ALBA: Novel Anomaly Location-Based Authentication in IoMT Environment Using Unsupervised ML

Fawaz J. Alruwaili[1 [0000−0003−0941−9855]],
Saraju P. Mohanty[1 [0000−0003−2959−6541]], and Elias Kougianos[2 [0000−0002−1616−7628]]

[1] Department of Computer Science and Engineering, University of North Texas, USA.
fawazalruwaili@my.unt.edu, saraju.mohanty@unt.edu
[2] Department of Electrical Engineering, University of North Texas, USA.
elias.kougianos@unt.edu

Abstract. Smartphones have become essential components in the Internet of Medical Things (IoMT), providing convenient interfaces and advanced technology that enable interaction with various medical devices and sensors. This makes smartphones serve as gateways for sensitive data that could potentially affect patients' health and privacy if compromised, making them primary targets for cybersecurity threats. Authentication is crucial for IoMT security, as its effectiveness relies on its resistance to any conditions of environment, device, or user. In this paper, we propose the Anomaly Location-based Authentication (ALBA) method using GPS technology and a lightweight unsupervised ML algorithm with more stable features. Our experimental results showed that the model successfully identified anomalous locations across three distinct datasets, demonstrating the adaptability of ALBA.

Keywords: Healthcare Cyber-Physical System (H-CPS) · Internet of Medical Things (IoMT) · Intelligent Security · Cybersecurity · Location-based Authentication

1 Introduction

The growth of IoT embedded systems and biosensors, has introduced the IoMT as a branch that integrates medical devices, applications, and networks to enhance the efficiency of healthcare system [1, 2]. The rapid advancements in mobile technology have enabled smartphones to become an important component of the IoMT network and a source of information due to the increasing complexity of software and hardware components and multiple interfaces in medical devices [3]. However, smartphones also introduce new security challenges due to the sensitive nature of medical data that they collect, making them a valuable target for cybersecurity threats [4, 5]. Therefore, ensuring the security of IoMT is crucial to mitigate risks and enhance the sustainability of healthcare.

Artificial Intelligence (AI) technologies have been advanced significantly and can be used to monitor and predict the behavior of entities within an IoT environment. However, data quality is crucial in machine learning (ML) for achieving accurate results. While many studies on behavioral authentication for smartphones have contributed valuable insights, research expectations have not met in terms of accuracy or considered IoT device security requirements under different conditions related to environment, device, and user. Therefore, effective IoMT security solutions require holistic security considerations, while maintaining user convenience

In this paper, we propose a behavior-based authentication method for smartphones in IoMT network using GPS sensors and an unsupervised ML model, which can be utilized as a additional security layer without requiring user intervention, and with more stable features. Fig. 1 depicts the basic overview of our proposed method. The method's efficacy was evaluated using three distinct datasets. The results demonstrate its adaptability to various realistic conditions, indicating its potential to be implemented in IoMT.

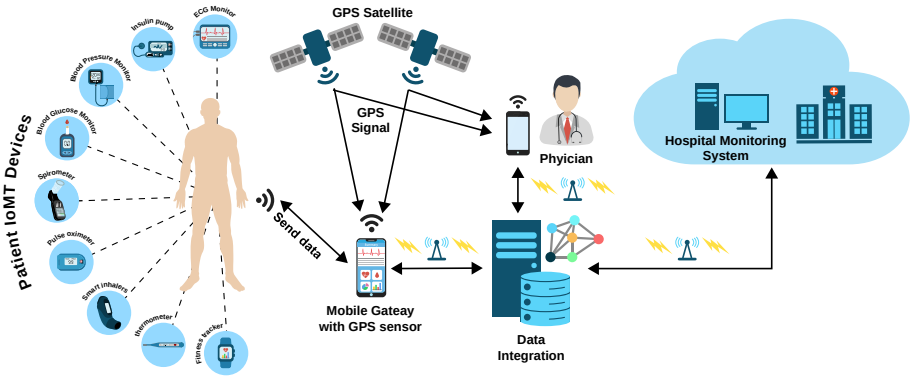


Fig. 1: Overview of the Proposed ALBA for IoMT.

The paper is organized as follows: Section 2 presents the literature review, while Section 3 introduces the novel contributions of this paper. The proposed method is presented in Section 4, and data preprocessing is detailed in Section 5. The ML model used in our method is described in Section 6. Experimental results are provided in Section 7, and the conclusion with future work are in Section 8.

