SUSTAIN Deliverable

D6.1 Behavioural biometrics based authentication software system design

Grant Agreement number	101071179
Action Acronym	SUSTAIN
Action Title	Smart Building Sensitive to Daily Sentiment
Type of action:	HORIZON EIC Grants
Version date of the Annex I against which the assessment will be made	28th March 2022
Start date of the project	1st October 2022
Due date of the deliverable	M15
Actual date of submission	12th December 2023
Lead beneficiary for the deliverable	AALTO
Dissemination level of the deliverable	Public

Action coordinator's scientific representative

Prof. Stephan Sigg
AALTO –KORKEAKOULUSÄÄTIÖ,
Aalto University School of Electrical Engineering, Department of Communications and Networking
stephan.sigg@aalto.fi

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Innovation Council and SMEs Executive Agency (EISMEA). Neither the European Union nor the granting authority can be held responsible for them.

European Innovation Council

Authors in alphabetical order		
Name	**Beneficiary**	**e-mail**
Name (from A's to Z's)	SHORT name of organization	name.name@org.de
Kahraman, Nihan	YTU	nicoskun@yildiz.edu.tr

Abstract

This deliverable presents the SUSTAIN project Task 6.1 work conducted during M1-M15 on the "Behavioural biometrics-based authentication software system" design. This system has been tailored for the operation on constrained distributed nodes in the building automation system. The aim was low FAR.

The YTU team designed a behavioural biometrics based authentication system for cryptographic key generation. In particular, the system was be tailored for the operation on constrained distributed nodes in the building automation system, which consists of a central coordinator (building automation controller) and many distributed participants (distributed sensor nodes). The team considered, in particular, multi-key scenarios and stressed low computational demand. Target FAR error rate and evolution time were required to be low.

The work will continue in T6.1 in joint work of AALTO and YTU, and the aim will be the designing of the hardware part of the authentication system for cryptographic key generation.

Contents

1 Introduction and purpose of this deliverable

This report is the first deliverable of Work Package (WP) 6: “Security”. The overall goal of WP6 is to design behavioural biometrics based secure encryption system using Garbled Circuit Protocol.

The purpose of this deliverable is to report the design of behavioural biometrics based authentication system for cryptographic key generation. The structure of the deliverable is as follows. First, we will give an overview of the behavioural biometrics based authentication system in SUSTAIN. Then, we will detail our designed behavioural biometrics based authentication software system and we will list challenges in behavioural biometrics based authentication. Finally, we will further detail our future strategies for software optimisation and for authentication hardware development in the project.

2 Overview of behavioural biometrics based authentication system in SUSTAIN

In this deliverable, the software of the behavioural biometrics based authentication system recommended for SUSTAIN has been implemented. The software's framework is depicted in Figure 1. Here, data is collected from several ambient sensors, and human activity recognition is conducted based on this data. Then, room biometric patterns were created by analyzing the activities of the rooms over time. Subsequently, the system identifies the room based on the input, which comprises the outcome of the activity and the time information derived from the analysis of the sensor data. In the future, the authentication system from this deliverable output will be used to generate particular keys for the recognized rooms.

