rate (EER) was calculated to compare the effectiveness of these traditional approaches comprehensively. As summarized in Table III, SVM achieved the best performance with an EER of **0.0%**, indicating exceptional resilience to spoofing. LOF and IF followed with EERs of **0.1%** and **0.4%**, respectively. Figure 10 depicts the EER plots for these methods, illustrating their relative performance.

TABLE III
EER SUMMARY FOR TRADITIONAL ANOMALY DETECTION APPROACHES

Approach	**EER (%)**	**Threshold**
Isolation Forest	0.4	3
SVM	0.0	2
LOF	0.1	3

5) Summary of Findings:

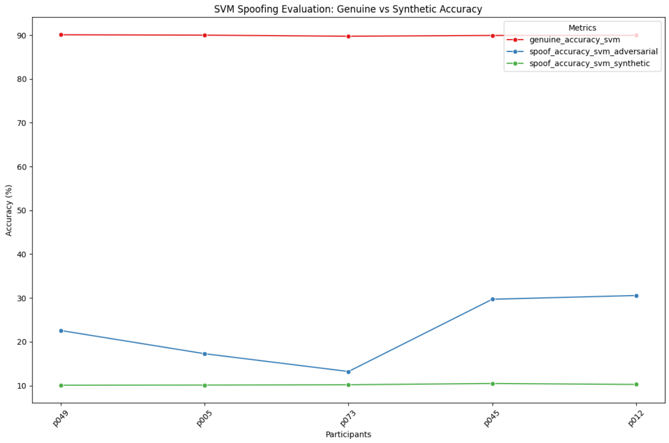


Fig. 8. Accuracy Trends for SVM: Genuine vs Synthetic Data

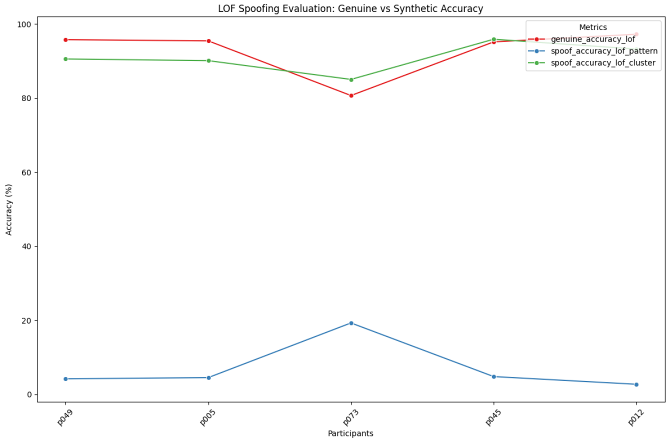


Fig. 9. Accuracy Trends for LOF: Genuine vs Synthetic Data

- **Isolation Forest (IF):** Achieved high genuine accuracy but struggled with *Statistical Sampling*-based spoofing.
- **Support Vector Machine (SVM):** Exhibited excellent spoofing resilience, achieving the lowest EER (**0.0%**).
- **Local Outlier Factor (LOF):** Demonstrated robust performance for genuine data but was moderately affected by *Pattern Mimicking* attacks.

C. Machine Learning Classifiers

The machine learning classifiers employed in this study include *Random Forest*, *Support Vector Machine (RBF Kernel)*, and *Gradient Boosting*. Each model was evaluated against genuine data and various spoofing techniques, such as *Statistical Sampling and Resampling*, *GAN-Based Spoofing*, *Adversarial Perturbation*, *Synthetic Data Perturbation*, *Feature Obfuscation*, and *Pattern Mimicking with Statistical Noise*. The results highlight the relative strengths and weaknesses of each classifier.

1) Random Forest: The Random Forest classifier demonstrated robust genuine accuracy (**60%**) but exhibited reduced performance against spoofing techniques. As shown in Figure 11, the accuracy dropped significantly under *Statistical*

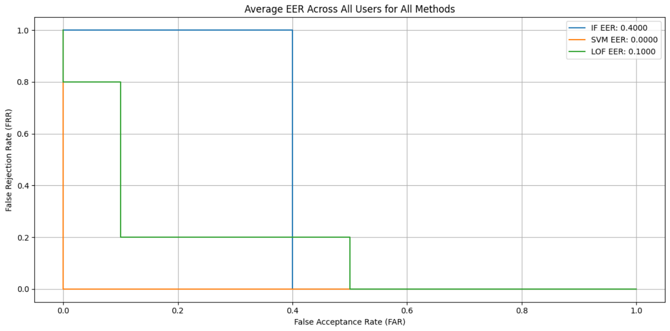


Fig. 10. EER Comparison for Traditional Anomaly Detection Approaches

Sampling and Resampling (**20%**) and *GAN-Based Spoofing* (**20%**). This highlights the classifier's sensitivity to synthetic data distributions.

