Gaining Insight into User Behaviour and Systematically Determing User Location via Bluetooth Low Energy Beacon Optimisation

Stefan Michael Anthony Benedict, Juan Carlos Augusto, Omer Faruk Kacar

Research Group on Development of Intelligent Environments,
Department of Computer Science
Middlesex University
London, United Kingdom
SB2528@live.mdx.ac.uk, J.Augusto@mdx.ac.uk, OK243@live.mdx.ac.uk

***Abstract*—The multiuser challenge within the field of Intelligent Environments, specifically concerning Indoor Positioning systems needs to be addressed. Solving this challenge is paramount for enabling customised services in indoor locations. This investigation aims to distinguish between multiple users in an Intelligent Environment and identify their specific locations at a given time by employing a visual interface to deploy localised and personalised services to specific individuals in real-time. The investigation is conducted in the Smart Spaces Lab of Middlesex University London (i.e., a fully functional Intelligent Environment). The investigation leverages the Lab's existing technology and uses BLE Beacons with a novel placement approach to complete the User Location challenge. User Data was also generated in the process, giving rise to many insights. On the other hand, Machine Learning was utilised to predict User Activity using the generated Data. The study also offers insight into the latest research concerning indoor positioning systems and their approaches. Additionally, the investigation benchmarks its approaches against the methods published in recent literature and reviews the limitations of this investigation, emphasising future work. Video-based evidence is provided to establish the investigation's authenticity and complement the description in this paper.**

Keywords—Intelligent Environments, Machine Learning, Activity Recognition, Data Generation

I. INTRODUCTION

"Intelligent Environment(s)" (IE) is an industry associated with a broad spectrum of domains ranging from Ambient Assisted Living, Smart Homes (SHs), and so forth [1], [2]. An IE can be defined as a physical space enhanced with computing, communication, and digital material [3]. A closely linked concept to IE is the "Internet of Things" (IoT) due to its association with smart devices (i.e., Smartphones, Smart Televisions, etc.) as well as the communication (i.e., Sending/ Receiving) of data and information within a network [4], [5].

Smart homes (i.e., a form of IE), equipped with smart technology, allow customers to experience customised services. Smart technologies enable individuals/users to be monitored and assisted, improving their quality of life and driving independence [6]. The technology is frequently utilised in the regulation of applications which include but are not limited to lighting, water, air conditioning, and home security [7].

However, enabling user-specific customised services within a multiuser IE depends on the identification of each user and their specific location within the IE [1]. This is a significant challenge within the IE industry, and solving it is paramount due to the dependency on providing smart customised services [1]. An ancillary benefit of solving the challenge is the copious quantity of generated data, giving insight into the user's behaviour (i.e., eating habits, sleeping patterns, activity patterns, etc.). This benefits additional stakeholders, which include but are not limited to energy providers, insurance providers, and ambient assisted living service providers [2], [8], [9]. For instance, this data can provide valuable insights into the health (i.e., progression of diseases) of individuals with special needs, the energy utilisation patterns of the residents, enabling the optimisation of consumption [2], [8].

In essence, the IE domain is a multifaceted industry with a significant global impact, currently on an upward growth trajectory, setting the stage for this investigation. The study utlises Bluetooth Low Energy (BLE) technology building on the approach described in [1], evaluating its performance more systematically with a strategic placement of beacons and a temporal factor while generating user data in the process, enabling User Insight generation and Activity Recognition (AR) via Machine Learning (ML). This approach is advantageous due to its low energy, low cost, small form factor, and flexibility in placement optimisation (i.e., ease of altering beacon locations).as opposed to other alternatives. The Investigation was carried out at the Smart Spaces Lab of Middlesex University London. A variety of fascinating yet challenging issues arise in this multi-user IE study which will explored in the next sections.

The remainder of the paper is segmented into seven sections. Section II focusses on the previous research work that is related to Positioning Systems specifically focusing on Indoor environments followed by AR with an emphasis on ML. Section III to VI presents the primary contribution of this investigation, which includes the data collection in an indoor localisation system allowing for multi-user capabilities in a smart home, obtaining insights from the generated data and respective contribution evaluations assessed in the Smart

Spaces Lab. Finally, sections VII and VIII discusses the limitations and conclusions of the investigation reported in this paper.

II. Related work

Positioning Systems

Positioning systems (PS) are pivotal in Smart Homes (SHs), and Assisted Living (i.e., forms of IE) for motives including but not limited to location awareness, personalised services, and data-driven services [1], [2], [10]. Holistically, PS can be categorised as Indoor Positioning Systems (IPS) and Outdoor Positioning Systems (OPS). OPS has been well established as opposed to IPS via technologies such as Global Positioning System (GPS) and Cellular Networks (CN).

