Journal of Information Systems Engineering and Management
2025, 10(24s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Application of Artificial Intelligence for Forming Evidence Base of Crimes in Smart Homes

Abdurakhimov Bakhtiyor[1] , Kurmaeva Malikakhon[2] , Makhmudova Miyasar[3]
[1] *Department of Applied mathematics National University of Uzbekistan, Tashkent, Uzbekistan*
[2] *Department of Cybersecurity and Forensics, Tashkent University of Information Technologies, Tashkent, Uzbekistan*
[3] *Tashkent State Agrarian University, Tashkent, Uzbekistan*

ARTICLE INFO

Received: 25 Dec 2024

Revised: 12 Feb 2025

Accepted: 24 Feb 2025

ABSTRACT

This paper presents a solution for forming an evidence base for crimes involving IoT devices in smart homes. It introduces artificial intelligence model assistant as the way to identify potential crime evidence. The crime evidence is identified by detecting abnormal behavior related to user and the sequence of data coming into the controller of a smart home [1]. The proposed approach analyzes data from smart home ecosystems and determines the likelihood of specific events being related to a crime [2-3]. The data is analyzed by artificial intelligence model bringing outputs in term of probabilities. The results are presented as percentages, reflecting the probability of connection of specific event and a sequence of events to a crime. Final calculations are brought to attention of a forensic specialist as a half sum of 2 probabilities. Using provided data the digital forensics criminalist can base further investigation and build theories on circumstances of a crime. This approach helps specialists to concentrate attention on valuable events by highlighting them and reducing the time spent reviewing data.

Keywords: anomaly, artificial intelligence (AI), ecosystem, evidence, Internet of Things (IoT), smart home.

INTRODUCTION

A smart home is a concept of living space equipped with intelligent technologies that provide automatic and/or remote control of various systems and devices. Such homes use integrated management systems to enhance comfort, energy efficiency, security, and convenience of living [4]. The control is maintained using smartphones, laptops, voice assistants or special interfaces. A smart home consists of different components communicationg with each other to automate processes happening in it. A smart home is one of the prominent examples of the application of the Internet of Things (IoT). In a smart home, many devices that are connected to the network and interact with each other are used [5-6] . The Internet of Things (IoT) is a concept that describes a network of physically connected devices (things) capable of collecting, exchanging, and processing data over the internet. These devices are equipped with sensors, processors, software, and communication modules, allowing them to interact with other devices and systems without human involvement [7-8]. Modern IoT technologies in smart home make user life easier but, at the same time, they initiate new risks associated with the security and possibility of usage for criminal purposes. IoT devices, such as surveillance systems, smart locks, and sensors, continuously generate vast amounts of data that can become important evidence in the event of crimes. However, standard methods of data processing and analysis are not always suitable for effectively identifying potential violations in real time. According to the Yahoo Business a smart home produces 1GB data in a week in general which is a complicated task for processing using todays solutions and tools [9]. As it is not only important to retrieve and analyze data from IoT devices in cases related to smart homes but also correlate the indexes to each other including highlighting the anomalies to gain a whole concept of the events happened the help of artificial intelligence-based solution is a direction for the development. In this regard, it becomes a relevant task, which will allow for the analysis of data from IoT systems to create a body of evidence, especially in the context of complex ecosystems of smart homes.

As of today, researchers are actively studying issues related to the security of IoT and smart homes, including anomaly detection, diagnostics, and data analysis methods. Many scientific papers focus on detecting security threats in IoT