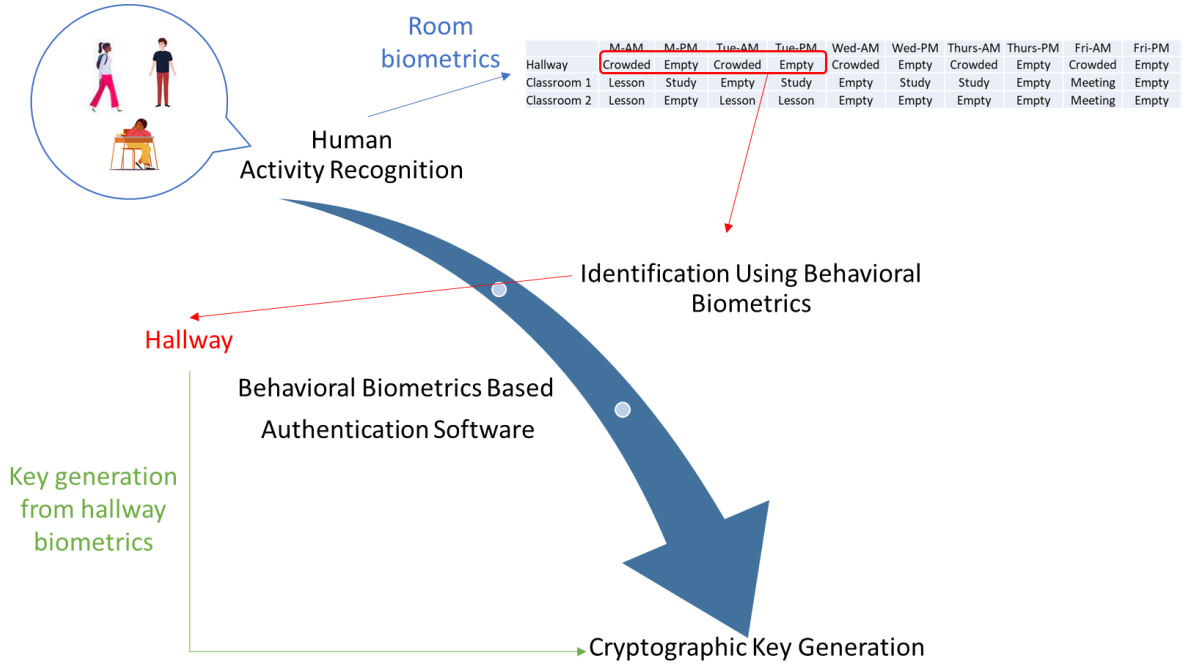

Figure 1. Behavioural biometrics based authentication system framework

3 Designed behavioural biometrics based authentication software system

The studies conducted during the development of the authentication software based on behavioural biometrics in SUSTAIN, can be categorized into three primary categories:

- **Human Activity Recognition (HAR) studies using ambient sensor data:**
  During these studies human activity recognition from continuous ambient sensor data set [1] was use for HAR. The provided dataset includes a collection of activity labels that can be utilized for tasks such as categorizing and tracking the actions of individuals. The dataset also encompasses ambient data gathered from the homes of volunteer residents, with data collection occurring constantly as inhabitants carry out their regular activities. Volunteer houses were equipped with ambient passive infrared (PIR) motion sensors, door/temperature sensors, and light switch sensors to collect data. In the experiments, feature vectors in the dataset was arranged using a sliding window of 30 sensor events from the raw data. The labeled activities were then classified.

  An ensemble learning approach was used with the aim of enhancing the recognition performance. Ensemble learning is a technique in machine learning that seeks to improve outcomes by integrating numerous learning algorithms. Temporal Convolutional Networks (TCN), Long short-term memory (LSTM), and Gated recurrent unit (GRU) methods were selected to be used with the ensemble learning method and voting was performed by weighting the class probabilities before averaging. All deep neural network (DNN) models were trained separately for each processed data subset of 30 different houses. The data for each subset was divided into 70% training and 30% test sets, and 20% of the training set was used for validation. Model training time was determined as 200 epochs and 50 epoch patience value was used to prevent overfitting.

  The performance of the proposed ensemble method was compared with three frequently used deep learning methods and the results are shown in Figure 2. From these results, it can be seen that the HAR performance achieved with ensemble learning is higher compared to classical DNN methods.

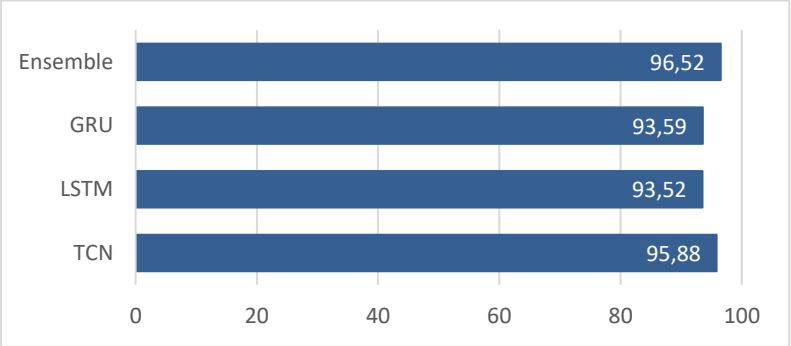

Figure 2. Average HAR accuracy of DNN models.

- **Sensor network setups and data set collection:**
  Sensor networks were deployed in several rooms to categorize activities and gather room biometrics. Subsequently, tests were conducted utilizing the data acquired from these sensors. Various sensor configurations were put in two distinct rooms, each serving various functions. Temperature, humidity, motion, and door sensors were used to collect data in the rooms at 10-minute intervals

over a period of 20 days. Figures 3 and 4 display the room plans and sensor positions for the purpose of examining the sensor placements in the rooms. Additionally Figures 5-8 shows the placed sensors in the rooms.