Shifting the focus to IPS, Wi-Fi technology has gained traction in IPS owing to the extensive usage of wireless local area networks (WLANs) in indoor spaces/locations and the prevalence of mobile devices (i.e., Smartphones, smart tablets, etc.) that are Wi-Fi-compatible, enabling a relatively low-cost means of user monitoring in the indoor environment [11]. The Wi-Fi fingerprinting localisation approach is widely used for location estimation as it does not rely on historical data concerning wireless access point (AP) distribution and it does not involve estimating a receiver's angle [11]. The approach predicts a device's location via received signal strength indicator (RSSI) readings of available APs and then associates it with a specific user. This technology triumphs over GPS when it is ineffective in scenarios such as IPS [11]. The methodology in the approach is similar to the CN's fingerprinting technique revolving around the The Received signal strength (RSS) metric of the RSSI approach.

The measure is commonly represented in decibel-milliwatt (dBm) or milliwatt (mW) units, which indicate the actual intensity of the signal received at the receiver (Rx) [12]. The RSS method can determine the distance between a transmitter (Tx) and an Rx [12]. Generally, a higher RSS value indicates a closer distance between the Tx and Rx [12]. When the transmission power (RSSI) at a reference node (RN) is known, the absolute distance can be estimated using various signal propagation models [12]. However, RSS readings can be affected by interferences present within an indoor location. This includes but is not limited to the presence of many objects (i.e., furniture, equipment, etc.), objects with high density, and reflective surfaces (i.e., metal, glass, etc.) [13].

Albeit, Wi-Fi fingerprint technology producing more accuracy, it is associated with higher implementation and maintenance expenses, including a periodic recalibration phase [14].

An alternative to RFID is a cost-effective and versatile technology utilised in IPS [13]. RFID technology can be classified into two primary categories: Active and Passive systems [12]. An Active RFID system necessitates a source of power, can operate a few 100 meters away (i.e., relative to the RFID reader), and functions in Ultra-High frequency (UHF) and microwave frequencies, rendering it suitable for IPS [12]. However, Active RFID systems are incapable of achieving accuracies finer than a meter and are not widely available on most portable consumer devices [12]. Conversely, a Passive RFID system functions without a source of power and functions in certain UHF and microwave frequencies but cannot be directly used for IPS purposes [12]. Interestingly, RSSI and RFID technology have been utilised synchronously in various IPS investigations [13], [15]. The studies were further optimised via means including grouping RFID tags in a "constellation" and Machine Learning (ML) in the approaches of Hatem et al. and Gomes et al., respectively [13], [15].

Further, BLE, a form of radio technology is also well suited for IPS, given its low energy consumption, small form factor, and low cost [1], [16]. In Consideration of the benefits of BLE, it is the predominant technology utilised in solving the multi-user challenge of this investigation. The small form factor of this technology is depicted in Figure 1 in terms of a BLE beacon powered by a CR2032 battery; interestingly, this is the specific type of beacon used in this investigation. The Figure also displays the BLE based android application from the approach of [1] used within this investigation.

BLE is a viable option similar to RFID technology when mitigating privacy concerns associated with video processing IPS solutions [1]. However, BLE is associated with more effort during installation and deployment [1]. While BLE may be used synchronously with a variety of localization approaches, including Received Signal Strength (RSS), Angle of Arrival (AoA), and Time of Flight (ToF), the preponderance of BLE-based localisation systems now in use are RSS-based due to their simplicity [12]. Consequently, the RSS approach is used in this study correspondingly.

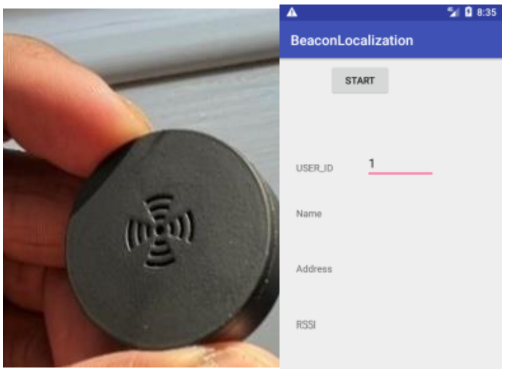

Figure 1: BLE Beacon and application Interface

Solving the challenge of Indoor user localisation inherently generates data, creating an opportunity for AR; the consequent section delves deep into this area with a specific focus on the application of ML within the space.

Activity Recognition

Recognising the activity of residents (i.e., Sleeping, Eating, etc.) at a given time is pivotal for enabling smart services (i.e., Lighting conditions, Temperature conditions, etc.) within an IE [17]. The need for AR is amplified in individuals with special needs (i.e., Individuals diagnosed with Dementia, etc.) [17].

According to the views of Bouchabou et al., there are two predominant methods for AR, specifically Vision-based systems and Sensor-based systems [17]. Sensor-based systems can be further categorised as Wearable Sensors, Object Sensors, and Ambient Sensors [17]. Vision-based AR uses cameras to monitor human behaviour and environmental changes [17]. Computer vision methods such as marker extraction, structure modeling, motion segmentation, action extraction, and motion tracking are used in this approach [17].