environments, as well as on the use of artificial intelligence and machine learning for processing big data. However, research aimed at building an evidence base using AI remains relatively new and relevant. Significant attention is paid to anomaly detection, but its implementation in the field of forensics for IoT devices is just beginning to develop. For instance, Salamh presents a forensic analysis framework called the FAHAD model for investigating home automation devices [10]. The study focuses on two smart home devices: the Kasa Smart Light Bulb and the Eufy Floodlight Camera. It explores how data is stored, transmitted, and can be extracted for forensic investigations. The research provides insights into potential security vulnerabilities and highlights methods for collecting and analyzing digital evidence from IoT devices in smart home environments. Abiodun Abdullahi Solanke in his article called «Digital Forensics AI: Evaluating, Standardizing and Optimizing Digital Evidence Mining Techniques» [11] describes how AI technologies can contribute to improving the forensic analysis of digital data. The approaches which allows machine learning algorithms to identified hidden patterns in digital artifacts are also reviewed. The work introduced by Yamana Salem, Magi Oda, and Amani Youssef Oda is dedicated to the development of a multi-level digital forensic framework for Internet of Things (IoT) devices that encompasses not only data collection but also validation and correlation of data between devices to establish a sequence of events within a single criminal activity [12]. They emphasize that it helps to decrease investigation time and increase accurancy of digital footprint documentation and detection presuming its integrity. The main focus is on the automation and standardization of processes for IoT expertise, which allows forensic examinations to be more comprehensive and reliable. The authors suggest new multilevel digital forensics framework (MAoIDFF-IoT) which key feature is its ability to cover all levels of IoT architecture and focus on artifacts of interest (AoI). The study introduces an Action/Detection Matrix, which helps to systematize and accelerate the process of forensic examination in IoT environments. The experiments conducted showed that the proposed framework outperforms existing methods in terms of usability, completeness of coverage, focus on AoI, and acceleration of the investigation process. Thus, the research represents a significant contribution to the field of digital forensics, offering an adapted and effective methodology for analyzing data generated by IoT devices.

However, these works address the issue at the stage of data acquisition from the sensors of Internet of Things devices before their processing by the central controller of the system. Moreover, although the use of artificial intelligence is mentioned as a method for gathering evidence through data correlation, there is no development of a solution in it. In this article the structured data from leading smart homes ecosystems including Google Home, Apple HomeKit and Amazon Alexa is used for analysis and evidence identification. The concept is focused on working with data which can be accessed from the resources and databases owned by the ecosystems. In such terms it is beneficial as it allows to concentrate on analysis of data which has been already processed and normalized. Such data has a common structure and is a suitable format for AI model to be built on. Also, the methodology which describes Artificial Intelligence usage as smart homes crime evidence detection tool is introduced. The methodology grasps the step-by-step guide to developing introduced solution from data processing till representing the result. At the same time in this paper the first iteration of the proposed artificial intelligence model development is described showing the result of experiments in which real data is used to calculate the probability of events correlation to the crime.

PROBLEM STATEMENT

The work presents a methodology for using AI in identification crime evidence in smart homes, as well as a model of AI that detects potential evidence from available data stored in smart homes ecosystems. It provides a probability percentage representing the relationship of events in particular and their sequences to anomalies which can be used as evidence to a crime. Structured data from leading smart home ecosystems on the market—Google Home, Apple Homekit, and Amazon Alexa—are used for analysis and evidence identification. The object of the research is the process of forming an evidence base for crimes committed using IoT devices in smart homes. The subject of the research is the use of artificial intelligence models for analyzing data from smart home ecosystems, with the aim of identifying anomalous actions and sequences of events.

The following tasks were identified under this work area:

1. This phase of the work is dedicated to the development of a methodology for using artificial intelligence to process and analyze data stored by smart home ecosystems. The key element of this methodology will be to create a system approach that ensures effective interaction with data coming from different IoT devices and allows to extract useful information for future use. The process will include several key steps:

- Data Acquisition: A step-by-step plan will be developed to extract data from the databases constituting the smart home ecosystem. This includes interaction with sensors, cameras, lighting control systems, temperature and other intelligent devices. This phase will focus on ensuring the correctness and security of data transmission.
- Primary data processing: The collected data will be subjected to a thorough initial processing, which includes normalization and structuring for further analysis. During the step data should be transformed into the form understandable for the system which includes dismiss of data meaningless for the processing and mapping of data indexes (database columns) to one introduced in the system. This will eliminate possible distortions and prepare data for more accurate and efficient analysis.
- Artificial intelligence analysis: Based on the cleaned and prepared data, artificial intelligence model will be applied to identify abnormal patterns using two approaches: anomalies found in data compared to the information related to user's normal behavior and those different from other sequences of events. AI will be trained based on data artificially generated or obtained from real smart home ecosystems' databases. They should contain user behavior records, changes in the environment indicators and events occurring in the house, with a goal to obtain accurate results that can be used for analysis of criminal acts and other events.
- Development of mathematical model: The outputs of the artificial intelligence model contain two probabilities values from two processes in it. For this step of research both results have equal weight on result. In this case, the decision on a specific event being related to a crime and considered evidence is calculated by a mathematical model as a half sum of both values. The decision on a model weight is a point of further discussions. Development of a model that will allow to identify accurately digital traces of cybercrimes. The system will give a probability that a certain set of data is relevant to a crime, which will help to identify key evidence quickly.
- Structured results display: The analysis results will be presented in a form that is easy to understand, which will ensure not only visibility but also simplicity of interpretation. These can be graphs, tables, time scales and other forms of visualization that will allow investigators and analysts to make decisions quickly and accurately. During this step of research, the result should be demonstrated in a form of console output including the event itself, its details (timestamps, intermediate results) and final calculated probability. Development of an interface that will present the results of artificial intelligence in a format that is easy to understand and use by investigative bodies. This will allow specialists to quickly navigate complex data, improving the speed and accuracy of investigation.