2 Related Research on Behavioral Authentication

Existing research on behavioral authentication has provided valuable insights using various techniques and sensors, such as Keystroke Dynamics (KD) [6, 7], Touch Gestures (TG) [8, 9], and Gait Behavior [10, 11]. However, some studies may not have fully met expectations in terms of accuracy and stability of authentication data, nor considered the holistic security needs covering diverse environmental, device, and user conditions. This limits their accuracy, suitability and effectiveness for the IoT devices, especially smartphones.

The limitations of these techniques are mainly due to internal and external factors, such as the variety of devices, where smartphones have touchscreens or keyboards with different shapes, layouts, and sizes [12]. Also, the specific language in which technique is applied affects the tested interval time between touches, where the user may be unfamiliar with some vocabularies. Additionally, there are external factors that affect these techniques, such as, environment, clothing, sickness, injuries, fatigue, emotional or mental status, and smartphone position. These limitations make the extracted features insufficient for behavioral-based authentication. Based on the above discussions, we conclude that existing approaches to

behavioral authentication in IoMT are still lacking and have limitations. Therefore, ALBA method aims to address these limitations and improve authentication data stability to be more accurate and usable in IoT devices.

3 Novel Contributions

3.1 Problem Addressed and Proposed Solution

Smartphones have revolutionized healthcare access due to thier advanced technology, where they used to collect and transmit sensitive medical data, making them vulnerable to cyber-attacks that compromise patient privacy and have life-threatening consequences. Therefore, securing smartphones within the IoMT network is essential. Various behavioral authentication methods for smartphones have been proposed to address vulnerabilities in traditional authentication factors. However, these methods face challenges in performance and accuracy due to factors impacting authentication data stability. Therefore, authentication methods must consider holistic security considerations, the nature of devices, targeted environments, and their applicability to available technologies.

ALBA exploits GPS technology in smartphones to authenticate users based on th ebehavior of their locations utilizing ML technology for analyzing and detecting anomalous locations, ensuring faster response times to security threats. ALBA overcomes limitations of behavioral features used in previous studies, and provides more stable behavioral features under different conditions related to environment, device, and user. GPS sensors can be embedded in multiple IoMT devices without requiring specific hardware design or size.

3.2 Novelty of the Proposed Solution

ALBA method provides several contributions: robustness by being less sensitive to internal/external factors and countering for GPS inaccuracies; increasing efficiency as GPS requires less features and a lightweight iForest algorithm that has low memory requirements, reducing computational demand and power consumption compared to existing behavioral methods; scalability and applicability with GPS integration in various IoMT devices without specific hardware requirements; enhancing user convenience as our method can be additional security layer along with existing authentication factors, reducing their constraints. These contributions make our proposed method more suitable for device technologies and more comprehensive in terms of security considerations in an IoMT environment. To the best of our knowledge, we have proposed the first behavioral authentication method integrating GPS and unsupervised ML technologies for IoMT security.

4 Proposed Authentication Mechanism

When user credentials are validated, authentication is based on comparing the current location with historical locations stored in the database within a given time frame. If the behavior of current location matches the behavior of the historical locations, it will be considered normal location. Otherwise is anomaly.

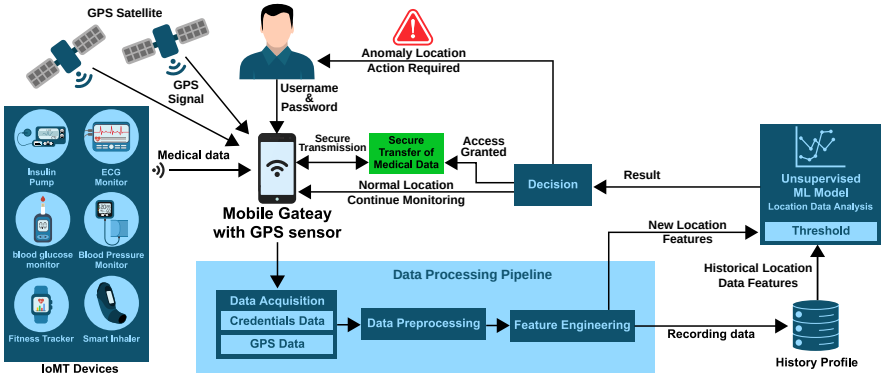


Fig. 2: ALBA Method Workflow for IoMT.