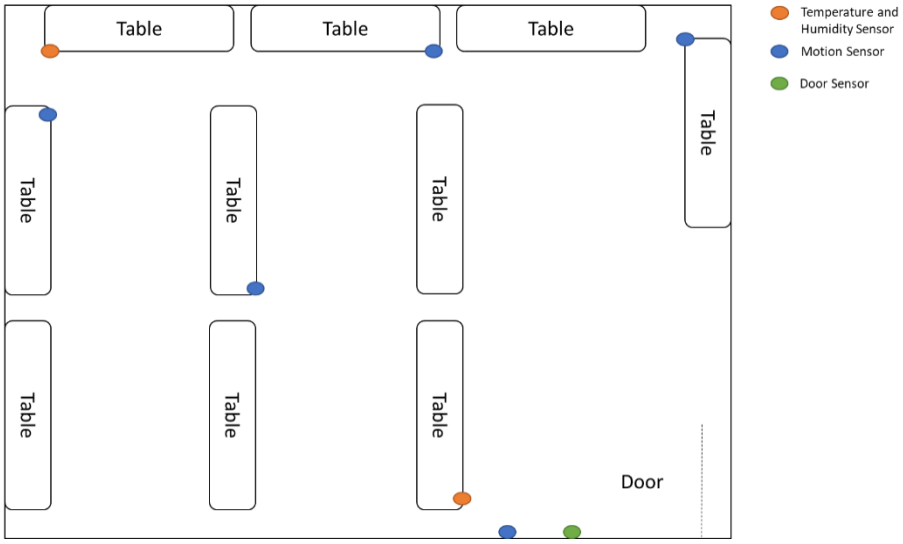

Figure 3. Room plans and sensor positions for Room 1

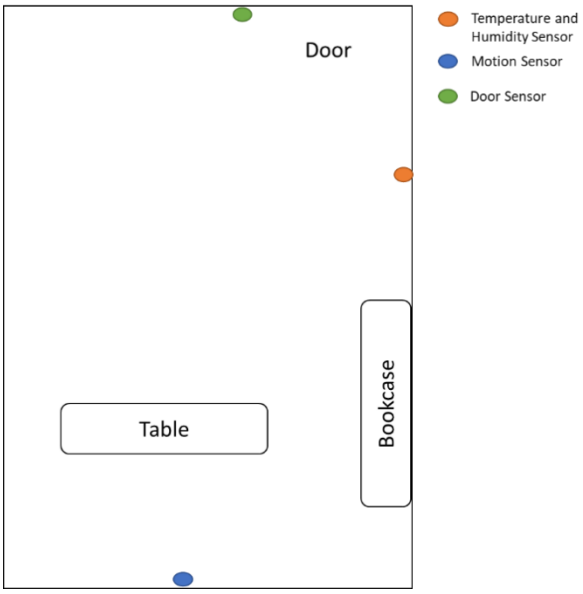

Figure 4. Room plans and sensor positions for Room 2

Figure 5. Motion sensor 1 in Room 1

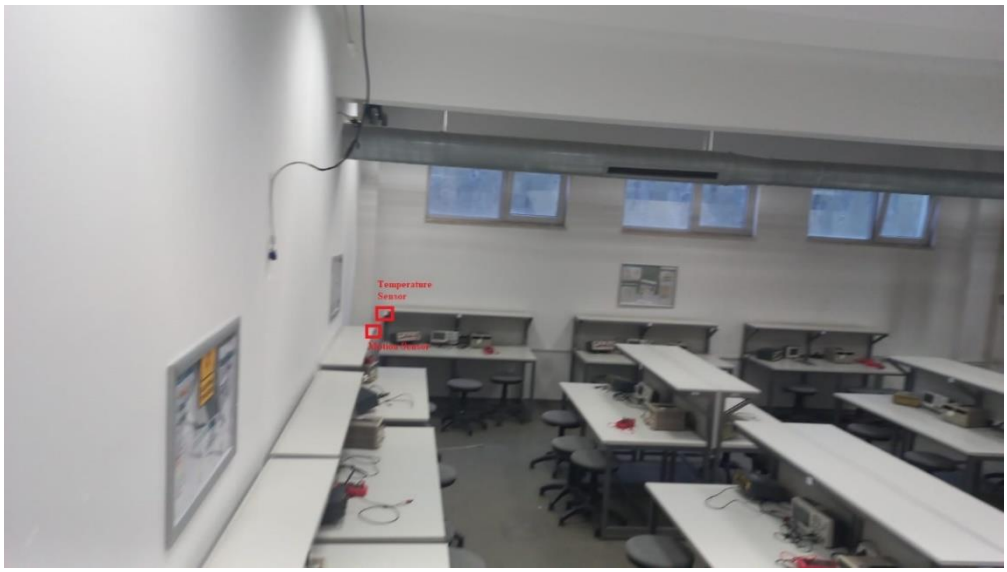

Figure 6. Temperature and motion sensor in Room 1

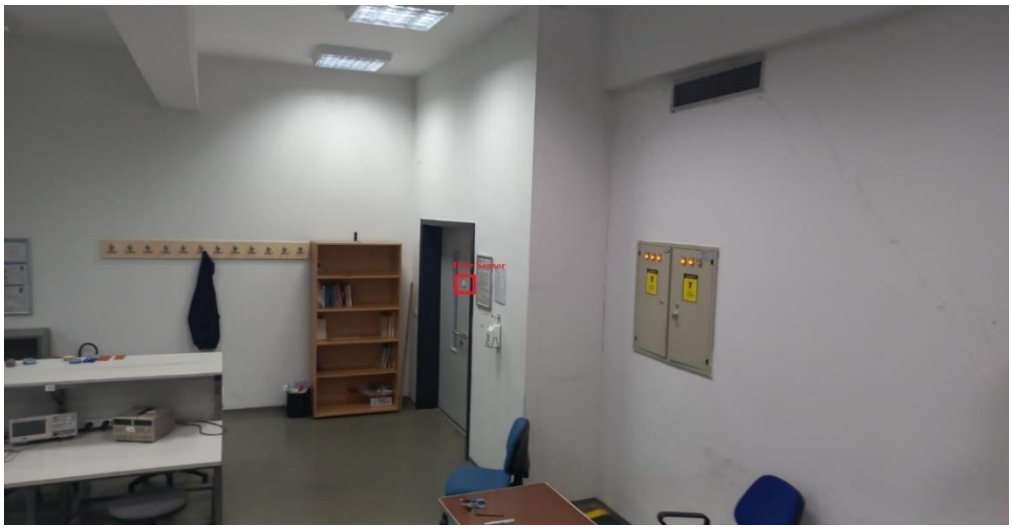

Figure 7. Door sensor 1 in Room 1

Figure 8. Motion sensor 2 in Room 1

The data set was generated using wireless Zigbee sensors, which were specifically chosen for this purpose. The collected data is then uploaded to the computer through wifi using the sensor hub. Subsequently, the data was systematically arranged and categorized based on the corresponding actions. The activity labels for Room 1 were "Lab", "Study", and "Empty". Room 2, being an personel office room, uses the labels "Study" and "Empty". The Zigbee sensor network setup is also given in Figure 9.