The development of this methodology will thus provide a basis for the establishment of a highly efficient system capable of extracting valuable information from complex data of smart home ecosystems and will provide new security and investigative capabilities.

2. This phase will involve the development and training of artificial intelligence (AI) for data analysis, which constitutes the smart home ecosystem. The challenge of AI is to effectively interpret them with advanced machine learning algorithms to extract useful information from these data and identify patterns that can indicate possible criminal actions. The first iteration of development should be finished with the prototype of the artificial intelligence model which is able to interpret input data and output valid probabilities. At this point, the success of the model is measured by its ability to process given data and result into the intermediate algorithm probabilities values and the value calculated by the mathematical model. Values should be normalized and correctly mapped to corresponding events or sequences. The functionality of the system should be tested by experimental means.

As a result of the implementation of this project, a system will be created that can effectively analyze data from smart homes and other IoT systems, which will significantly improve the quality of investigations into crimes in the field of cybersecurity and law enforcement.

GOALS

The aim of this study is to develop a methodology, artificial intelligence (AI) and mathematical model for analyzing data generated in smart home ecosystems. Its main purpose is to ease the work done by digital forensics criminalists, decrease the time needed for investigation of digital materials, point out the most important parts of data to increase

the investigation success rate and decrease the error, mislead and overlooking of crucial evidence. Special attention is given to the evaluation of the probability of certain events recorded in these systems as they can serve as legal evidence for criminal acts.

MAIN PART

The smart home ecosystem is a network of devices and technologies that work together to automate and simplify home management. The central part of the smart home is often a central controller or app that connects all devices and allows them to be controlled remotely via a smartphone or voice commands using assistants. This study focuses on the three leading ecosystems by the number of users: Google Home, Apple HomeKit, and Amazon Alexa.

1. Google Home: Managed through Google Assistant, it is ideal for users of the Google ecosystem.
2. Apple HomeKit: A system designed for users of Apple devices. It offers high security and automation, as well as excellent integration with Apple products such as iPhone, iPad, Apple TV, and HomePod.
3. Amazon Alexa: Based on Amazon's voice assistant Alexa, it offers extensive compatibility with various smart devices from different manufacturers.

Each of these three systems has over 100 million users.

1. Google Home: Operated through the Google Assistant and is ideal for users of the Google ecosystem. Users can control various devices including thermostats, lights, security cameras, voice commands and mobile applications. Has over 100 million users.

2. Apple HomeKit: This is a system oriented to users of Apple devices. It offers high security and automation, as well as excellent integration with Apple products such as iPhone, iPad, Apple TV and HomePod. HomeKit provides an easy-to-use interface for controlling smart devices such as lighting, security and thermostats, and supports a high degree of data protection. However, the number of compatible devices in the HomeKit ecosystem is less than that of competitors. Has about 100 million users.

3. Amazon Alexa: Based on the Amazon Alexa voice assistant and offers extensive compatibility with various smart devices from different manufacturers. Alexa stands out for its ability to integrate with the Amazon ecosystem, which allows convenient shopping and also good support for music services. Alexa also supports a variety of devices, including smart speakers and displays, and offers extensive options for home automation settings. Has over 200 million users.