Privacy is the predominant drawback of vision-based systems in AR, as in the case of user localisation systems [17]. Alternatively, a Sensor-based system can present itself in many forms, including motion sensors, accelerometers, temperature sensors, and so on [17]. Due to the less intrusive nature of Sensor-based systems, they are more accepted, and fortunately, the current affordability of sensors has spurred favourable conditions for AR within IE [17]. In essence, the role of these systems is to generate the input, which then needs to be funneled into AR models to predict the activity of the residents [17].

AR within IE is a challenging task due to the complexity and variability of individual activities [17]. This is because every individual is unique, with distinct lifestyles, abilities, and habits leading to a plethora of diverse daily activities [17]. Moreover, an individual's activities are also linked with temporal drift, further exacerbating this challenge [17]. The resident's routines and behaviour may evolve, causing a discrepancy between the training data and the current generated data over time [17]. Consequently, it can be deduced that a lucrative AR model is scalable while being adaptable [17]. In alignment with recent literature, AR can be solved with two core techniques, specifically Knowledge-Driven Approaches (KDA) and Data-Driven Approaches (DDA) [17]. DDA is based on user-generated data and ML approaches, while KDA requires expert domain knowledge and logical reasoning to produce rule-based models [17].

This investigation at its core relies on DDA in unison with robust ML models explicitly used for AR, including the Naïve Bayes classifier (NB), Decision Trees (DT), Random Forest (RF), and Logistic Regression (LR) [17], [18]. DT is a non-parametric technique for complex pattern recognition, especially in jobs requiring the classification of many patterns and features [19]. The DT model's primary goal is the sequential assessment of a decision function in a way that lessens the degree of uncertainty in recognising an unidentified pattern [19]. Similarly, RF is a non-parametric technique utilised for classification and regression tasks [20]. The RF model compromises a group of hyperparameters (i.e., number of trees, randomness, etc.) that must be defined for optimum performance [20]. RF models have been used for two core purposes, specifically for future data prediction and determining the contribution significance of the predictor variables against the response variable [20]. Another popular ML classification approach is LR [21]. It is employed in many fields since it is easy to grasp and the outputs are interpretable, which permits a "what-if analysis" [21]. On the other hand, the NB approach assumes that all predictor variables are independent of one another given the response variable and then deduces the probability that an instance belongs to a specific class [22]. Given its practical competency in the real world, this investigation has incorporated the NB classification approach [22].

The subsequent sections initiate this investigation's main contribution, beginning with the Data Collection, leading to the User Location Prediction, AR, and Finally the User Insights segment.

III. DATA COLLECTION

This investigation is built upon by leveraging the well-established Software (i.e., Web Servers, Databases, an Android application, etc.) and Hardware architecture (i.e., Smart sensors, smart switches, BLE Beacons, etc.) of the Smart Spaces Lab of Middlesex University London.

As displayed in Figure 2. All locations, except for Room 04, Room 05, and the Shower, are used for Data Collection. The data that can be generated in the Smart Spaces Lab utilising its technology, as displayed in Figure 2. All locations, except for Room 04, Room 05, and the Shower, are used for Data Collection. This is indicated by the "N/A" stamps on the Figure.

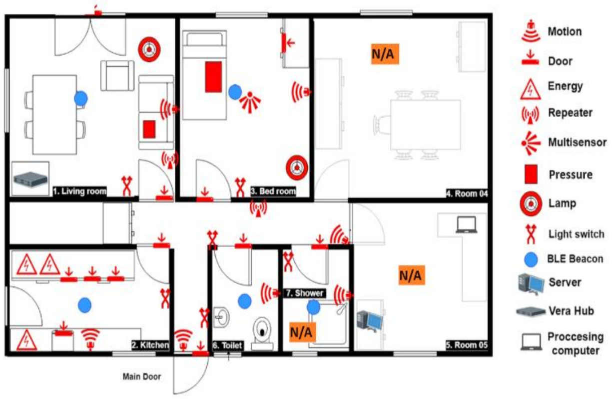

Figure 2: Smart Spaces Lab Map

Further, Table 01 illustrates a sample of the type of Data used within this research by utilising the sensory equipment in the Lab and their respective descriptions. The complete version of the table can be found in [23].

Table 1: Data Sample

Feature Name	Description	Unit/ Type
Specific User	The Specific User within the IE	N/A
Time Stamp	Time of the activity	s/min
Type of Day	Weekday/Weekend	N/A
Activity	The activity (i.e., eating, sleeping, etc.) that the user carries out	N/A
Location	Location of the user	N/A
Routine	Morning/Evening/Night	N/A
Bedroom Data		
Bed Pressure Sensor	Senses if the user is on the bed	Binary
Bedroom Motion Sensor	Senses the movement of the user	Binary
Bedroom Door Sensor	Senses if the door has been opened	Binary
Bedroom Light Switch	Senses if the lamp is turned on	Binary
Energy Sensor 1	Senses if an electrical appliance is turned on	Binary
Bedroom BLE Beacon	Used to determine the location of the user	mW
Living room Data		
Couch Pressure Sensor	Senses if the user is on the couch	Binary
Energy Sensor 2	Senses if an electrical appliance is turned on	Binary
Living room Motion Sensor	Senses the movement of the user	Binary
Living room Door Sensor	Senses if the door has been opened	Binary
Garden Door Sensor	Senses if the door has been opened	Binary