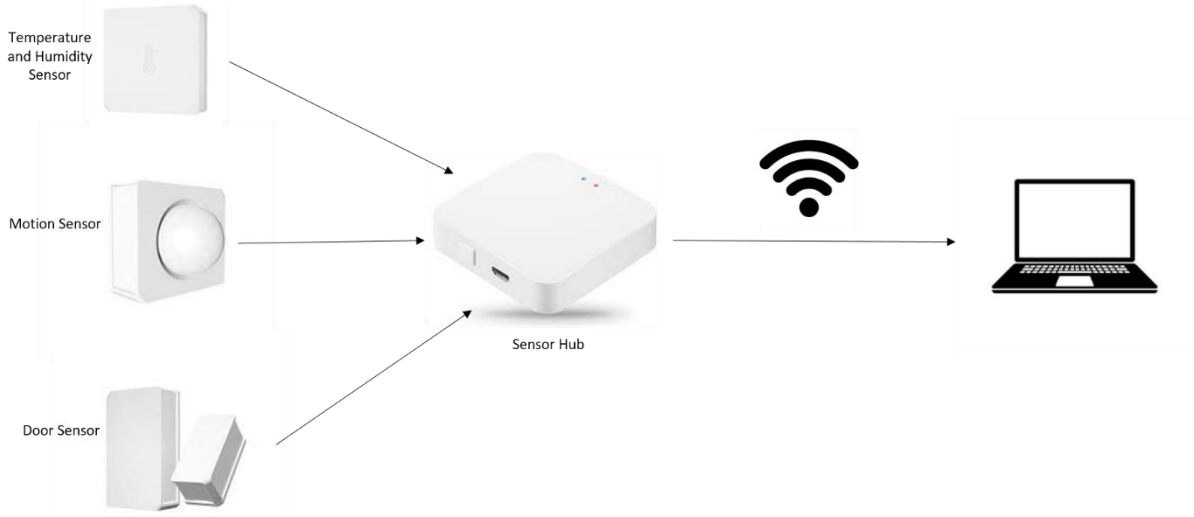

Figure 9. Sensor network setup

- **Room behavioural biometrics based authentication using collected sensor data:**
  During studies on room behavioural biometrics based authentication, activity classification was initially performed using data collected from the sensors located in the rooms. The classes for the preliminary tests were determined as "lesson", "study", and "empty". The performance of activity classification was evaluated using several classifiers. Subsequently, another classification technique

was devised that used the evaluation of time and day information to determine the room. This was achieved by utilizing the outputs of the activity classification models specific to each room. The framework of room authentication is given in Figure 10.