Despite the different operating principles, the ecosystems of Google Home, Amazon Alexa and Apple HomeKit store similar data in a similar format, especially in the context of what is needed for smart home automation. This includes:

1. User data:
   - User information: name, email, account settings, preferences.
   - Action history: how often and when the user uses devices, what commands were executed.
   - Preferences settings: user preferences, such as lighting modes, temperature, preferred scenarios.
2. Device data:
   - Device type: information about connected devices (e.g., smart bulbs, thermostats, cameras).
   - Device status: information about the current state of the device (on/off, temperature, battery level etc.).
   - Device settings: Device-specific settings (e.g., brightness for light bulbs, temperature for thermostats).
   - Device logs: information about the device's operation, errors, notifications.
3. Sensor and display data:
   - Temperature, humidity: data from thermostats, temperature sensors, humidity.
   - Motion sensors: information about movement in the room.
   - Sensor readings: such as CO2 levels, air quality, light level and other parameters.

- Energy consumption: data on the energy consumption of smart devices (e.g., smart sockets, electrical appliances).

4. Scenario and automation data:
   - Automatic scenarios: information about the automation scenarios created (for example, lighting on schedule or temperature change under certain conditions).
   - Event log: record all actions occurring in the system (switching on and off devices, activation of scripts).
5. Safety data:
   - Surveillance video: if the device is video-enabled, the camera recordings can be stored in the cloud (usually for a certain amount of time).
   - Alarm data: information about the activation of motion sensors, sound or opening of doors.
   - Access history: records of when and who accessed the system (for example, through smart locks or security systems).
6. Analytical and reporting data:
   - Energy use: data to analyze energy consumption, helping to optimize costs.
   - Performance data: analysis of how often devices are used, their efficiency and optimal settings.

The data presented in the tables of the ecosystem databases of smart home, make it possible to determine the standard behavior of the user, compare it with the data coming from the devices for the detection of anomalies. Also, the sequence of events can give important information on which a model can calculate its probability relation to a potential crime. These data will be used to teach the model and also analyzed when it is used.

Methodology

To effectively form an evidence base for crimes in Internet of Things (IoT) systems of smart homes using artificial intelligence (AI), the following key stages can be identified [13-14]:

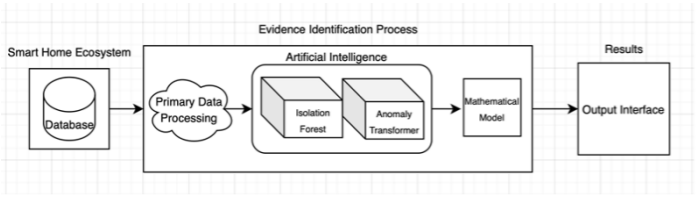

Fig.1 - Methodology for using artificial Intelligence to Identify evidence

1. Data extraction. Access to device data from storage, including event logs, sensor readings, and user behavior data, is essential. The data stored can vary depending on the database type and a company, but the usage of many columns is minimized to decrease the necessity to customize the system according to differences existing. For this purpose, the following database tables of the smart home ecosystem are used:
   - Events. This table includes all events happening in a smart home including user actions, changes in devices, actions happening. For instance, the opening of a door, temperature changes in a room, user authorization are the data stored in events table of an ecosystem. The data stored in events table is crucial for comparison with the usual user experience as well as other events to detect the anormal behavior and patterns. This is the base of the research and the main data resource.
   - Devices. The devices table consists of columns related to the device information. Those can be an id as a table key, title, description of a device, its status, manufacturing company and timestamps of creation and deletion. This data is important to correlate the events stored in database with the device the signal came from. The device data allows the model to base the calculations by introducing new characteristics and grouping data by the type of the device.
   - Schedules. Schedules is a table which stores the timetable of events registered by the user. Usually, this table represents the activities a user programmed for the convenience of usage. For instance, a user made a

schedule of smart lighting to be turned off automatically after 9 p.m. every day. This record allows the system to have an insight of a usual behavior of its owner which is important to identify when the events in a smart homes ecosystem differ from the expected ones. This data consists of the user id (primary key of a table users), device id (primary key of a table devices), timestamp of a schedule being registered, the time and occurrence frequency of an action, and the action itself or its code/id.