Fig. 2 illustrates the proposed authentication method, where the mobile device collects data from different resources (user, medical devices, and GPS satellites) and transmits it to the medical server's IoMT system for verification. Specifically, when users connects their medical devices to the server, the IoMT system prompts the user for username and password. The GPS data are then verified and analyzed using ML algorithms to detect whether the current location is anomalous or not. If the location is normal, the verification process is successful, and user's medical devices will be connected to the medical server, allowing secure transmission of medical data for doctor diagnosis. The user also will be able to access health record. Historical location data is pre-processed before being stored to reduce computational time and resource consumption, which positively impacts power consumption during future authentication processes. The result of data analysis determines whether to continue monitoring user's current location or take appropriate action in case of any deviation from the expected behavior, such as limiting system functionality until additional authentication is provided or sending alerts through other channels.

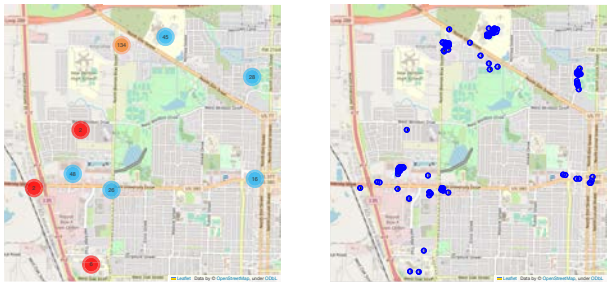
4.1 Data Collection

The real-world dataset was collected over 27 days using the Google Maps app on an iPhone 11 Pro, with 359 locations visited during various times of the day and using different navigation modes. Fig. 3 (a) illustrates a sample of the recorded locations density, with reduced clutter to improve readability. Fig. 3 (b) shows the recorded locations individually.

Data accuracy is crucial in ML, and significantly impact model performance. Therefore, data collection process was monitored daily to ensure the accuracy.

4.2 Datasets Description

The effectiveness of ALBA was evaluated using three distinct datasets: (1) a real-world dataset which was collected for this study with 359 observations recorded at irregular intervals over 27 days using an iPhone 11 Pro to evaluate ALBA under real-world scenarios, (2) a public dataset that was utilized to evaluate the performance of ALBA on different real-world data and scenarios, and to ensure its generalizability. It was obtained from Kaggle [13] with 40,603



(a) Clustered locations (b) Individual locations

Fig. 3: Sample of collected locations - Real-world Dataset.

observations collected in October 2014 using an Android device, and (3) a virtual dataset which was created with 3,359 observations using Python programming language, including 5 anomalous locations with different regular patterns and sudden changes. It was utilized for evaluation under specific realistic scenarios not clearly present in the previous datasets.

5 Data Pre-processing

In our experiment, location features, such as latitude, longitude, and timestamp features are selected from the original dataset, while others are redundant. Day and hour features were extracted from timestamps to improve anomaly detection by identifying deviations during specific times or days. In addtion, latitude and longitude features are standardized to be on comparable scales. These features are then combined using PCA to create a single location feature, enabling easier analysis and visualization of location patterns and anomalies.

6 Isolation Forest model

The iForest is an unsupervised ML algorithm used for anomaly detection without pre-labeled data. It has several advantages that make it suitable for our proposed method, including its low memory requirements, reducible model sensitivity, adaptability to data distribution changes.

Anomalies are isolated by building decision trees (DTs), which are combined to produce prediction. DTs are constructed by recursively selecting a random feature and a split value within the range of the selected feature. The iForest has a linear time complexity O(n), as it isolates anomalies instead of normal observations [14], where anomalies are expected to be fewer [15], tending to be closer to the root. The number of DTs affects the model performance and accuracy, but also impacts computational time and resource requirements. The optimal choice relies on the dataset, available resources, and multiple experiments.