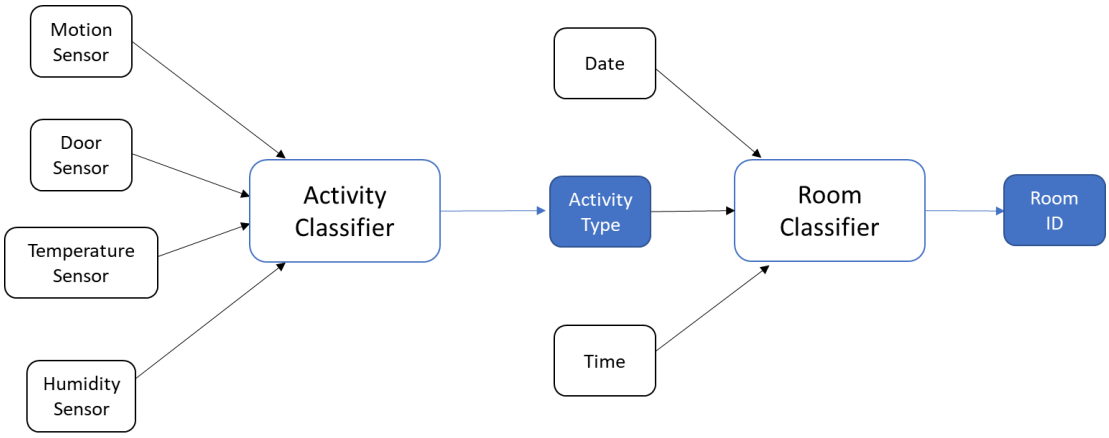

Figure 10. Framework of room behavioural biometrics based authentication

In the studies carried out for human activity recognition study and room biometry extraction, the data received from the sensors were first arranged and normalized. Later, conventional machine learning algorithms and various classifiers available in the literature were used to perform the classification process. In the following section, detailed information is given about the classifiers used within the scope of the project.

AdaBoost Classifier: Adaboost or Adaptive Boosting algorithm is an ensemble boosting classifier that combines multiple classifiers to increase the accuracy of classifiers and aims to obtain a powerful classifier from lower performing classifiers. It was first proposed by Yoav Freund and Robert Schapire in 1996. The training sets used in this method must have various weights and interact. In its working principle, it tries to achieve maximum performance by assigning higher weights to both incorrectly classified data and classifiers with high performance.

Bagging Classifier: Bagging Classifier method is the general name given to classifiers that basically create subsets of the data in the dataset and assign a basic classifier to each subset and then use methods such as voting or averaging.

k-Nearest Neighbour Classifier: k-Nearest Neighbor classifier works as a classifier based on voting and distances in the feature space. In the created system, it is decided which class the new data belongs to by considering the location in space of the representations obtained from the data in the data set. The new data is assigned to the class with the most votes by looking at the k data classes located nearby a newly introduced data. Although Euclidean distance is commonly used as a distance measurement, different distance calculations can also be used for this classifier.

Nearest Centroid Classifier: In this classifier, unlike the k-NN classifier, the centroid of each class is taken as the basis for classification and the new data is assigned to the class that is closest to the centroid of the class.

Logistic Regression: Logistic Regression is a regression method used for classification. It is used to classify categorical or numerical data. It works when the dependent variable, that is, the result, can only take 2 different values. It is widely used in linear classification problems. For this reason, it is very similar to Linear Regression. The biggest difference between logistic regression and linear regression is the way it uses the line that separates the two classes. While linear regression uses the "Least Squares Method" to draw the optimal line, logistic regression uses the "Maximum Likelihood" method.

Linear Discriminant Analysis: Linear Discriminant Analysis (LDA) is used in feature reduction processes as well as as a classifier. LDA Classifier is a linear classifier that uses the Bayes rule and fits class conditional densities to the data set. Assuming that each class shares the same covariance matrix, the classifier determines a Gaussian distribution for each class and performs the classification process using this distribution.

Quadratic Discriminant Analysis: Quadratic Discriminant Analysis Classifier is a classifier that uses the Bayes rule and fits class conditional densities to the data set. Assuming that each class shares the same covariance matrix, the classifier determines a Gaussian distribution for each class. Unlike LDA, this classifier uses a quadratic decision boundary.

Decision Tree: A decision tree is a flowchart-like diagram that maps all potential solutions to a particular problem. It is often used by organizations to help determine the optimal course of action by comparing all the possible outcomes of making a series of decisions. Its structure consists of decision nodes and leaf nodes. They have a predefined target variable. Due to their structure, they offer a strategy that goes from top to bottom.

Extra Trees Classifier: It is a type of classifier that consists of randomized decision trees and these decision trees are applied to sub-data sets in the data set. The averaging process used in the method aims to increase the performance of the system and solve the overfitting problem.

Random Forest: Random Forest (RF) is a machine learning algorithm used to solve regression and classification problems that combines the output of multiple randomly generated decision trees. The algorithm aims to increase the classification value during the classification process by producing more than one decision tree. Random forest algorithm is the process of combining many decision trees that work independently of each other and selecting the value with the highest score among them. The main difference between the decision trees algorithm and random forest is that in the random forest algorithm, the process of finding the root node and splitting the nodes is random.