2. Data Preprocessing for AI Analysis. The next step involves cleaning and normalizing the collected data by removing gaps, outliers, and inconsistencies. Also, data should be mapped to expected value as column and table naming in database can vary depending on an ecosystem. Necessary tables should be chosen, mapped to the names introduced in the system. The same operation is processed with the tables' columns. Additionally, data from various tables can be merged into a unified structure to ensure compatibility for subsequent analysis.
3. Data Analysis Using AI. The prepared data is then analyzed by artificial intelligence model. This solution proposes two elements of analysis:
    - Isolation Forest is an algorithm used to detect anomalies in comparison to the normal user behavior. The Isolation Forest algorithm is based on the concept of isolating anomalous points in the data [15-18]. Its key principles include:
        - Number of Splits for Isolation: A split in Isolation Forest is a randomly selected threshold applied to a randomly chosen feature, dividing the dataset into two subsets. Each split helps progressively isolate individual data points.Anomalous points require fewer splits in a randomly constructed tree compared to normal points, making the algorithm effective in detecting them [19].
        - Anomaly Score: The degree of anomaly is determined by the average path length (h(x)), calculated across multiple trees and normalization correcting factor (c(n)), related to the number of points in sample. The anomaly score represented by a normalization factor (s(x,n)) related to the sample size (n) is computed as:

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

If anomaly score approaches 1, the point is considered anomalous. If s(x,n) approaches 0.5, the point is likely normal.

- Anomaly Transformer in unsupervised mode for anomalies detection in sequences of events. Anomaly Transformer — machine learning algorithm, designed for detecting anomalies in time series in an unsupervised mode [20-22]. It focuses on differences in associations between normal and anomalous points in a time series, efficiently identifying deviations [24-26]. Inside trasnformer Anomaly-Attention Discrepancy Analysis is used, which compares the importance of each point in the time series with its neighbors. The main idea is that if the attention given to the current point significantly deviates from expectations (compared to its "normal" neighbors), this point may be considered anomalous. The Anomaly Transformer employs two key components [26-28]:
    - Attention Map (A): Represents a matrix that quantifies the relationship between each time step in the sequence based on similarity scores between query and key vectors. Each element (A(i,j)) is computed using a scaled dot-product attention mechanism, where the similarity between a query vector at time step (Q_i) and a key vector at time step (K_j) is determined by their dot product ($Q_i K_j^T$). This similarity score is scaled by the square root of the key dimensionality (d_k) to maintain numerical stability and is then passed through a softmax function to ensure proper weighting across all time steps [29-30].

$$A(i,j) = Softmax(\frac{Q_i K_j^T}{\sqrt{d_k}}) \quad (2)$$

- Loss-function (attention discrepancy analysis): Loss-function is divided into 2 components [31]:

1) $\mathcal{L}_{recon}$: minimizes the reconstruction error of the time series.

2) $\mathcal{L}_{prior}$: measures the discrepancy between the attention distribution of the current point and the expected distribution.

Common loss function with the balancing coefficient(λ):

$$\mathcal{L} = \mathcal{L}_{recon} + \lambda\mathcal{L}_{prior} \quad (3)$$

4. The output of both algorithms are called the anomaly scores (anomaly_scores) and assigned to each point/event in the time series based on its degree of anomaly gained in previous steps. These scores are normalized(normal_scores) to fall within the range [0, 1] using the minimum (min_score) and maximum (max_score) possible scores:

$$normal\_scores = \frac{anomaly\_scores - min\_score}{max\_score - min\_score} \quad (4)$$

5. The results for each point/event (normal_scores) are then converted into percentage probabilities (probabilities_percent):

$$probabilities\_percent = normal\_scores \times 100 \quad (5)$$

The result of the algorithms (P) work is the half sum of the anomaly probability related to the user's normal behavior (P_e) and the sequential probability for each point/event (P_s).

$$P = \frac{P_e + P_s}{2} \quad (6)$$

6. The results are represented in a form of event id from a table “Events”, the name of an event, its independent probability (P_e) obtained by Isolation Forest algorithm usage, its sequential probability (P_s) and a combined probability (P). Those outputs are shown as separate rows in a console [32].

7. The forensic specialist makes the final decision.

   1) Based on the provided probability percentages and data, the specialist decide whether to consider a digital indicator or not: if it falls below a predefined threshold and lacks clear relevance to the crime, the result is disregarded [19]. Otherwise, it can be further investigated and accepted as evidence.

   2) However, even if the probability is high but the anomaly and event list is incomplete or irrelevant, the specialist rejects the result or perform further analysis.

   3) If the indicator aligns with the specialist's plan and includes comprehensive correlated data, the result is accepted as evidence.