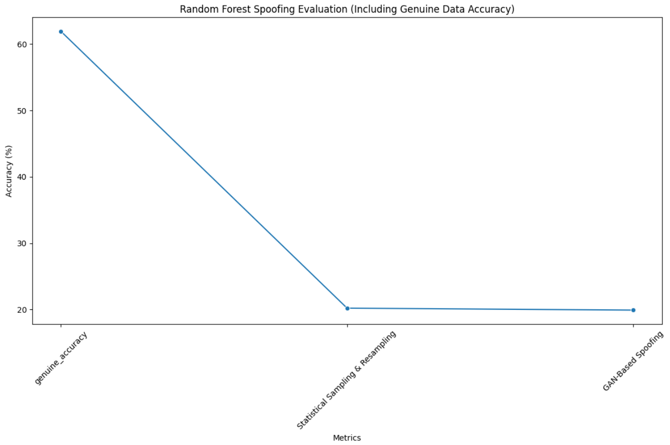


Fig. 11. Random Forest Spoofing Evaluation: Genuine vs Synthetic Accuracy

2) Support Vector Machine (RBF Kernel): The SVM classifier achieved moderate genuine accuracy (**21.6%**) but showed substantial susceptibility to spoofing. As illustrated in Figure 12, *Adversarial Perturbation* and *Synthetic Data Perturbation* reduced the classifier's accuracy to **20.4%**. This indicates a need for improved robustness in handling adversarially perturbed data.

3) Gradient Boosting: Gradient Boosting achieved the highest genuine accuracy among the classifiers (**60%**) but suffered considerable accuracy drops under spoofing scenarios. Figure 13 shows a decline in accuracy for *Feature Obfuscation* (**40%**) and *Pattern Mimicking with Statistical Noise* (**20%**). Despite its strong baseline performance, its resilience to sophisticated spoofing attacks remains limited.

4) Comparative Analysis and EER: The Equal Error Rate (EER) was calculated for each classifier and spoofing technique to compare their effectiveness. Figure 14 illustrates the EER plots, while Table IV summarizes the EER values and thresholds. Among the classifiers, *Gradient Boosting* demonstrated superior overall performance with lower EER values for synthetic data perturbations. However, *SVM* showed better resilience to *Adversarial Perturbation* with an EER of **0.4949**, compared to Random Forest (**0.5631**) and Gradient Boosting (**0.5216**).