Living room Light Switch	Senses if the lamp is turned on	Binary
Living room BLE Beacon(s)	Used to determine the location of the user	mW

The first phase of Data Collection revolved around formulating Hypothetical scenarios within an IE, specifically a SH. The generated scenarios account for a multitude of elements, including two users, type of day (i.e., Weekday or Weekend), a variety of activities, a selection of Routines, and so forth. This specific data can be accessed in [23]. Next, the Hypothetical scenarios were acted out in person in the Smart Spaces Lab, and the generated data, which was logged on to the relevant databases, was downloaded via a cross-platform database software and then compiled using a secondary software. The final feature space consisted of 33 features and 2678 instances. Due to the number of observations being more significant than ten times that of the features, the data is considered low dimensional, which is beneficial for ML. This is observed and discussed in the upcoming segments of this investigation. The generated data can be obtained from [23]. The data is further explored via visualisations in the User Insights segment of the study. The next section discusses the User Location prediction element of the investigation, beginning with its implementation and concluding with its evaluation. Both The User Location prediction and ML AR were implemented successfully.

IV. USER LOCATION PREDICTION

A.Implementation

Extant BLE Beacon Technology of the Smart Spaces Lab was leveraged while harnessing the approach of M. Quinde et al. coupled with data-driven distinctive contributions (i.e., Beacon placement optimisation, User Location visualisation) within this study to achieve User Localisation.

The approach of M. Quinde et al. used two beacons with a distance of 1m between each other (at the center of each location) in all the experimentation rooms within the Smart Spaces Lab [1]. The IPS's core component is an Android application that analyses the BLE beacons deployed in the lab, filters out the beacon with the strongest signal strength, and associates it with a specific room (i.e., the RSS approach) [1]. The primary assumption of this approach is that the residents of an IE have access to a smart device (i.e., smartphone, smart tablet, smartwatch, etc.) and always carry it in person while engaging in activities [1]. This enables each resident to be assigned a unique identifier within the application [1]. The implemented approach emphasises the importance of the strategic placements of the beacons within the experimentation locations [1]. Optimising the Beacon location was implemented via quantitative means of Analysis and Testing. The experimentation was based on three distinct approaches: Single Beacon, Dual Beacon, and Dual Beacon (Optimised) location layouts. Both the Single Beacon and Dual Beacon approach position the beacons on the ceiling of the respective rooms. In contrast, the Dual Beacon (Optimised) approach strategically positions the beacons in regions linked with a higher probability of more significant activity (i.e., areas more likely for the residents to spend more time).

Additionally, the experimentation accounted for distinct acceptable time frames (i.e.,5 seconds and 15 seconds) of user location recognition when transitioning from one location to the next, with three iterations of each transition. Intriguingly, the "Location transitioning" element of the experimentation is considered as future work of the approach of M. Quinde et al. [1]. After a thorough evaluation, the Dual Beacon (Optimised) approach triumphed over the alternatives. Further details on the approaches can be found in [23].

The data (i.e., The MAC address of the closest beacon) was retrieved from the Beacon localisation database of the Smart Spaces Lab through SQL querying and stored as a variable within a script leveraging the Python programming language The Beacon MAC Address was associated with distinct locations (i.e., Bedroom, Living Room, Toilet, and Kitchen) within the Lab. Next, the User's location was visualised via a maroon dot which would move to specific hard-coded locations/ coordinates within the map of the Lab depending on the Beacon MAC Address. The map of the Lab spanned the entire area of the interface, and the User Location dot was overlayed upon the map, while the Middlesex University Logo was used as the icon of the interface. Furthermore, the dynamic Interface for the User Location of this study can be observed in [23] and is evaluated in the subsequent section.

B.Evaluation

The summarised performance metrics of the three distinct placement approaches of the investigation experimented with two different time frames (i.e., 5 seconds and 15 seconds) can be found in [23]. The techniques used explicitly used in this study are the 1 Beacon, 2 Beacons, and 2 Beacons (optimised) approaches.

An approach's efficacy is considered superior if it is associated with higher values of the performance metrics used in the study. Considering the Accuracy metric, the 2 Beacons (optimised) approach outperforms the alternatives. Intriguingly, the acceptable time frame attribute had an impact on the Accuracy across all approaches and metrics except Precision. Regarding the 2 Beacons (optimised) approach, there is a spike in accuracy from 58.33% to 80.56%. There are better factors than the Precision metric in accessing the best approach in this study. This is because the experimentation is based on whether the system recognises a user's transition from one location to another. Regarding the Recall and F1 score metrics, the 2 Beacon (optimised) approach again overpowers the alternatives with values of 0.806 and 0.893, respectively. The values correspond to the 15-second time frame. Table 2 simultaneously illustrates the performance metrics of recent literature in the field of IPS and the 2 Beacon (Optimised) approach. The approaches include BLE, RFID, and Wi-Fi-based technologies/techniques.

