

Machine Learning Approaches for Surveilling Keyboard Activity on Mobile Devices through Acoustic and Inertial Sensors

Nirasha Kulsooriya^{1*}

*Department of Information and Communication Technology
University of Sri Jayewardenepura
Homagama, Sri Lanka
nirashakulsooriya@sjp.ac.lk*

Asanka Sayakkara²

*Department of Computation and Intelligent Systems
University of Colombo School of Computing
Colombo, Sri Lanka
asa@ucsc.cmb.ac.lk*

Abstract—This study investigates keystroke detection on mobile devices using multimodal sensor data, combining acoustic signals with inertial sensor readings taken from an accelerometer and a gyroscope while the device operates in fully silent mode. The primary objective was to evaluate the feasibility of uniquely identifying keypresses when a background application has access only to microphone and motion sensor permissions, which are commonly allowed without raising user suspicion. To achieve this, an end-to-end pipeline was developed, including signal preprocessing, feature extraction, and machine learning models for classification of keypresses typed on a virtual keyboard. An integrated dataset was created, including audio recordings in .mp3 format with 30 keypress samples for each alphabet letter and the space key. To address background noise and minor distortions in the raw recordings, several preprocessing techniques such as normalization, trimming, filtering, and segmentation were applied to enhance signal quality. Initially, audio-only models achieved limited performance, with Random Forest yielding 21% accuracy. After refining preprocessing steps, hyperparameter tuning, and model optimization, the accuracy improved significantly to 94% for audio-based classification. Models trained on inertial sensor data reached up to 80% accuracy, demonstrating the potential of motion-based features in keystroke identification. These results highlight the complementary strengths of acoustic and motion data and underscore the security risks of sensor-based side-channel attacks, emphasising threats to user input confidentiality, device integrity, and privacy protection.

Index Terms—keystroke, acoustic signals, inertial sensor, segmentation, random forest

I. INTRODUCTION

Smartphones have become an integral part of modern life, permeating almost every aspect of personal and professional activities. According to recent statistics [1], over 5.44 billion people worldwide approximately 68% of the global population use smartphones, a trend that has significantly transformed personal communication as well as sectors such as healthcare, education, and commerce. The proliferation of affordable smartphones has accelerated adoption in developing regions, helping to bridge the digital divide and enhance global connectivity. Mobile data traffic has also surged as a result, with the report predicting a fivefold increase by

2027 [2]. Furthermore, the rise of mobile applications has driven the expansion of the digital economy, as app-based services such as mobile banking, online shopping, and digital health solutions become ubiquitous. The global mobile app market alone was valued at over \$200 billion in 2022 and is projected to continue growing exponentially [2].

The trends in the smartphone market reveal a rapidly evolving landscape driven by technological advancements, changing consumer preferences, and global economic forces. In recent years, the smartphone market has experienced a noticeable shift toward 5G technology, with an estimated 1.5 billion 5 G-enabled devices expected to be in use globally by 2025. This transition has spurred innovation across industries, as faster internet speeds and lower latency enable new applications, such as augmented reality (AR), virtual reality (VR), and advanced mobile gaming experiences. Alongside 5G, the market is witnessing a surge in demand for foldable smartphones, which offer consumers larger screen sizes without sacrificing portability.

The security concerns in mobile devices have become a critical issue in today's increasingly connected world. As smartphones and tablets are now ubiquitous, storing vast amounts of personal and professional data, they have become prime targets for cybercriminals. According to recent studies [3], [4], mobile malware attacks increased by over 50% in 2022, with hackers exploiting vulnerabilities in operating systems, apps, and even hardware to gain unauthorized access to sensitive information.

A. Background

1) *Mobile Devices and Security*: Mobile devices are now essential for communication and accessing sensitive data, but their growing capabilities increase exposure to security threats. While encryption and authentication protect software and networks, these measures are often insufficient against sophisticated attacks exploiting hardware and physical

vulnerabilities.

2) *Inertial Sensors*: Modern smartphones and wearables use embedded sensors, especially accelerometers and gyroscopes, to enable motion detection and advanced functionalities. While these sensors enhance user experience, their typically unrestricted access creates security and privacy risks, making them potential targets for unintended data leakage or malicious exploitation.