Support Vector Machines Classifier: Support Vector Machines (SVMs) are one of the supervised learning methods generally used in classification problems. Support Vector Machines (SVMs) are an algorithm that can be used for regression and outlier detection, apart from classification problems. Draws a line to separate points placed on a plane. It aims to have the maximum distance for the points of both classes of this line. It is suitable for complex but small and medium-sized data sets. While Support Vector Machines (SVMs) show high performance in solving linear data in classification problems, they can also achieve high performance in classifying non-linear data by using different kernel types.

EXtreme Gradient Boosting (XGM) Classifier: XGBoost uses a gradient-based sampling method to select data instances for each tree, which enhances training efficiency and reduces the impact of noisy data. To prevent overfitting and improve generalization performance, L1 and L2 regularization are employed to regulate the complexity of the decision trees, and a small learning rate is added to the predictions of each tree to reduce its impact on the final predictions.

Light Gradient Boosting (LGBM Classifier): LightGBM is optimized for the fast and efficient processing of large-scale datasets with high-dimensional features. It groups highly correlated features into exclusive feature bundles, which results in fewer splits and improved efficiency.

Both XGBoost and LightGBM use a histogram-based approach to select the best split points for tree construction, which involves binning feature values into discrete intervals (histograms) and selecting the optimal split points based on the histograms, rather than considering every possible split point. This approach can significantly reduce the time and memory required for tree building, especially for large datasets.

Stochastic Gradient Descent (SGD) Classifier: All classifiers that use the SGD algorithm are called SGD Classifiers. The important thing is that the linear classification algorithm uses SGD during its training.

Ridge Classifier: This classifier first transforms the binary targets into {-1, 1} and then optimizes the same target as above by treating the classification problem as a regression task. The predicted class is the output that the regressor gives as a prediction. For multi-class classification, the problem is treated as multiple output regression and the predicted class corresponds to the output with the highest value.

Perceptron: Perceptron is another simple classification algorithm that is suitable for large-scale learning, does not require a learning rate, does not need to be regularized, and updates its model only in case of errors.

Passive Aggressive Classifier: This classifier, which is similar to the perceptron classifier in that it does not require a learning rate and is suitable for large-scale learning, differs from the perceptron in its need for regularization.

Gaussian Naive Bayes (Gaussian NB) Classifier: Naive Bayes classifiers are simple models based on probability theory that can be used for classification. They arise from the assumption of independence between input variables. Although this assumption is not true in the vast majority of cases, they generally perform very well on most classification tasks, so they are quite popular. Gaussian Naive Bayes incorporates another (often false) assumption: that the variables represent a Gaussian probability distribution. Although it is difficult to accept that so many incorrect assumptions can lead to such good performances, it is a very good working classifier and is widely used by researchers.

Bernoulli Naive Bayes (Bernoulli NB) Classifier: Bernoulli NB implements the naive Bayes training and classification algorithms for data that is distributed according to multivariate Bernoulli distributions.

Dummy Classifier: Dummy classifier is an algorithm that is generally used to compare with other classifiers and performs classification by ignoring input values.

Evaluation metrics commonly used in the literature were used to observe the performance of the installed systems. Descriptions and calculations of the metrics used in model evaluations for this deliverable are provided below.

Accuracy: It is an evaluation metric that gives the ratio of data correctly predicted by the model in a data set to the whole data set.

$$Accuracy = \frac{True\ Predicted\ Data}{All\ Data}$$

Precision: In solving a classification problem, precision equals the number of true positives divided by the total number of data labelled as belonging to the positive class.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall: In solving a classification problem, recall equals the number of true positives divided by the the total number of elements that actually belong to the positive class (True Positive + False Negative).

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Balanced Accuracy: The accuracy metric alone is not sufficient, especially in studies using unbalanced data sets. In these cases, a metric that averages the recall value obtained for each class is used to measure system performance. The name of this metric is balanced accuracy.

$$Balanced\ Accuracy = mean\left(\sum_{i=1}^{Number\ of\ Classes} \frac{Relevant\ Retrieved\ Instances(i)}{All\ Retrieved\ Instances(i)}\right)$$

F-Score: The metric in which precision and recall metrics are used and the harmonic average of the two is taken is called the F-Score metric.

$$F\ Score = 2\ x\ \frac{Precision\ x\ Recall}{Precision + Recall}$$

In the experiments carried out, activity recognition was first made for two different rooms with different sensor setups, using instantaneous data collected at 10-minute intervals. The classification performance analysis results obtained using various classifiers and evaluation metrics in the tests carried out using the collected sensor data for Room 1 and Room 2 are given in Table 1-2, respectively.