Table 2:Beacon Performance

IPS Approach	Accuracy	Precision	Recall	F1-score	Recognition proximity
BLE RSSI [1]	91.01%				Room-Level
RFID Clustering and Hierarchical Classification [13]	99.36%	99%	99%	99%	5 cm

RFID Constellation [15]					81 cm
Wi-Fi Fingerprint [24]	86%				Room-Level
BLE 2 Beacon (Optimised)	80.56%	100%	81%	89%	Room-Level

A comprehensive review of the literature showed that it posed a challenge in the performance comparison of the distinct approaches. This challenge stems from various factors, including but not limited to missing performance metrics, different locations of IPS implementation, and varied methods of estimating the performance metrics. Accuracy is the primary indicator of performance used for analysis within this section due to the unavailability of other metrics. All the models perform with an Accuracy of over 80%, with the RFID Clustering and Hierarchical Classification approach associated with the highest accuracy at 99.36% and the BLE 2 Beacon (Optimised) approach with the lowest at 80.56%.
Although the BLE RSSI approach and the BLE 2 Beacon (Optimised) approach have been implemented in the Smart Spaces Lab at Middlesex University London, the methods used to acquire the respective performance metrics are different. The BLE 2 Beacon (Optimised) approach accounts for the transition between locations and considers acceptable time frames for user recognition within an IE, which is not implemented with the BLE RSSI approach. Further discrepancies between the different techniques include the IPS's recognition proximity. For instance, the RFID Clustering and Hierarchical Classification approach had a location precision of 5 cm and 81 cm for the RFID Constellation approach, while the other techniques can only achieve room-level accuracy. The next segment focuses on the evaluation and results of the AR element of this study. Interestingly, the RF model outperformed the alternatives with performance metrics over 99%.

V. ACTIVITY RECOGNITION WITH ML

A. *Implementation*

The ML component of this investigation is based on six models, specifically DT, RF, LR, NB, a Dummy classifier (DC) model, and a tuned RF model. The classes predicted from the models included Eating, Resting, Sleeping, and Miscellaneous activities. After the models were executed, their performances were evaluated to distinguish the best model. The best model in this investigation was quantified as the RF model and was tuned.

The computational hardware utilised in this segment of the investigation consisted of an 8th generation Intel Core i5 processor, 16 GB RAM and an 8 GB GPU. The pipeline was implemented using the Python programming language. The data is then split into the independent and Target variables (i.e., User Activity) as this is a classification-based ML challenge. Due to the reason that ML models work purely on numerical input data, the data had to be preprocessed before the implementation of the ML pipeline. The preprocessing was executed by converting the Boolean variables (i.e., Lights, motion sensors, etc.) of the data into numeric format (i.e., from True or False to 1 or 0) and using the approach of Label encoding for the remaining independent categorical variables, which include the "Weekday/Weekend", "Location" and "Routine" attributes. It was observed that there are 2,678 instances corresponding to 32 independent variables. Next, the data was split into Train and Test data with a training size of 30% corresponding to 1874 instances. The next and final phase of this was to implement the ML pipeline of the DT, RF, LR, and NB models. It is imperative to emphasize that the data was further preprocessed before executing the DT through scaling the data for better performance. This is not required for the remaining models. Moreover, a random state parameter was included in the relevant models, including the LR, DT, and RF models, ensuring the models' reproducibility. The subsequent segment discusses the most accurate model's dominant features. To determine the most influential features of the ML element of the investigation, the model with the highest accuracy needed to be assessed. This task was implemented by estimating the accuracy of all the models (i.e., DT, RF, LR, NB, and DC), Table 3 summarises the accuracy of the models.

Table 3: Machine Learning Model Performance

Model	**Accuracy of Train data**	**Accuracy of Test data**
DT	1.000	0.998
RF	0.999	0.999
LR	0.957	0.964
NB	0.735	0.720
DC		0.241

As illustrated in Table 3, RF can be deemed the model with the highest accuracy on test data and the best Bias-Variance trade-off. The DT, RF, LR, and NB models have an excellent Bias-Variance trade-off (i.e., The ability to be less susceptible to noise and identify the underlying patterns) and perform significantly better than the baseline/DC model. The performances of these models are further discussed in detail in the next segment.
On the other hand, Figure 3 displays the most important features of the RF model in descending order. The top three attributes associated with the high influence on the model are the Location, Time, and Bedroom Bed Pressure. The contributions are over 25%, just under 20%, and just over 10%, respectively. Strikingly, only five features are associated with a contribution of over 5%.