Numerical Experiments

A study was conducted to evaluate the performance of artificial intelligence and the expert model. The experiment consisted of following steps:

1. Simulating the operation of a smart home ecosystem. Initially, a database simulating the smart home ecosystem's database was created and populated. Tables for events, users, devices, and schedules were set up with the data which was artificially generated. Each 10 seconds the database received data describing a “signal” from one of 7 device used in experiment. Also, each device had different options for an event type which helped with the diversity of data. The devices in simulation included:

- Thermostat
- Light Bulb
- Camera
- Doorbell
- Door lock
- Smoke detector
- Smart Assistant

2. Artificial intelligence operation. Once the database was populated, the artificial intelligence processes were initiated, sequentially employing two algorithms: Isolation Forest and Anomaly Transformer.

Results of Isolation Forest: The output includes the percentage probability of an event's anomaly relative to the user's normal behavior as reflected in the schedules table. The results consist of the event's index in the database, the event action, and its probability.

```
ai_digital_forensics    | {'id': 1, 'event': 'Light turned on', 'event_probability': '93.55%'}
ai_digital_forensics    | {'id': 2, 'event': 'Speaker played music', 'event_probability': '0.68%'}
ai_digital_forensics    | {'id': 3, 'event': 'Speaker played music', 'event_probability': '0.68%'}
ai_digital_forensics    | {'id': 4, 'event': 'Speaker played music', 'event_probability': '0.68%'}
ai_digital_forensics    | {'id': 5, 'event': 'Camera activated', 'event_probability': '0.0%'}
ai_digital_forensics    | {'id': 6, 'event': 'Lock engaged', 'event_probability': '30.86%'}
ai_digital_forensics    | {'id': 7, 'event': 'Battery low', 'event_probability': '56.47%'}
ai_digital_forensics    | {'id': 8, 'event': 'Lock engaged', 'event_probability': '30.86%'}
ai_digital_forensics    | {'id': 9, 'event': 'Light turned on', 'event_probability': '93.55%'}
```

Fig.2 - Results of Isolation Forest

Results of Anomaly Transformer: The output is the percentage probability of a sequence of events being anomalous relative to other events. Events are divided into sequences of 21 elements and displayed as the sequence index and its probability.

```
ai_digital_forensics    | {'sequence_index': 0, 'sequence_probability': '0.0%'}
ai_digital_forensics    | {'sequence_index': 1, 'sequence_probability': '100.0%'}
```

Fig.3 - Results of Anomaly Transformer

Final stage: Processing results with a mathematical model. In this step, each event was evaluated individually. Based on the event's index, its relation within a specific sequence and its probability were determined. Using the sequence probability, the mathematical model calculated the final result for each point.

```
ai_digital_forensics    | {'id': 1, 'event': 'Light turned on', 'event_probability': '93.55%', 'sequence_probability': '0.0%', 'combined_probability': '46.77%'}
ai_digital_forensics    | {'id': 2, 'event': 'Speaker played music', 'event_probability': '0.68%', 'sequence_probability': '0.0%', 'combined_probability': '0.34%'}
ai_digital_forensics    | {'id': 3, 'event': 'Speaker played music', 'event_probability': '0.68%', 'sequence_probability': '0.0%', 'combined_probability': '0.34%'}
ai_digital_forensics    | {'id': 4, 'event': 'Speaker played music', 'event_probability': '0.68%', 'sequence_probability': '0.0%', 'combined_probability': '0.34%'}
ai_digital_forensics    | {'id': 5, 'event': 'Camera activated', 'event_probability': '0.0%', 'sequence_probability': '0.0%', 'combined_probability': '0.0%'}
ai_digital_forensics    | {'id': 6, 'event': 'Lock engaged', 'event_probability': '30.86%', 'sequence_probability': '0.0%', 'combined_probability': '15.43%'}
ai_digital_forensics    | {'id': 7, 'event': 'Battery low', 'event_probability': '56.47%', 'sequence_probability': '0.0%', 'combined_probability': '28.23%'}
ai_digital_forensics    | {'id': 8, 'event': 'Lock engaged', 'event_probability': '30.86%', 'sequence_probability': '0.0%', 'combined_probability': '15.43%'}
```

Fig.4 - Final results

CONCLUSIONS

In this study, a methodology for using artificial intelligence (AI) to establish an evidence base for crimes committed using IoT devices in smart homes is proposed. The developed system includes the stages of obtaining data from smart home ecosystem databases, their preprocessing, analysis using AI, finalizing of results using a mathematical model and presenting the results in a readable clear form for specialists. The proposed approach allows to effectively identify anomalous actions and sequences of events that can serve as evidence for crimes committed using IoT devices. The analysis results are provided in percentage terms, reflecting the likelihood of events being related to offences, which facilitates the process of establishing an evidence base for specialists. The system developed contributes to the efficiency of investigations in the context of the growing proliferation of smart home technologies and the Internet of Things. Finally, this research presented an innovative methodology based on the use of artificial intelligence (AI) and mathematical modelling to analyze data generated by smart home ecosystems.