Table 1. Classification performance analysis for Room 1 using instantaneous data

Method	Accuracy (%)	Balanced Accuracy (%)	F1 Score (%)
LGBMClassifier	100.00	100.00	100.00
XGBClassifier	100.00	100.00	100.00
BaggingClassifier	100.00	100.00	100.00
ExtraTreesClassifier	99.88	99.57	99.88
RandomForestClassifier	99.88	99.57	99.88
DecisionTreeClassifier	99.65	98.70	99.66
LabelPropagation	99.07	95.20	99.06
LabelSpreading	99.07	95.20	99.06
ExtraTreeClassifier	99.07	94.32	99.06
QuadraticDiscriminantAnalysis	97.22	88.89	97.27
CalibratedClassifierCV	97.45	86.75	97.43
SVC	97.57	86.74	97.53
KNeighborsClassifier	97.57	86.35	97.51
Perceptron	96.99	85.41	97.00
LinearSVC	97.22	85.00	97.18
LinearDiscriminantAnalysis	96.64	84.61	96.69
PassiveAggressiveClassifier	96.76	84.55	96.76
LogisticRegression	97.11	84.12	97.05
GaussianNB	96.18	83.21	96.29
SGDClassifier	96.88	81.92	96.76
BernoulliNB	95.83	80.19	95.93
AdaBoostClassifier	95.60	79.50	95.63
RidgeClassifier	96.41	78.86	96.23
RidgeClassifierCV	96.41	78.86	96.23
NearestCentroid	94.21	75.85	93.90
DummyClassifier	86.69	33.33	80.51

Table 2. Classification performance analysis for Room 2 using instantaneous data

Method	Accuracy (%)	Balanced Accuracy (%)	F1 Score (%)
LinearSVC	99.73	98.95	99.73
PassiveAggressiveClassifier	99.73	98.95	99.73
BernoulliNB	99.73	98.95	99.73
CalibratedClassifierCV	99.73	98.95	99.73
SVC	99.73	98.95	99.73
SGDClassifier	99.73	98.95	99.73
GaussianNB	99.73	98.95	99.73
KNeighborsClassifier	99.73	98.95	99.73
LabelPropagation	99.73	98.95	99.73
LabelSpreading	99.73	98.95	99.73
QuadraticDiscriminantAnalysis	99.73	98.95	99.73
Perceptron	99.73	98.95	99.73
LogisticRegression	99.73	98.95	99.73
NearestCentroid	99.73	98.95	99.73
XGBClassifier	99.60	98.87	99.60
RandomForestClassifier	99.60	98.87	99.60
AdaBoostClassifier	99.60	98.87	99.60
BaggingClassifier	99.60	98.87	99.60
ExtraTreesClassifier	99.60	98.87	99.60
ExtraTreeClassifier	99.60	98.87	99.60
DecisionTreeClassifier	99.60	98.87	99.60
LGBMClassifier	99.60	98.87	99.60
LinearDiscriminantAnalysis	98.65	91.10	98.60
RidgeClassifier	98.65	91.10	98.60
RidgeClassifierCV	98.65	91.10	98.60
DummyClassifier	93.13	50.00	89.81

After instantaneous data experiments are carried out, activity recognition was then made for two different rooms with different sensor setups, using data collected at 10-minute intervals for the last 60 minutes. The classification performance analysis results obtained using various classifiers and evaluation metrics in the tests carried out for Room 1 and Room 2 are given in Table 3-4, respectively.

Table 3. Classification performance analysis for Room 1 using 60 minutes data

Method	Accuracy (%)	Balanced Accuracy (%)	F1 Score (%)
LGBMClassifier	100.00	100.00	100.00
BaggingClassifier	100.00	100.00	100.00
XGBClassifier	100.00	100.00	100.00
DecisionTreeClassifier	100.00	100.00	100.00
RandomForestClassifier	99.88	98.99	99.88
ExtraTreesClassifier	99.54	97.58	99.54
ExtraTreeClassifier	98.73	94.42	98.76
LabelPropagation	98.84	94.12	98.84
LabelSpreading	98.84	94.12	98.84
SGDClassifier	98.15	92.91	98.17
LogisticRegression	98.61	92.73	98.60
QuadraticDiscriminantAnalysis	97.22	92.52	97.30
LinearSVC	98.49	92.37	98.49
SVC	98.73	92.12	98.70
CalibratedClassifierCV	98.38	91.36	98.36
Perceptron	98.38	90.06	98.34
LinearDiscriminantAnalysis	97.80	89.57	97.82
KNeighborsClassifier	97.91	89.27	97.92
GaussianNB	95.48	88.11	95.75
AdaBoostClassifier	97.22	85.53	97.27
RidgeClassifier	97.68	85.30	97.59
RidgeClassifierCV	97.68	85.30	97.59
PassiveAggressiveClassifier	97.45	84.27	97.37
BernoulliNB	94.79	84.06	95.01
NearestCentroid	92.82	78.90	92.63
DummyClassifier	85.40	33.33	78.67