3) *Side Channel Attacks*: Side-channel attacks are a class of security threats that exploit indirect information leakage from a system's physical implementation rather than directly attacking its software or cryptographic algorithms. Side-channel attacks exploit unintended signals, such as acoustic emissions or inertial sensor readings, to infer sensitive information like keystrokes, passwords, and other confidential user inputs on mobile devices as smartphones and wearables become increasingly sensor-rich and integrated into daily activities, understanding the feasibility, mechanisms, and potential impact of such attacks is essential for designing more secure and resilient mobile platforms. These studies highlight the need for improved protective measures beyond conventional software and network security.

As mobile devices are appearing in almost all daily activities, users are frequently engaging with software/app installation and uninstallation processes. While installing an app on our devices, we used to allow the permissions for several components in Micro-Electro-Mechanical System (MEMS) and other integrated components. In a moment of utilising an app (i.e. a malware) which has already gained permission only for the microphone and the inertial sensors, there may be a security concern in uncovering the keypresses by keeping track of inertial sensor readings and the audio signal emanating with each keypress in the device's fully silent mode. Based on these assumptions, the study has been conducted.

II. RELATED WORKS

An acoustic side-channel attack is one of the most prominent fields of study in academia nowadays. Everywhere we are dealing with many systems, most of which are powered by AI. So, each individual now interacts with AI platforms to accomplish daily tasks without fully considering the potential drawbacks or vulnerabilities associated with their use, unlike the more cautious approach followed in the past.

Balgani K.S. et. al [5] have conducted a study to detect acoustic side-channel attacks while entering a Personal Identification Number (PIN) or password into the ATM for money withdrawal. Here, 58 participants have used two commercially available metal keypads to try out the PINs with 4 - 5 digits, while money withdrawal with 3 attempts by analyzing two considerable distances setting up external microphones for better accuracy and efficiency. In line with

the accuracies they have obtained [5], the better robustness could be seen as 93% and 95% for PIN 4 and 5-digits entering within a distance of 0.3 meters. Meanwhile, the PinDrop system has performed with 44% accuracy overall in any PIN entry while maintaining the 2 meter distance.

Mamta B. et al [6] have examined an elementary review on keyboard acoustic side-channel attacks on Android phones in the AI era of digital banking. They thoroughly analyzed the safety risks and holes in the context of online banking with the side-channel attacks that are possible and identical to the context addressed previously.

In 2019, an observation was conducted [7] on recovering the acoustic wave generated when a finger touched the screen on an Android smartphone and a tablet. The main way of recording the acoustic signal generated when touching the screen was through the built-in device's microphone. An assumption has been made that some malicious application has already been installed on your device without knowing you, in that case, the text or PIN once entered by a user and that text can be inferred from the relevant device. The experiment was done in three (03) steps (1) detecting a single-digit inference (2) PIN inference (3) Letter and word inference. For each step, different numbers of samples were gathered under different conditions.

Another recent study was conducted by [8] on keystroke leakage when typing something while using the virtual keyboard on smart televisions. They have mainly considered on most popular smart TV brands: Apple and Samsung. On smart TVs, a hardware remote controller is used for navigating to each keypress on the virtual keyboard, where some set of actions can be repetitively done while typing something that generates some acoustic signal most probably identical to the action taking place currently: (1) selecting a key, (2) moving a cursor (3) deleting a character. Murali and Appaiah [9] have experimented with detecting the keyboard side-channel attached to smartphones using sensor fusion. This study can be defined as an advanced step taken in the acoustic side channel attack detection context. Rather than using only the acoustic waveform for the analysis, they have looked at the gyroscopes to accelerometers, which are highly responsive sensors equipped with a smartphone. De Souza Faria and Kim [10] have done a study on identifying the text/ PIN when a user is pressing a particular key using the characteristic of the sound that is emitted. The main significance of this study, when comparing the studies done by various researchers, is the usage of the classical frequency spectrum technique, which can be expressed as the transfer function between the two signals which were captured by placing two microphones inside the space of the device.