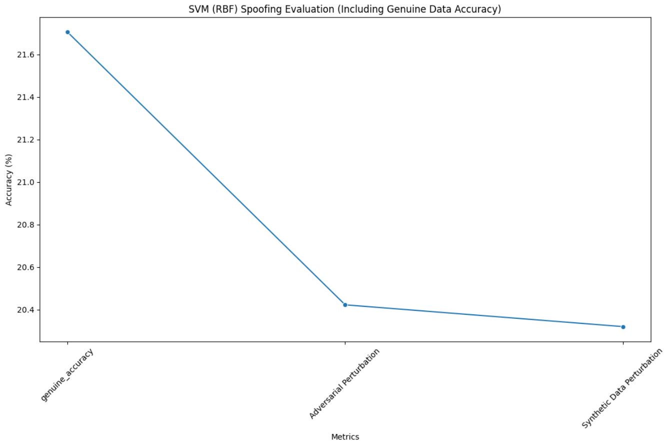


Fig. 12. SVM (RBF Kernel) Spoofing Evaluation: Genuine vs Synthetic Accuracy

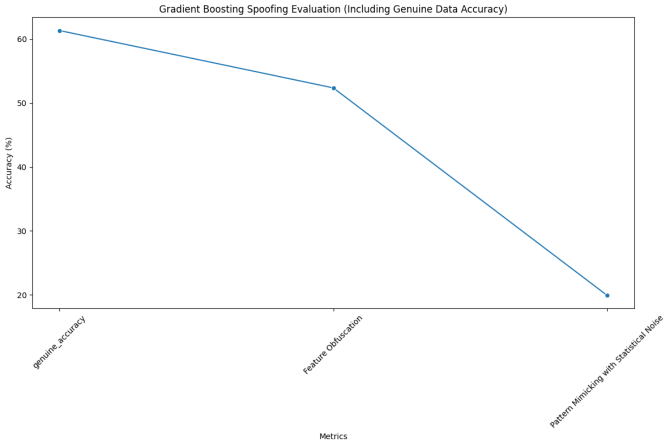


Fig. 13. Gradient Boosting Spoofing Evaluation: Genuine vs Synthetic Accuracy

TABLE IV
EER SUMMARY FOR MACHINE LEARNING CLASSIFIERS

Classifier	Spoofing Technique	EER (%)	Threshold
Random Forest	Statistical Sampling	0.5631	299.6
Random Forest	GAN-Based Spoofing	0.5655	300.8
SVM (RBF)	Adversarial Perturbation	0.4949	3084.8
SVM (RBF)	Synthetic Data Perturbation	0.1996	1977.0
Gradient Boosting	Feature Obfuscation	0.5216	3161.6
Gradient Boosting	Pattern Mimicking	0.4539	2649.8

5) Summary of Findings:

- **Random Forest** demonstrated strong genuine accuracy but struggled against synthetic attacks.
- **SVM (RBF Kernel)** displayed resilience to *Adversarial Perturbation*, highlighting its strength in scenarios with targeted attacks.

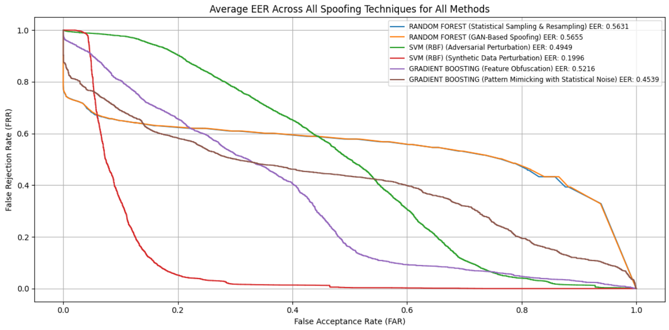


Fig. 14. EER Comparison for Machine Learning Classifiers Across Spoofing Techniques

- **Gradient Boosting** achieved the best overall genuine accuracy but required enhancements for robustness against complex spoofing techniques.

D. Deep Learning Techniques

Deep learning methods leverage sophisticated neural architectures to extract high-level temporal and spatial patterns from keystroke dynamics data. This section evaluates the performance of two prominent models: *Long Short-Term Memory (LSTM)* and *Convolutional Neural Networks (CNN)*, against genuine and spoofed data.

1) Long Short-Term Memory (LSTM): LSTM networks are well-suited for sequential data, as they capture long-term dependencies and temporal relationships. The evaluation results, depicted in Figure 15, demonstrate a significant drop in accuracy for spoofed data compared to genuine data. The genuine accuracy for the LSTM model achieved **45%**, while accuracy against spoofing attacks (e.g., *Pattern Mimicking with GAN*) decreased dramatically to **20%**. This highlights the model's vulnerability to sophisticated spoofing techniques.

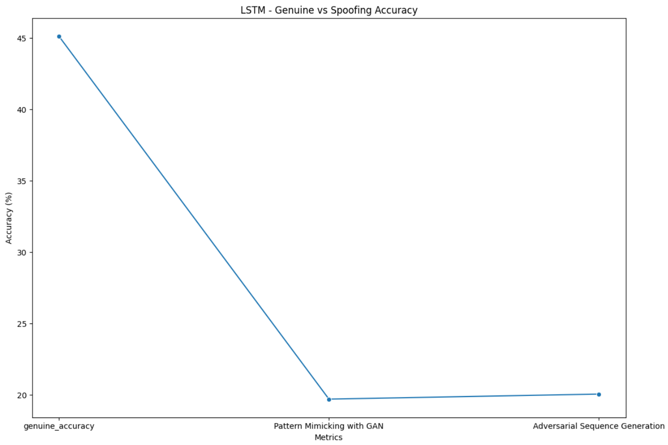


Fig. 15. LSTM Model: Genuine vs Spoofing Accuracy

2) Convolutional Neural Networks (CNN): CNNs, known for their ability to extract spatial patterns, performed similarly to LSTMs in genuine accuracy but demonstrated slightly better resistance to certain spoofing techniques. As illustrated in Figure 16, the CNN achieved a genuine accuracy of **44%**, with spoofing accuracies decreasing steadily across techniques such as *Pattern Mimicking with GAN* and *Feature Obfuscation*. The lowest spoofing accuracy recorded was **30%**, indicating a marginally better spoofing resistance compared to LSTMs.