Table 4. Classification performance analysis for Room 2 using 60 minutes data

Method	Accuracy (%)	Balanced Accuracy (%)	F1 Score (%)
GaussianNB	93.51	89.11	94.47
QuadraticDiscriminantAnalysis	92.70	88.69	93.89
LinearSVC	97.43	87.45	97.45
BernoulliNB	97.43	87.45	97.45
CalibratedClassifierCV	97.43	87.45	97.45
SVC	97.43	87.45	97.45
SGDClassifier	97.43	87.45	97.45
RidgeClassifierCV	97.43	87.45	97.45
RidgeClassifier	97.43	87.45	97.45
LinearDiscriminantAnalysis	97.43	87.45	97.45
LogisticRegression	97.43	87.45	97.45
NearestCentroid	97.16	87.30	97.21
KNeighborsClassifier	97.16	81.08	97.03
Perceptron	96.89	78.45	96.69
BaggingClassifier	96.89	78.45	96.69
AdaBoostClassifier	96.76	78.38	96.58
LabelSpreading	96.76	78.38	96.58
LabelPropagation	96.76	78.38	96.58
PassiveAggressiveClassifier	96.22	78.09	96.12
RandomForestClassifier	96.76	77.13	96.52
ExtraTreesClassifier	96.76	77.13	96.52
LGBMClassifier	96.62	77.06	96.41
XGBClassifier	96.35	73.19	95.99
ExtraTreeClassifier	96.22	71.87	95.81
DecisionTreeClassifier	96.22	71.87	95.81
DummyClassifier	94.86	50.00	92.36

Based on the experimental results, it is evident that instantaneous data is adequate for Room 2, which has 2 classes, while for Room 1, with 3 classes, the performance of the models improves with 60 minutes of data. Therefore, it was determined that in future studies on HAR, time sequences would be preferred over other methods, despite the increase in computational cost, as the number of activity classes increases. Furthermore, upon closer analysis of the tests, it was observed that the impact of motion sensors on the model was more pronounced in comparison to other ambient sensors. Henceforth, forthcoming research will focus optimization of feature weights on the model for the hardware implementation. In subsequent research, the DNN techniques employed in the tests conducted on the ambient sensor data set will be further examined after augmenting the our dataset. Specifically, these approaches will be evaluated for their efficacy in human activity recognition (HAR) and room authentication. Additionally, ensemble learning will be employed in conjunction with selected DNN techniques. Finally, the findings will be applied to the hardware system.

4 Challenges and future strategies for software optimisation and for authentication hardware development

In the studies carried out within the scope of the project, behavioural biometrics of the rooms within the faculty were identified by using data from motion, temperature, humidity and door sensors. The test results obtained by classifying and interpreting the data obtained from sensors installed in different areas in a room with classical machine learning algorithms are promising. Various plans were made to develop the behavioural biometric system created as a continuation of the project. Detailed information about these plans is given in the following items.

- In the first phase of future studies, it is aimed to increase both the number of sensors installed in the rooms within the faculty and the diversity of sensors. When the studies in the literature are examined, it is observed that the performance of sensor-based biometric systems is improved by using more sensors.
- The data obtained from the sensors were used both instantaneously and sequence-based within the scope of this study. It is known that in learning-based classifiers and sequence-based systems, there is a positive correlation between increasing data and the performance of the system. For this reason, another future plan of the researchers is to obtain longer-term records from the data obtained from the sensors in the future and to use the long-term records obtained to increase the system performance.
- In order to increase the generalization feature of the system and enable it to recognize more activities, it is aimed to collect data from rooms with different scenarios and to detect different activities. In this way, the behavioural biometric authorization system created will be ready for use for wider applications in future versions. In the studies planned to be carried out in this context, the primary goals are the integration of different types of sensors in different laboratories and usage areas, recording from these sensors, and increasing the generalization feature of the system with the long-term recordings obtained.
- One of the processes mentioned in future studies is taking longer-term recordings from different sensors. In this case, with the increase in the data in the records obtained, there may be a possibility that the classical machine learning algorithms currently used in the behavioural biometric system will be insufficient. In case such a situation is encountered, it is among the possibilities that the models used in the software can be replaced with a deep learning model that has been widely used recently.
- One of the works that the YTU team will carry out within the scope of the project is the integration of the created software into the hardware. In this context, the created software must be optimized before being integrated into the hardware. This is because the processing power of the computers on which the models are run and trained and the development boards are not the same. For this reason, both reducing the size of the resulting models by applying processes such as pruning and integrating the optimized models into the development cards form the basis of the second work package.

[1] Cook,Diane, Crandall,Aaron, and Thomas,Brian. (2019). Human Activity Recognition from Continuous Ambient Sensor Data. UCI Machine Learning Repository. https://doi.org/10.24432/C5D60P.