A recent systematic literature review done by Zunaidi, Sayakkara, and Scanlon [11] stated a proper definition of acoustic side-channel attacks, where we can take advantage

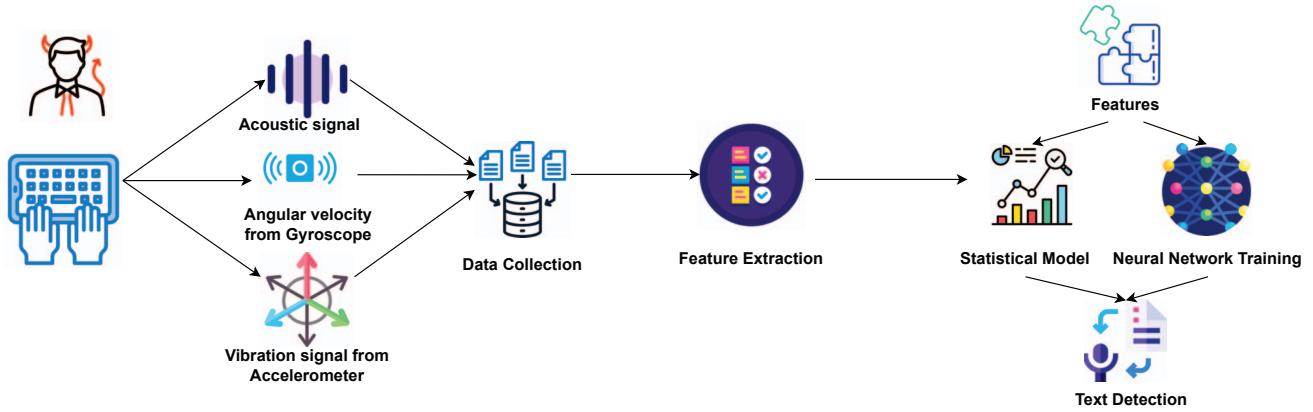


Fig. 1: The overall proposed methodology.

of the sound produced by computers or other peripheral electronic devices, such as keyboards, ATMs, and PIN pads. By utilising optical microphones, acoustic waves and vibrations can be measured while typing something, mainly focusing on comparing and contrasting the unique characteristics of an acoustic signal generated through fingerprints, where it is possible and simple to detect the acoustic wave generated by pressing the space bar in a keyboard.

According to the existing studies reviewed, a gap could be found in experimenting with the keystroke detection using acoustic signals and inertial sensor readings, assuming a malware/app has gained the permission to access both the microphone and both the accelerometer and gyroscope when the mobile device is in its totally silent mode. Through that, the feasibility was identified and the experiments were executed.

III. METHODOLOGY

The primary objective of this study is to systematically analyze the variations in sensor (accelerometer and gyroscope) data and audio data, determining their effectiveness in classifying keypress events under controlled conditions. The experiment follows a structured approach, involving data collection, preprocessing, feature extraction, model training, and performance evaluation. The overall proposed methodology of this research is illustrated in Fig. 1.

A. Data Gathering Process

This study captures keypress events in silent mode using multiple sensors and a controlled setup to ensure high-quality data. The built-in accelerometers and gyroscopes of modern mobile devices detect subtle changes in motion during typing. The accelerometer measures linear acceleration, while the gyroscope records angular velocity, providing motion-based signals that reveal keystroke behavior without relying on audible sound. A microphone is also used to capture any residual acoustic signals, such as slight mechanical vibrations

or surface interactions, even when the device is in silent mode.

A custom Flutter-based mobile app called "Sensor Activity" was developed to collect synchronized sensor and audio data with precise timestamps during structured typing sessions. Experiments were conducted in the Advanced Digital Multimedia Technology Centre (ADMTC) studio at UCSC, considering a quiet environment to minimise external noise and ensure clean recordings. This controlled environment guarantees that the collected data accurately represents keypress signals.

B. Data Pre-processing

To enhance the quality of collected data, noise reduction techniques are applied. Filters are used to filter out unwanted background noise and sensor artefacts as shown in Fig. 2. This preprocessing step helps isolate the relevant keypress signals while eliminating distortions caused by environmental factors or unintended device movement. The feature extraction phase focuses on identifying critical characteristics from the processed data. Key features such as vibration patterns, frequency domain characteristics, and inter-keypress time intervals are extracted, providing valuable input for classification models.

C. Data Analysis and Feature Extraction

To gain meaningful insights, Exploratory Data Analysis (EDA) is performed using detailed statistical and graphical techniques to uncover hidden patterns, trends, and correlations between keypress events and sensor readings. Visual tools like heat maps help identify relationships among features such as vibration intensity, inter-keypress intervals, and subtle acoustic variations. Statistical summaries, including measures such as mean and standard deviation, provide a clearer understanding of the data distribution and reveal potential outliers. After EDA, feature selection identifies the most significant attributes for classifying keypresses. Principal Component Analysis (PCA) then reduces dimensionality by retaining only informative features, enhancing model performance, and avoiding overfitting.