Secondly, the conducted experiment has shown that the results of the AI resources are plausible and can be valuable at this point of research. The model successfully processed the data bringing normalized values within the expected

range. The algorithms within the model had no difficulties with calculating the probabilities with a small sample of data accessible. Results are logical and understandable from a human perspective.

NOVELTY

The novelty of the work lies in the development of an integrated methodology that combines artificial intelligence (AI) and expert evaluation for analyzing data from smart home ecosystems to establish an evidence base for crimes committed using IoT devices.

Unlike existing studies, which primarily focus on individual aspects such as data collection or anomaly detection, the presented methodology covers the entire cycle: from obtaining data from smart home databases to their analysis using AI and presenting the results in a form convenient for specialists. This allows for more effective detection and interpretation of anomalous actions and sequences of events potentially related to crimes, which was not fully implemented in similar systems before.

The proposed expert (mathematical) model enables the quantitative assessment of the likelihood that certain events recorded in smart home ecosystems are related to criminal activity. This model ensures objectivity and accuracy in the process of forming an evidence base, which represents a significant contribution to the field of digital forensics.

RESULTS

The study has yielded the following results:

1. Development of methodology for the use of AI. The system approach to data processing and analysis from smart house ecosystems has been developed, including the stages of data acquisition, primary processing, analysis and basic visualization.
2. The primary AI model has been developed and trained, capable of analyzing smart home data and detecting anomalous actions potentially related to criminal activity.
3. Creating a mathematical model based on the assumption of equilibrium of both AI algorithms for evaluation of evidence: A model has been developed that allows to estimate the probability that certain events are related to offences, which facilitates the process of establishing an evidentiary basis. At this moment the results of the mathematical model is the half sum of the results of two AI algorithms.

Thus, the proposed methodology and developed tools demonstrate high potential in analyzing smart home data to identify potential crimes, It opens up new opportunities for digital forensics and security in the face of growing IoT devices.

FUTURE WORK

Future research are planned to be focused on the following areas:

1. Improving artificial intelligence (AI) performance. Optimization of machine learning algorithms to increase the accuracy and speed of data analysis from smart house ecosystems is prioritized. Special attention will be given to the adaptation of models to the specifics of data coming from different IoT devices, as well as the introduction of self-learning methods to improve AI's adaptability to new types of threats. Moreover, the AI model processing will be enhansted by using additional parameters of decision. In future iterations the model should not only compare events by its devices and types but also consider the time frames, changes of user normal behaviour, and the frequency of events happening.
2. Development of methods for evaluating the effectiveness of AI. Systematic approaches to AI performance measurement, including metrics on accuracy, completeness and response time, are planned. This allows an objective measurement of the effectiveness of the proposed solutions and identify areas for further improvement.
3. Improvement of the mathematical model for the final result. More sophisticated models are expected to be developed that can take into account multiple factors and their interrelationships in assessing the likelihood of criminal acts. This will improve the reliability of the conclusions and reduce the number of false responses. This task includes considering the types of events, time frames, the influence of one algorithm to another. The result of impoving mathematical model will consist of number of rules and cases considering which the model should give more accurate results.
4. Visualisation improvement. Part of future work will be dedicated to building a user friendly design. The console version will be transformed into the intuitive interface. It will have expanded functionallities among which the

open data review. There a user can look at the analized data upclose and understand the logic behind given numbers. The result of the system will be supported by the intermediate results of algorithms, records of database and characteristics influencing the outputs of mathematical model.

5. Enhanced experiments. Further development of the forensics system will require more sophisticated testing. This will include wider devices range, types of events and more datasets varieties. With this goal, the datasets will be formed to cover different crime cases expecting different results from the system measuring its accuracy. Moreover, it is planned to organize the work of real smart home to make it possible to test the system in real-life environment.

The implementation of these directions will create a more effective and reliable system for establishing an evidence base for crimes committed with IoT devices in smart homes.