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{1}$$

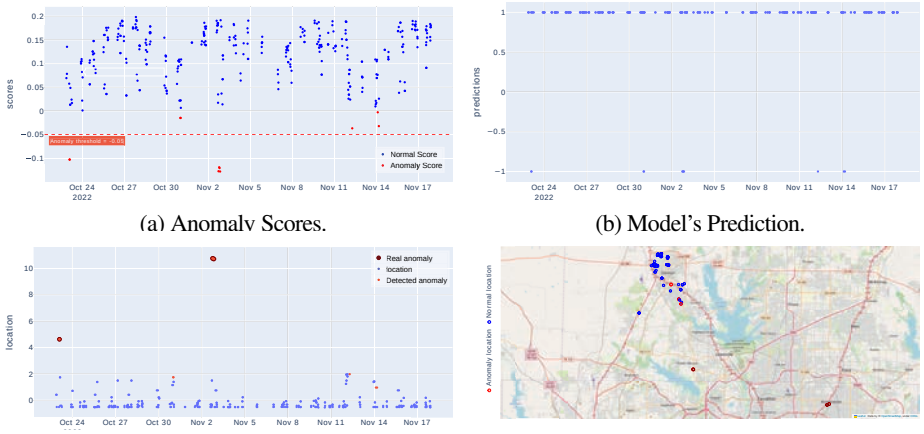
The base 2 in the exponential function is to ensure the score is between 0 and 1. The number of observations is indicated by n, where $h(x)$ is the path length of a point x, and $E(h(x))$ its expected average path length. The constant $c(n)$ is the average path length of terminal nodes

in DTs, used to scale and normalize scores. Utilizing a suitable threshold value is essential for accurately identifying security threats, especially in the IoMT where false positives can disrupt medical operations and compromise patient safety. Identifying an optimal threshold requires iterations, evaluating results, and refining the value with domain knowledge and expert input.

7 Experimental Results

7.1 Real-world Dataset Results

In our experiment, the iForest model was trained on location data in the real-word dataset. The results showed that the model successfully calculated anomaly scores as shown in Fig. 4 (a), identifying 11 anomaly scores represented in red dots as negative values, while positive vlaues (blue dots) represent the normal scores. The farther from 0, the more anomalous (or normal) a location is. The model's prediction is shown in Fig. 4 (b) as a binary series of -1 for anomalies and 1 for normal points.



(a) Anomaly Scores. (b) Model's Prediction.

(c) Anomalous Locations (d) Result on map.

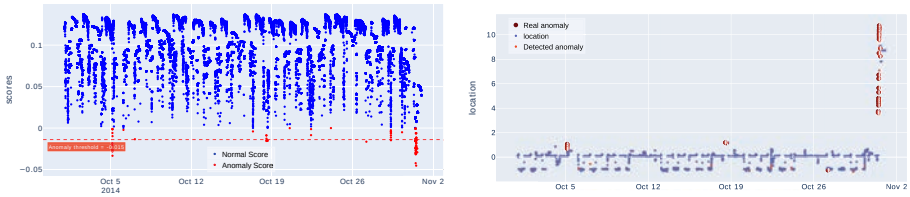
Fig. 4: Real-world dataset Results - A threshold of -0.05

The model's sensitivity was controlled by utilizing a anomaly threshold of -0.05 after conducting multiple experiments and evaluating performance, leading the model to identify 6 real anomalous locations with a significant deviation indicated by dark red in Fig. 4 (c). For contextual anomalies, they occurred during certain hours or days, which makes them different from other locations. However, the model considered them as normal based on the determined threshold as they have slight deviation. For more insight ont he spatial distribution of anomalous and normal locations, they were projected on the map depicted in Fig. 4 (d).

7.2 Public Dataset Results

The model was trained on the public dataset, and all anomaly scores were calculated successfully as shown in Fig. 5 (a). There were 121 anomalous scores (red dots), where 58 of them

were above the utilized threshold of -0.015, representing real anomalous locations as shown in Fig. 5 (b) represented by dark red dots.



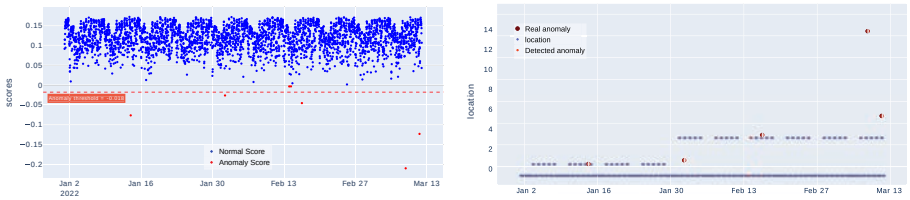
(a) Anomaly Scores. (b) Anomalous Locations

Fig. 5: Public Dataset Results - A threshold of -0.015

We can notice that some of dark red locations have a significant deviation from the behavior of other locations, while the red locations have a slight deviation, and which cosidered normal based on the determined threshold. However, the public dataset locations are shown in a clear daily pattern with some significant deviation, especially on the right side. Most of the deviations occur at the beginning or end of the week or during the weekends, which is a reasonable pattern.