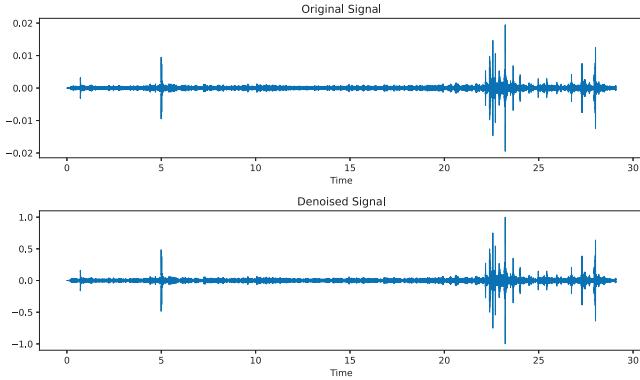


Fig. 2: The original vs. denoised audio signals

D. Classification Model Training using Machine Learning

To achieve accurate keypress classification, various supervised machine learning models are developed and tested. Random Forest is used for its strength in handling complex data patterns and reducing overfitting, while Support Vector Machines (SVMs) are explored for their ability to detect subtle variations by finding optimal decision boundaries. Neural Networks are also implemented to learn nonlinear relationships within the high-dimensional feature space. These models are trained on labelled keypress data to recognize unique patterns for different keypress events. Feature engineering is crucial for boosting accuracy, incorporating domain-specific attributes such as vibration signatures and acoustic profiles. To evaluate the effectiveness of the model, metrics such as accuracy, precision, recall, and F1-score are used to balance false positives and negatives, ensuring reliable performance on unseen data.

IV. EXPERIMENTS

In this section, the experimental setup, data preprocessing with filtering and segmentation, feature extraction process, and the machine learning model training were outlined while utilizing both acoustic and inertial sensor data for each step separately.

A. Experimental Setup

The experiment was carried out in two different environments, ensuring that the performance of the models was tested under different acoustic conditions: (1) quiet environment (e.g., ADMTC studio) and (2) moderate noise environment (e.g., at home).

For this experiment, a Redmi Note 13 Pro Android smartphone was used to capture relevant sensor and audio data. In addition, a hybrid mobile app named "Sensor Activity" was installed on the device, allowing both Android and iOS devices to account for variations in hardware and software. In each keypress event, the audio signal, inertial sensor readings, timestamp, typed character, and saved file location of the audio file were recorded for both soft and hard presses in JSON format. As the initial version of data collection, all keypress events collected solely by ourselves.

B. Data Pre-processing

During the time of inertial sensor data preprocessing, the main focus was to prepare accelerometer and gyroscope readings for each key press, captured along the X, Y, and Z axes with timestamps and corresponding characters. Initially, a low-pass filter was applied to eliminate high-frequency noise, ensuring smoother signal patterns. Missing values, often resulting from sensor latency, packet loss, or CPU-related constraints, were handled by identifying NaN entries and replacing them with the mean value to preserve continuity. In certain experiments, outlier removal techniques were also employed to discard extreme sensor readings that could distort the signal pattern. These steps ensured a clean and consistent dataset for downstream feature extraction and machine learning analysis. Acoustic data preprocessing involved enhancing raw audio signals captured during key presses to ensure high-quality input for machine learning models. The process began by converting .m4a files into .wav format using the pydub library for compatibility with audio processing tools such as librosa. Then each signal was trimmed by removing the initial 5 seconds and the final 2 seconds to isolate the relevant portions containing the 30 key presses. The audio was then segmented into 30 equal chunks, each representing an individual key press. To ensure consistency, amplitude normalization was performed by scaling each waveform between -1 and 1. Finally, noise reduction techniques were applied to minimize environmental and device-related noise, enhancing the clarity of each keypress signal while keeping low-frequency information loss in mind as a trade-off.

C. Feature Extraction

1) *Inertial Sensor Feature Extraction:* For the inertial sensor data, features were extracted from both accelerometer and gyroscope readings along the X, Y, and Z axes. Time, frequency, and statistical domain features were considered to capture the dynamic motion patterns during each keypress. Basic raw features such as A_x , A_y , A_z , G_x , G_y , and G_z were used, along with the calculation of the mean value to reflect time-domain variations. These extracted features helped in identifying subtle motion differences associated with each key press, supporting effective machine learning classification.