REFRENCES

[1] Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.

[2] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems, 29*(7), 1645-1660.

[3] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys, 41*(3), 1-58.

[4] Ashton, K. (2009). That "Internet of Things" thing. *RFID Journal*.

[5] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems, 29*(7), 1645-1660.

[6] Zanella, A., Bui, N., Castellani, A., Vangelista, M., & Zorzi, M. (2014). Internet of Things for smart cities. *IEEE Internet of Things Journal, 1*(1), 22-32.

[7] Miorandi, D., Sicari, S., De Pellegrini, F., & Chlamtac, I. (2012). Internet of Things: Vision, applications and research challenges. *Ad Hoc Networks, 10*(7), 1497-1516.

[8] Mann, D., & Frolow, M. (2015). *Smart homes and the Internet of Things: Opportunities, trends, and issues*. Springer.

[9] How much data does a smart home generate? (n.d.). Retrieved from https://finance.yahoo.com/news/much-data-smart-home-generate-083203809.html.

[10] Salamh, F. E. (2020). A forensic analysis of home automation devices (FAHAD) model: Kasa smart light bulb and Eufy floodlight camera as case studies. *International Journal of Cyber Forensics and Advanced Threat Investigations*.

[11] Solanke, A. A. (2022). Digital forensics AI: Evaluating, standardizing, and optimizing digital evidence mining techniques. *KI - Künstliche Intelligenz*.

[12] Salem, Y., & Owda, A. Y. (2024). Towards Digital Forensics 4.0: A multilevel digital forensics framework for Internet of Things (IoT) devices. *International Journal of Wireless and Microwave Technology*.

[13] Chakhkiev, M. T., Tramov, I. B., & Gasanov, N. (2021). Databases in the software and hardware complex for managing a smart home system. *StudNet, 4*(3).

[14] Kafle, K., Moran, C., Mandhar, S., Nadkarni, A., & Poshivanik, D. (2019). A study of data store-based home automation. *Proceedings of the Ninth ACM Conference on Data and Application Security and Privacy*.

[15] Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. *Proceedings of the 2008 IEEE International Conference on Data Mining (ICDM)*.

[16] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys, 41*(3).

[17] Filonov, P., et al. (2020). *Anomaly detection principles and algorithms*. ISBN: 978-1733455934.

[18] Bishop, C. M. (2006). *Pattern recognition and machine learning*.

[19] Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. ISBN: 978-1492032649.

[20] Zong, B., Song, L., & Ye, J. (2018). Deep autoencoding Gaussian mixture model for unsupervised anomaly detection. *Proceedings of the 2018 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 118-127.

[21] Dai, X., & He, X. (2018). Anomaly detection with LSTM autoencoders. *2018 IEEE International Conference on Big Data (Big Data),* 1485-1494.

[22] Li, X., & Zhang, L. (2020). Anomaly transformer: A transformer-based framework for anomaly detection in time series. *Proceedings of the 2020 Conference on Neural Information Processing Systems (NeurIPS).*

[23] Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications, 60*, 19-31.

[24] Bergmann, P., Fauser, B., Sattlegger, D., & Sick, B. (2019). Intrinsic outlier detection with LSTMs for one-class classification. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV),* 6557-6566.

[25] Xia, F., & Liu, X. (2021). A survey on anomaly detection for time series data. *IEEE Access, 9*, 22258-22272.

[26] Kiran, B. R., & Reddy, R. K. (2020). Time series anomaly detection using transformer-based models. *International Conference on Computational Intelligence and Data Science (ICCIDS),* 333-343.

[27] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR), 41*(3), 1-58.

[28] Brownlee, J. (2019). *Machine learning mastery with Python: Understand your data, create accurate models, and work projects end-to-end*. Machine Learning Mastery.

[29] Li, J., & Li, W. (2019). Attention-based anomaly detection in time series data. *arXiv:1904.04203*.

[30] Choi, E., Bahadori, M. T., & Sun, J. (2017). A survey of deep learning for anomaly detection. *arXiv:1709.03947*.

[31] Gupta, A., Jones, M., & Taylor, K. (2019). Applying machine learning techniques for anomaly detection in IoT data. *IEEE Internet of Things Journal*.

[32] Johnson, T. (2018). Probabilistic inference in artificial intelligence systems: Fundamentals and applications in forensics. *ACM Computing Surveys*.