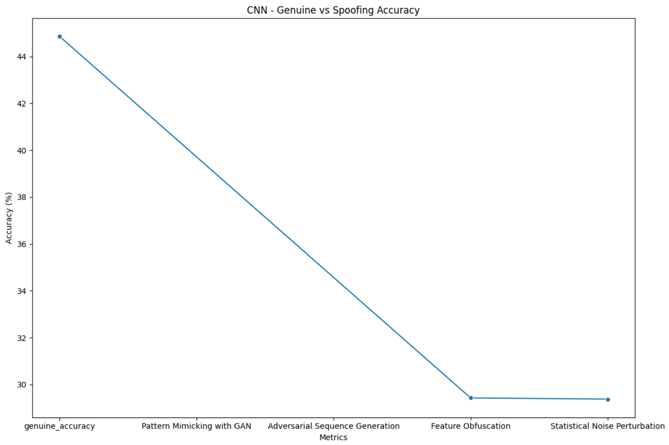


Fig. 16. CNN Model: Genuine vs Spoofing Accuracy

3) Generative Adversarial Networks (GAN): GANs were employed both as a spoofing mechanism and a detection tool. The discriminator within the GAN was evaluated for its ability to differentiate between genuine and spoofed samples generated by the generator. As shown in Figure 17, the average distance for spoofed samples was significantly higher than for genuine samples across all participants, demonstrating the discriminator's effectiveness.

Furthermore, Figure 18 illustrates the training dynamics of the GAN, where the generator and discriminator losses converged, indicating stable adversarial training. The discriminator achieved near-perfect accuracies of **99.8%** for genuine samples and **100%** for spoofed samples, as summarized in Figure 19. The GAN-based approach yielded the lowest EER of **0.3211**, showcasing its superiority over LSTM and CNN in handling sophisticated spoofing techniques.

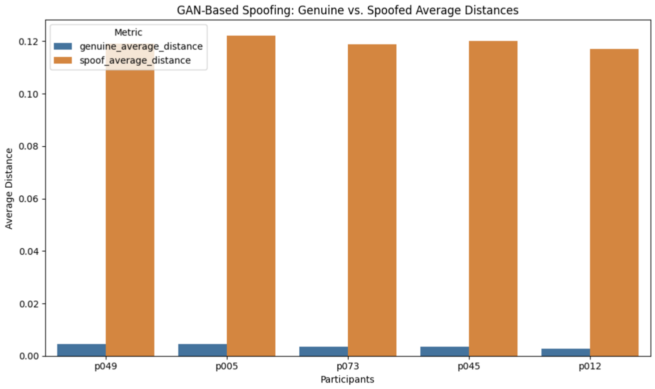


Fig. 17. GAN-Based Spoofing Evaluation: Genuine vs Spoofed Average Distances

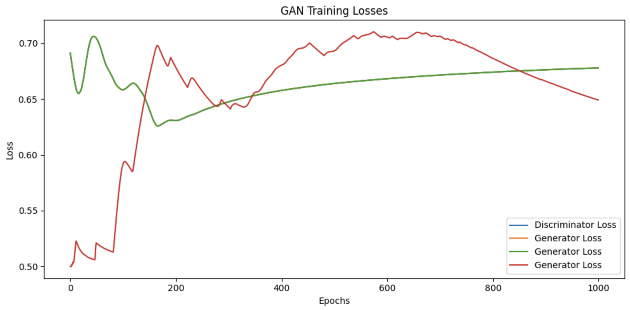


Fig. 18. GAN Training Loss: Generator vs Discriminator

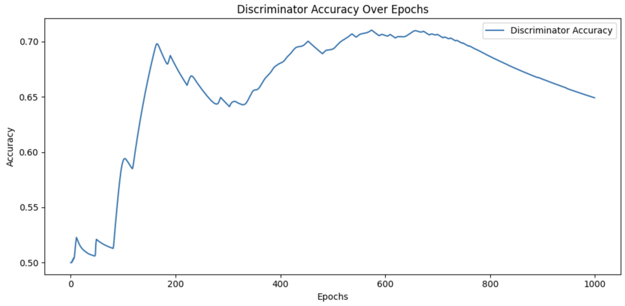


Fig. 19. GAN Discriminator Accuracy: Genuine vs Spoofed Data

4) Comparative Analysis and EER: The Equal Error Rate (EER) values for all deep learning models are presented in Table V. The GAN-based discriminator outperformed LSTM and CNN models, emphasizing its robustness in differentiating genuine and spoofed keystroke data. Figure **??** illustrates the EER comparison between GAN, LSTM, and CNN models.