2) *Acoustic Signal Feature Extraction:* In the case of acoustic data, both statistical and time-domain features were extracted from the preprocessed audio segments. Thirteen Mel-Frequency Cepstral Coefficients (MFCCs) were computed to represent the shape of the audio spectrum, which is essential for recognizing acoustic events. Additionally, features such as Mean Absolute Value (MAV), Waveform Length (WL), Sum of Squared Energy (SSE), Zero Crossing Rate (ZC), and Root Mean Square (RMS) energy were calculated to capture the intensity, complexity, and energy variations of each keypress sound. These features allowed better differentiation of keypress types for classification tasks.

D. Data Analysis

Data analysis in this study involved processing the extracted features from both inertial sensor and acoustic data to identify patterns and evaluate the effectiveness of clustering-based models in detecting keypress events. For inertial sensor data, sensor fusion techniques were applied by combining accelerometer and gyroscope readings to calculate magnitude-based features. Clustering was then performed using the Gaussian Mixture Model (GMM) and the Density-Based Spatial Clustering of Applications with Noise (DBSCAN). GMM helped differentiate between keypress and non-keypress motions by modelling the data as a combination of Gaussian distributions, while DBSCAN effectively identified clusters based on density, also detecting noise and outliers. For the acoustic data, MFCC and statistical features were extracted and normalized prior to the application of clustering algorithms, including K-Means, Gaussian Mixture Models (GMM), and DBSCAN. Across all clustering approaches, a consistent pattern was observed: most keypress events were grouped into a single dominant cluster, indicating a high degree of similarity in the signal characteristics throughout the dataset.

E. Machine Learning Model Training

For inertial sensor data, machine learning models were trained using accelerometer and gyroscope features, including Raw axes and computed magnitudes. The dataset was split using an 80/20 train-test ratio to ensure balanced evaluation. Model optimization was performed using 3-fold cross-validation within the GridSearchCV procedure. The classifiers included Support Vector Machines (linear, polynomial, and RBF kernels), K-Nearest Neighbours, and Random Forest. These models learned motion patterns associated with soft keypress events, enabling accurate character classification using inertial signals captured during each keypress. The ROC curve of SVM on inertial sensor data can be shown as in Fig. 3.

For acoustic signals, MFCC and statistical audio features were extracted from each keypress and normalized before modelling. The dataset was divided using an 80/20 train-test split. For classical machine learning models, 3-fold cross-validation was applied during hyperparameter tuning to improve generalization. Multiple classifiers were evaluated, including SVM (RBF kernel), KNN, Random Forest, Logistic Regression, and XGBoost. Additional accuracy-enhancement techniques included feature standardization, label encoding, and confusion-matrix-based error analysis. Advanced models, such as an LSTM-based RNN and an ensemble Voting Classifier, were also tested to better capture temporal and nonlinear acoustic patterns.

V. RESULTS

A. Inertial Sensor Data Accuracy Analysis

In the analysis of inertial sensor data, three machine learning models SVM, KNN, and Random Forest were trained using accelerometer and gyroscope readings captured during soft