Albeit the RF model exhibited superior performance on both training and testing data, it was tuned for further optimisation. The model was tuned on the maximum features hyper-parameter and used the Out-of-bag (OBB) score as an indication of performance. The maximum features hyperparameter with the best performance was established to be 6 with 0.998 accuracy. In the succeeding phase leverages the tuned RF model and the DC model to visualise the ML element of the study via an interface that demonstrates the capability of the tuned model over the baseline using the Python programming language. Further, the total training time was 1.11 seconds for all the models with the Tuned RF Model contributing 0.49 seconds to the metric. This was estimated by running a timer during the training phase using the time module of the Python programming language. The total memory storage of the ML pipeline was estimated at just over 630 kilo Bytes by accounting for memory usage of the training data utilised, and ML models, which were implemented using the sys module of the Python programming language.

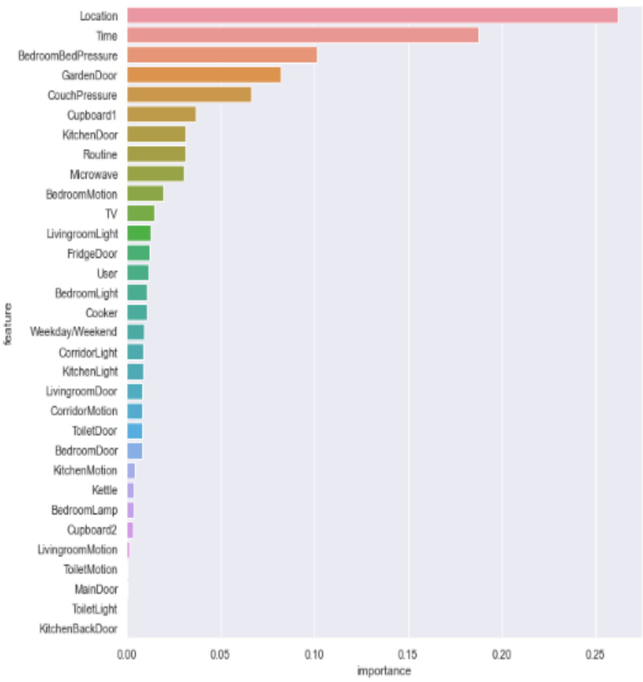

Figure 3: Random Forest Model's Most Influential Features

.

Figure 4: Machine Learning Interface

Figure 4 displays the interface which benchmarks the performance of the tuned RF model against the baseline DC model using a visualisation approach. The Users' activities are represented by distinct icons, as showcased at the bottom of the interface in the Legend section. The backend of the interface uses test data to update the respective front-end icon of the Predicted Activity (i.e., model prediction) and the Real Activity, enabling performance comparison. A video demonstrating the interface can be accessed through [23]. In a real-world scenario, this can be the real activities of the user where it is beneficial to be monitored for dementia in a care home but would be dependent on the captured data. It should be noted that the implemented system does not account for anomalies within test data and would lead to a misclassification of the resident’s activity. Moreover, the System was tested with humans in the real physical Smart Spaces Lab. However, those participating in the validation do not live full time in the Smart Home due to legal constraints imposed by the University Campus. Consequently, the ML-based User Activity prediction approach used in this investigation is extensively evaluated against alternative approaches in the next segment.

B. *Evaluation against published literature*

As established, the study explores a pipeline of six ML models for User Activity Prediction/ AR. An approach's efficacy is considered superior if it is associated with higher values of the performance metrics used in the study, identical to the User Location element. All the approaches (i.e., DT, RF, LR, NB, and Tuned RF) perform better than the baseline/DC model. Further, the DT, RF, LR, and Tuned RF have all the performance metrics (i.e., Accuracy, Precision, Recall, and F1-Score) above 95%. On the other hand, the NB model is only associated with values of just over 70%. However, it still performs better than the DC model, where all its performance metrics are under 30%. Notably, the RF models are quantified to be the best-performing techniques. Interestingly, the tuning phase of the RF model shows no improvement in performance. Due to the Superior comparative performance of the RF model, it will be used to benchmark the investigation’s performance against recent Literature concerning User Activity Prediction in IEs.

Figure 5 simultaneously illustrates the performance metrics of recent literature in AR and the tuned RF approach. The approaches include K-Nearest Neighbor (KNN), Support Vector Machine (SVM), RF, Gradient Boosting Decision Tree (XGBoost), NGBoost, DT with entropy, DT, LR, NB, and Stochastic Gradient Descent (SGD) ML techniques. The research approaches used correspond to a study by Dakota State University, as observed in [18]. This study considers three distinct data sets and three distinct methods of analysis pertaining to the classification of classes of the data sets for AR [18]. This classification includes Individual activities, which use the activities of the data sets as is, Group activities which group the activities into common categories (i.e., standing and sitting categorised as stationary activities, etc.), and Common activities, which only consider the mutual activities [18]. However, this segment only considers the dataset that exhibited the best performance, specifically the “Pamap2” dataset from the University of California and the “Individual activities” approach, as it resembles the methodology of this investigation [18]. Figure 5 indicates that most approaches based on [18] generate identical performance of over 99% across all assessed metrics (i.e., Accuracy, Precision, Recall, and F1-Score) when benchmarked against the RF approach in this investigation. The exception is the NGBoost, LR, NB, and SGD approaches, which generate between 90% and 94% on the performance metrics.