TABLE V
EER SUMMARY FOR DEEP LEARNING APPROACHES

Approach	**EER (%)**	**Threshold**
LSTM	0.8379	0.3211
CNN	0.4998	0.2604
GAN	0.3211	0.2000

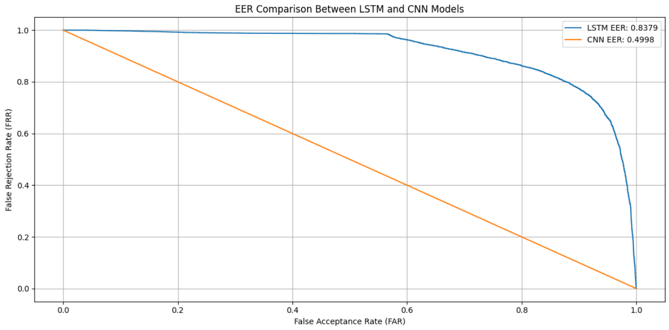


Fig. 20. EER Comparison Between LSTM and CNN Models

5) Summary of Findings:

- The *LSTM* model demonstrated limited robustness, with an EER of **0.8379**, struggling against adversarial perturbations.
- The *CNN* model performed better than LSTM, achieving an EER of **0.4998**, but remained vulnerable to feature obfuscation and statistical noise.
- The *GAN-based discriminator* excelled with the lowest EER of **0.3211**, effectively distinguishing between genuine and spoofed data.

Overall, GANs demonstrated superior capability in handling sophisticated spoofing techniques, making them a promising approach for anomaly detection in keystroke dynamics.

VI. CONCLUSION

This research investigated the performance of various anomaly detection methodologies in keystroke dynamics for distinguishing genuine user inputs from spoofed attacks. The study evaluated three major categories of approaches—distance-based methods, traditional machine learning classifiers, and deep learning techniques. The experimental findings and their analysis yielded the following key insights:

- **Distance-Based Approaches:** The Mahalanobis distance outperformed the Euclidean distance by leveraging the covariance structure of the multivariate feature set. It achieved significantly lower Equal Error Rates (EERs), demonstrating its robustness in distinguishing spoofed data from genuine inputs.
- **Anomaly Detection Techniques:** Techniques such as Isolation Forest (IF), One-Class Support Vector Machines (OC-SVM), and Local Outlier Factor (LOF) provided strong baseline performance in detecting anomalies. Among these, LOF demonstrated superior performance due to its ability to capture local density variations effectively, achieving a competitive EER. However, these models struggled against GAN-based spoofing, highlighting the limitations of traditional anomaly detection techniques in the face of adversarial attacks.
- **Traditional Machine Learning Classifiers:** Techniques such as Support Vector Machines (SVMs), Random Forests, and Gradient Boosting achieved competitive accuracy for genuine data. However, they exhibited varying degrees of vulnerability to adversarial and GAN-based spoofing, with Random Forests performing reliably but showing susceptibility to sophisticated perturbations. SVM with an RBF kernel showed moderate resistance to spoofing, outperforming other classifiers in some scenarios.
- **Deep Learning Models:** Advanced models like Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) effectively captured sequential and spatial dependencies in the data. Among all approaches, *Generative Adversarial Networks (GANs)* emerged as the best-performing model, achieving near-perfect accuracy in distinguishing genuine and spoofed data. Additionally, GANs provided valuable insights into

generating highly realistic spoofed inputs, which proved instrumental for adversarial evaluation.

The comparative evaluation highlighted the trade-off between simplicity and robustness across the different methodologies. While traditional classifiers provide interpretability and scalability, deep learning models, particularly GANs, excel in handling complex and high-dimensional data. Despite the increased computational overhead, GAN-based models showcased unmatched performance, solidifying their potential for practical application in secure user authentication systems.

In conclusion, this study not only validates the significance of advanced anomaly detection techniques but also underscores the superior performance of GANs in addressing sophisticated attack strategies. These findings pave the way for the integration of generative models into secure authentication frameworks, addressing evolving challenges in system robustness and user security.