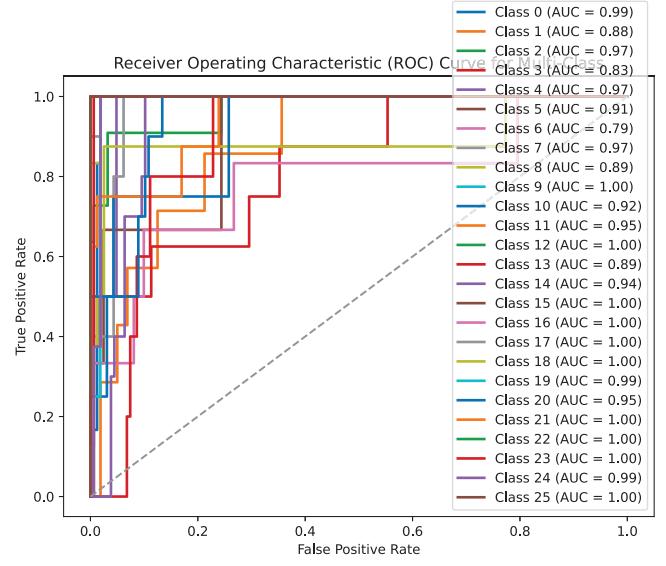


Fig. 3: The ROC curve of SVM on inertial sensor data

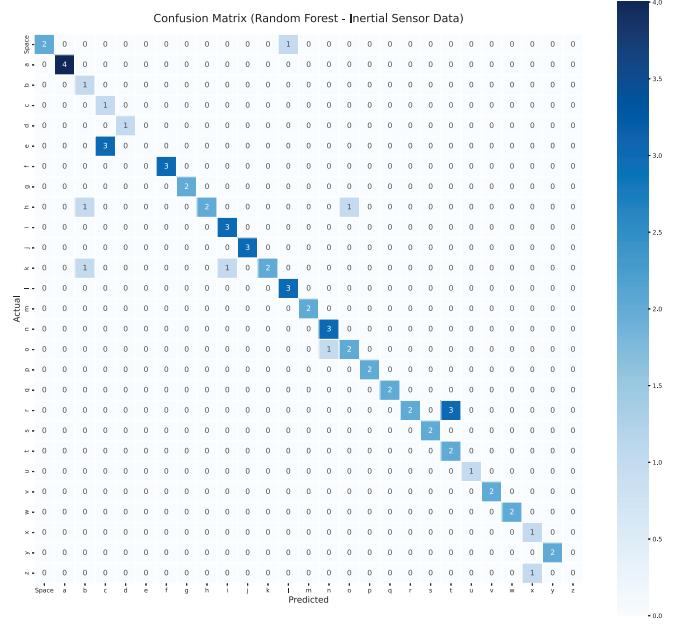


Fig. 4: The confusion matrix of random forest using inertial sensor data

keyresses. The confusion matrix of the random forest model is captured as shown in Fig. 4.

Among the models, Random Forest achieved the highest accuracy at 80%, followed by KNN at 66% and SVM at 65%, as shown in Table 1. Across all models, certain letters like "a" and "j" were consistently well-detected, while others, such as "e" and "z", were often misclassified. The Random Forest model demonstrated superior performance not only in terms of accuracy but also in detecting a broader range of characters with higher F1 scores. A binary tree representation was also introduced using F1 score weights to explore efficient letter

TABLE I: Summary of Accuracies of Inertial Sensor Models

Model	Accuracy
Support Vector Machine	65%
K-Nearest Neighbour	66%
Random Forest	80%

TABLE II: Summary of Accuracies in Acoustic Models

Model	Accuracy
Support Vector Machine	79%
K-Nearest Neighbour	90%
Random Forest	94%
Recurrent Neural Network	40%

prediction strategies, further confirming the strength of inertial sensor data in silent-mode keypress detection.

B. Acoustic Signal Data Accuracy Analysis

As the preliminary experiment for acoustic data, five machine learning models SVM, KNN, Logistic Regression, Random Forest, and XGBoost were initially trained, with Random Forest achieved the highest accuracy of 21%, followed by XGBoost (20%) and Logistic Regression (18%). While these results confirmed the feasibility of using acoustic signals for keypress detection, the accuracy levels were relatively low. Then, an improved version of the audio signal analysis experiments was carried out that demonstrated a significant enhancement in classification performance compared to the preliminary results. As shown in TABLE II, the Random Forest model achieved the highest accuracy of 94%, followed by KNN with 90% and SVM with 79%, indicating the effectiveness of optimized preprocessing, feature selection, and model tuning techniques. Although the RNN model showed relatively lower accuracy at 40%, it still highlights the potential for future deep learning-based improvements. These results confirm that, with proper refinement, acoustic signals can be effectively utilized for accurate keypress detection.

VI. DISCUSSION

Initially, due to time and technical limitations, the experiments were carried out using a small dataset and a single user and device. Considering the limitation of the dataset conventional machine learning model training was tried such as SVM, KNN, and random forest. Although an RNN model was tried, a considerable level of accuracy could be obtained. The experiments conducted on inertial sensor data showed acceptable performance, with the Random Forest model achieving the highest accuracy of 80%, while SVM and KNN reached 65% and 66%, respectively. Although some letters, such as "a" and "j", were highly detectable, several others were misclassified, showing the influence of typing variability and device noise.