Model	Accuracy	Precision	Recall	F1-Score
KNN	0.999	1	1	1
SVM	0.999	1	1	1
RF	0.999	1	1	1
XGBoost	0.999	1	1	1
NGBoost	0.936	0.93	0.93	0.93
DT	0.999	1	1	1
DT with entropy	0.999	1	1	1
LR	0.92	0.92	0.92	0.92
NB	0.901	0.91	0.90	0.90
SGD	0.9	0.9	0.9	0.9
RF	0.997	0.996	0.997	0.997

Figure 5: ML Performance for AR (State of the Art)

Hence, in theory, ML has demonstrated theoretical feasibility within the field of AR. Interestingly, Section VI delves into the User Insights segment of investigation. This section presents the insight potential of the Data Collected in section III. The investigation is finally concluded in section VII.

VI. USER INSIGHTS

This section is dedicated to establishing the potential for obtaining insights from the generated data and presenting it in a Visual format. The Python programming language was used for Data Visualisation, Data Manipulation and Analysis. This section considers insights relating to the user's overall weekly Activity breakdown, Location breakdown, Motion breakdown and Energy Consumption. The insights correspond from Figures 6 to 9 respectively. To compute the overall weekly metrics the data had to be processed and quantified from weekday and weekend data. This means factoring and summing up the weekday data five times and the weekend data twice. Further, it is crucial to note that each instance within the data was assumed to have occurred in 1 minute when it was 1 second. This was done to make the investigation more identical and closer to real life. Some insights derived from Figure 6 include but not limited to The Weekday contribution to the Weekly engagement is higher for all activities except Leisure for both users, highlights the overall time spent by each User in each location over a period of a Week. It provides insights, which include User 1 spending more time in the Bedroom, Garden, and living room, while User 2 spends more time the Corridor, Kitchen, and Toilet. as displayed in Figure 7. The time spent in the Kitchen and the Garden is approximately identical for both users. Figure 8 on the other hand displays User 1 is more active (i.e., moves more) than User 2, the data is obtained from the motion sensor data from the Smart Spaces Lab. Energy is defined as the overall amount of work created or generated, whereas power is the rate of energy transmission over time. Utilising the definitions and the power consumption values of the respective electrical appliances of the IE, the energy consumption for each instance corresponding to each specific electrical appliance was calculated. Figure 9 aims to determine the type of day (i.e., Weekday or Weekend) with the highest Energy consumption and associate it with a specific User. It can be observed that User 1 has a higher energy consumption on Weekdays and Weekends with a value of around 1200 Watt-hours and just over 1000 Watt-hours, respectively. However, User 2 has a higher energy consumption on the Weekends than the Weekdays, with a value of just over 1000 Watt-hours and around 900 Watt-hours. Further, it is also apparent that User 1's Energy consumption is higher than that of User 2, given that both of its curves are higher than that of User 2, This was considering the Lights, TV, Cooker, Kettle and the Microwave to have a power of 10, 100, 1500, 2000 and 1200 Watts respectively, and considering the Energy in terms of Watts per minute.