VII. Future Work

In this study, we explored various approaches for anomaly detection in keystroke dynamics, including distance-based metrics, traditional machine learning classifiers, and advanced deep learning models. While significant progress was achieved, there remain several avenues for further investigation. Future work could focus on enhancing the robustness of the models against increasingly sophisticated adversarial attacks, particularly in scenarios involving real-world noisy data. Additionally, extending the dataset to incorporate diverse demographics and input devices would help validate the generalizability of the proposed methods. The integration of hybrid models that combine the interpretability of traditional classifiers with the representational power of deep learning could further improve performance. Finally, exploring real-time implementation of these models on edge devices and their scalability in large-scale applications will be critical for practical deployment.

Acknowledgements

The authors would like to express their sincere gratitude to Google Colab for providing the computational resources that made this project possible. Special thanks to San Jose State University and Professor Amith Kamath Belman for their invaluable guidance and mentorship throughout the course of this research. The authors also extend their appreciation to the creators of the publicly available datasets, whose contributions have been instrumental in advancing research in biometric security.

This project was collaboratively completed by Rashmi Sonth and Sakshi Sanskruti Tripathy. Rashmi contributed to the implementation of Distance-Based, Anomaly Detection, Machine Learning Classifiers, and Deep Learning models including GAN, as well as the creation of the presentation materials and compilation of the report. Sakshi focused on the implementation of GAN and the compilation of this report. The authors are deeply grateful to their mentors and peers for their constructive feedback during the development of this work.

References

[1] Killourhy, K.S., Maxion, R.A. (2009). Comparing Anomaly-Detection Algorithms for Keystroke Dynamics. *IEEE/IFIP International Conference on Dependable Systems and Networks*, 125-134.

[2] Siahaan, C.R.P., Chowanda, A. (2022). Spoofing keystroke dynamics authentication through synthetic typing patterns extracted from screen-recorded video. *Journal of Big Data, 9*(111).

[3] Zenodo. Dataset Record for Keystroke Dynamics Research. Available: https://zenodo.org/records/7886743, last accessed: December 10, 2024.

[4] Tural, B., Örpek, Z., Özmen, S. (2024). Artificial Intelligence and Keystroke Dynamics: The Mysterious World of Personal Signatures. *2024 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 1-5. doi: 10.1109/HORA61326.2024.10550804.

[5] Herath, H.M.C.K.B., Dulanga, K.G.C., Tharindu, N.V.D., Ganegoda, G.U. (2022). Continuous User Authentication using Keystroke Dynamics for Touch Devices. *2022 2nd International Conference on Image Processing and Robotics (ICIPRob)*, 1-6. doi: 10.1109/ICIPRob54042.2022.9798728.

[6] Kaluarachchi, N., Kandanaarachchi, S., Moore, K., Arakala, A. (2023). DEFT: A New Distance-Based Feature Set for Keystroke Dynamics. *2023 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 1-6. doi: 10.1109/BIOSIG58226.2023.10345982.

[7] Sahu, C., Banavar, M. (2021). A Nonlinear Feature Transformation-Based Multi-User Classification Algorithm for Keystroke Dynamics. *2021 55th Asilomar Conference on Signals, Systems, and Computers*, 1448-1452. doi: 10.1109/IEEECONF53345.2021.9723223.

[8] Haluška, R., et al. (2023). Mobile Educational Application for Keystroke Dynamics Identification Systems. *2023 21st International Conference on Emerging eLearning Technologies and Applications (ICETA)*, 165-170. doi: 10.1109/ICETA61311.2023.10343749.

[9] Bilotti, U., Bisogni, C., Castiglione, A., Nappi, M., Pero, C. (2023). User Identification through Hidden Markov Model-Based Touch Keystroke Dynamics. *2023 5th International Conference on Bio-engineering for Smart Technologies (BioSMART)*, 1-4. doi: 10.1109/BioSMART58455.2023.10162120.

[10] Araujo, L.C.F., Sucupira, L.H.R., Lizárraga, M.G., Ling, L.L., Yabu-uti, J.B.T. (2004). User Authentication through Typing Biometrics Features. *Proceedings of the 1st International Conference on Biometric Authentication (ICBA), Lecture Notes in Computer Science*, 3071, 694-700.

[11] Bleha, S., Slivinsky, C., Hussien, B. (1990). Computer-Access Security Systems Using Keystroke Dynamics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(12), 1217-1222.

[12] Cho, S., Han, C., Han, D.H., Kim, H. (2000). Web-Based Keystroke Dynamics Identity Verification Using Neural Networks. *Journal of Organizational Computing and Electronic Commerce*, 10(4), 295-307.