In contrast, acoustic signal models performed much better after preprocessing and feature extraction. Random Forest again gave the highest accuracy of 94%, followed by KNN at 90% and SVM at 79%, while RNN underperformed at 40%. Since the precision, recall and F1-scores should be mentioned

separately for each alphabetical letter and space key, only the accuracies have been used for comparisons.

VII. CONCLUSION AND FUTURE WORKS

This study investigated whether virtual keyboard keypresses can be detected in silent mode using inertial sensor data and residual acoustic signals. Data collected under different noise levels and device placements was used to train SVM, KNN, and Random Forest models, with Random Forest achieving the best results around 80% accuracy from inertial data and up to 94% from acoustic data. Key features came from accelerometer and gyroscope variations.

Although the models performed well, their reliability varied across users and devices due to typing style and hardware differences. The findings highlight privacy risks if malicious apps exploit sensor access. Future work should include more users and devices, explore multimodal methods, and evaluate countermeasures to reduce sensor-based side-channel threats.

REFERENCES

- [1] S. Kemp, "Digital 2023: Global overview report – DataReportal – Global Digital Insights," DataReportal, 2023. [Online]. Available: <https://datareportal.com/reports/digital-2023-global-overview-report>. [Accessed: 15-Nov-2024].
- [2] Ericsson, "Ericsson Mobility Report, June 2022," 2022. [Online]. Available: <https://www.ericsson.com/en/press-releases/2022/6/ericsson-mobility-report-5g-to-top-one-billion-subscriptions-in-2022-and-4-4-billion-in-2027>. [Accessed: 15-Nov-2024].
- [3] Zimperium, "Zimperium Report, 2022," 2022. [Online]. Available: <https://zimperium.com/blog/global-mobile-threat-report-key-insights>. [Accessed: 15-Nov-2024].
- [4] Zscaler, "Zscaler Report, 2022," 2022. [Online]. Available: <https://www.zscaler.com/resources/industry-reports/Zscaler-ESG-Report-2022.pdf>. [Accessed: 15-Nov-2024].
- [5] K. S. Balagani et al., "We can hear your PIN drop: an acoustic side-channel attack on ATM PIN pads," in *Lecture Notes in Computer Science*, 2022, pp. 633–652. doi: 10.1007/978-3-031-17140-6_31. [Online]. Available: <https://doi.org/10.1007/978-3-031-17140-6-31>.
- [6] M. B. et al., "Keyboard acoustic side channel attacks on Android phones in the AI era of digital banking: An elementary review," Vol. 12, No. 17s, 2023.
- [7] I. Shumailov et al., "Hearing your touch: A new acoustic side channel on smartphones," arXiv preprint arXiv:1903.11137, Cornell University, 2019. [Online]. Available: <https://arxiv.org/pdf/1903.11137.pdf>.
- [8] T. Kannan et al., "Acoustic keystroke leakage on smart televisions," in *NDSS Symposium*, 2024. doi: 10.14722/ndss.2024.24072. [Online]. Available: <https://doi.org/10.14722/ndss.2024.24072>.
- [9] N. Murali and K. Appaiah, "Keyboard side channel attacks on smartphones using sensor fusion," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018. doi: 10.1109/glocom.2018.8647336. [Online]. Available: <https://doi.org/10.1109/glocom.2018.8647336>.
- [10] G. De Souza Faria and H. Y. Kim, "Differential audio analysis: a new side-channel attack on PIN pads," *International Journal of Information Security*, vol. 18, no. 1, pp. 73–84, 2018. doi: 10.1007/s10207-018-0403-7. [Online]. Available: <https://doi.org/10.1007/s10207-018-0403-7>.
- [11] M. R. Zunaidi, A. Sayakkara, and M. Scanlon, "Systematic literature review of EM-SCA attacks on encryption," arXiv preprint arXiv:2402.10030, Cornell University, 2024. doi: 10.48550/arxiv.2402.10030. [Online]. Available: <https://doi.org/10.48550/arxiv.2402.10030>.
- [12] A. S. George and S. Sagayarajan, "Acoustic eavesdropping: How AIs can steal your secrets by listening to your typing," Zenodo preprint, CERN, 2023. doi: 10.5281/zenodo.8260814. [Online]. Available: <https://doi.org/10.5281/zenodo.8260814>.