7.3 Virtual Dataset Results

In the virtual dataset, there were 7 anomaly scores as depicted in Fig. 6 (a). Based on the calculated anomaly scores, the model successfuly identified 7 anomalous locations as illustarated in Fig. 6 (b)



(a) Anomaly Scores. (b) Anomalous Locations.

Fig. 6: Virtual Dataset Results - A threshold of -0.018

The model effectively detected all 5 known anomalous locations, which are represented by dark red dots. However, there were 2 false positives identified by the model, which were due to the adjustment of the model's parameters to optimize accuracy for the virtual dataset. Despite this, by adjusting the threshold value to 0.018, the model effectively reclassified these false positives as normal locations (red dots), demonstrating its ability to adapt and perform well on the given dataset.

8 Conclusion and Future Work

Integrating GPS and ML technlogoies can enhance the security of traditional authentication factors. The behavioral patterns in proposed ALBA are more stable and accurate compared

to previous studies, which depend on other behavioral patterns that can be affected by various factors. The experimental results across diverse datasets validate the model's ability to detect location deviations from normal patterns, ensuring effective authentication. For future work, we suggest exploring the integration of additional behavioral patterns and data types to further improve effectiveness and robustness of authentication process.

References

1. S. P. Mohanty, U. Choppali, and E. Kougianos. Everything you wanted to know about smart cities: The internet of things is the backbone. *IEEE Consumer Electronics Magazine*, 5(3):60–70, 2016.
2. A. Ghubaish, T. Salman, M. Zolanvari, D. Unal, A. Al-Ali, and R. Jain. Recent advances in the internet-of-medical-things (IoMT) systems security. *IEEE Internet of Things Journal*, 8(11):8707–8718, 2021.
3. A. G. Mutyara, B. A. Farras, L. P. Sari, S. Achmad, and R. Sutoyo. The influence of smartphone applications on human healthcare. In *International Conference on Informatics Electrical and Electronics (ICIEE)*, pages 1–6, 2022.
4. G. M. E. ur Rahman, R. I. Chowdhury, A. Dinh, and K. A. Wahid. A smart sensor node with smartphone based iomt. In *IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia)*, pages 92–95, 2019.
5. W. Iqbal, H. Abbas, M. Daneshmand, B. Rauf, and Y. A. Bangash. An in-depth analysis of iot security requirements, challenges, and their countermeasures via software-defined security. *IEEE Internet of Things Journal*, 7(10):10250–10276, 2020.
6. T. L. Lin and Y. S. Chen. A chinese continuous keystroke authentication method using cognitive factors. In *IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW)*, pages 1–2, 2019.
7. H. M. C. K. B. Herath, K. G. C. Dulanga, N. V. D. Tharindu, and G. U. Ganegoda. Continuous user authentication using keystroke dynamics for touch devices. In *2nd International Conference on Image Processing and Robotics (ICIPRob)*, pages 1–6, 2022.
8. Y. Ouadjer, M. Adnane, and N. Bouadjenek. Feature importance evaluation of smartphone touch gestures for biometric authentication. In *2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH)*, pages 103–107, 2021.
9. A. Suharsono and D. Liang. Hand stability based features for touch behavior smartphone authentication. In *3rd IEEE International Conference on Knowledge Innovation and Invention (ICKII)*, pages 167–170, 2020.
10. P. Musale, D. Baek, and B. J. Choi. Lightweight gait based authentication technique for iot using subconscious level activities. In *IEEE 4th World Forum on Internet of Things (WF-IoT)*, pages 564–567, 2018.
11. L. He, C. Ma, C. Tu, and Y. Zhang. Gait2vec: Continuous authentication of smartphone users based on gait behavior. In *IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 280–285, 2022.
12. S. Roy et al. A systematic literature review on latest keystroke dynamics based models. *IEEE Access*, 10:92192–92236, 2022.
13. J. SimonD. Mobile location history of 10/2014. Kaggle, 2017. [Accessed: Dec-28-2022].
14. L. Zhang and L. Liu. Data anomaly detection based on isolation forest algorithm. In *International Conference on Computation, Big-Data and Engineering (ICCBE)*, pages 87–89, Yunlin, Taiwan, 2022.
15. E. Marcelli, T. Barbariol, V. Savarino, A. Beghi, and G. A. Susto. A revised isolation forest procedure for anomaly detection with high number of data points. In *IEEE 23rd Latin American Test Symposium (LATS)*, pages 1–5, Montevideo, Uruguay, 2022.