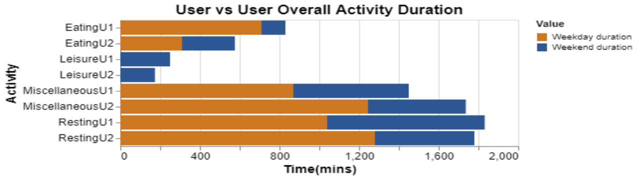

Figure 6: User Overall Activity Duration

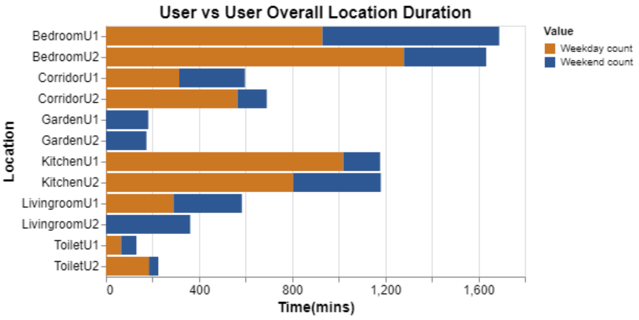

Figure 7: User Overall Location Duration

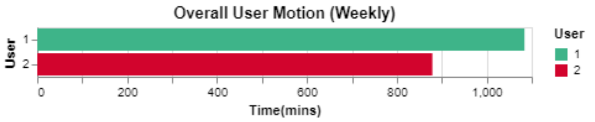

Figure 8: Overall User Motion

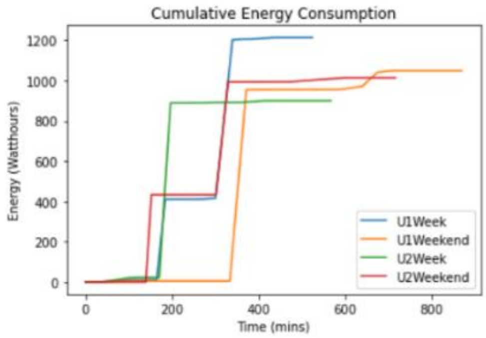

Figure 9: Cumulative Energy Consumption

VII. CONCLUSIONS

To provide customised indoor services, it is imperative to address the multiuser challenge within Intelligent Environments, especially regarding indoor positioning systems. This research uses BLE beacons and a visual interface to track the real-time whereabouts of several users at Middlesex University London's Smart Spaces Lab. It creates user data, uses machine learning to forecast user activity, investigates indoor positioning research, compares to recent literature, identifies the study's shortcomings, and makes recommendations for future work and is also backed up by video evidence.

Evidently, the BLE Beacon placement plays a crucial role in the performance metrics of the approach within the scope of this investigation. Despite the strategic placement efforts such as region experimentation and increasing the number of beacons, the approach is limited in performance compared to the alternatives, as discussed in detail in the specific Evaluation segment. Hence, alternative efforts need to be taken to assess the feasibility of performance improvement, which is regarded as further work of this investigation. Unexplored strategies include incorporating ML, assessing the performance variation with other form factors of BLE-compatible smart devices (i.e., wearables, etc.), and Securing the Beacons in such a way that is fixed in place/ immobile (i.e., static BLE enclosure). This preserves the performance consistency of the system. Since the current system uses a Velcro (i.e., Hook and loop) mechanism to attach the beacons each time a beacon is reattached (i.e., change of battery, etc.), the system's homogeneity is altered slightly. Monitoring of the alternative technologies within the Smart Spaces Lab, User Identification in corridor regions, and the implementation of the visual interface as a standalone application that will be beneficial in Remote Patient Monitoring use cases (i.e., disease progression, etc.) are also considered as further work. On the other hand, The AR segment performed on par with recent literature and the preceding section of the investigation showcased the capability of Insight generation both in terms of quality and variety. This elucidates that both sections depend on the quality of data collected. This data's primary implication and limitation is that it is Fictitious, albeit it has been generated through the reenactment of hypothetical scenarios in the Smart Spaces Lab. Collecting more data with real users over a long period (i.e., Months, Years, etc.) and presenting this on live dashboards where there is a summary of data over a configurable period (i.e., weeks, months, etc.) can be deemed as future work enabling realistic insights in real-time. However, this work is associated with the data collection implementation challenges, which are the main bottlenecks (i.e., Privacy concerns, etc.).

REFERENCES

[1] M. Quinde, J. Giminez-Manuel, C.L. Leonard and J.C. Augusto, "Achieving multi-user capabilities through an indoor positioning system based on BLE beacons," in - 2020 16th International Conference on Intelligent Environments (IE), 2020, DOI: 10.1109/IE49459.2020.9155011.

[2] J.G. Gimenez Manuel, J.C. Augusto and J. Stewart, "AnAbEL: towards empowering people living with dementia in ambient assisted living," Universal Access in the Information Society, vol. 21, (2), pp. 457-476, 2022.

[3] W. Minker and T. Heinroth, "Chapter 18 - spoken dialogue systems for intelligent environments," in Human-Centric Interfaces for Ambient Intelligence, H. Aghajan, R. L. Delgado and J. C. Augusto, Eds. 2010., DOI: 10.1016/B978-0-12-374708-2.00018-8.

[4] D. K. Alferidah and N. Jhanjhi, "Cybersecurity impact over bigdata and IoT growth," in - 2020 International Conference on Computational Intelligence (ICCI), 2020, DOI: 10.1109/ICCI51257.2020.9247722.

[5] S. Voulgaris et al, "Blockchain technology for intelligent environments," Future Internet, vol. 11, (10), pp. 213, 2019.

[6] D. Marikyan, S. Papagiannidis and E. Alamanos, "A systematic review of the smart home literature: A user perspective," Technological Forecasting and Social Change, vol. 138, pp. 139-154, 2019,. DOI: 10.1016/j.techfore.2018.08.015.

[7] A. Rego, P. L. Ramírez, J. M. Jimenez, and J. Lloret, "Artificial Intelligent System for multimedia services in Smart Home Environments," Cluster Computing, vol. 25, no. 3, pp. 2085–2105, 2021. DOI:10.1007/s10586-021-03350-z.

[8] M. Z. Fakhar, E. Yalcin and A. Bilge, "A survey of smart home energy conservation techniques," Expert Syst. Appl., vol. 213, pp. 118974, 2023,. DOI: 10.1016/j.eswa.2022.118974.

[9] M. Eggert, "Understanding the acceptance of smart home-based insurances," 2019.

[10] M. Anjum et al, "RSSI Fingerprinting-Based Localization Using Machine Learning in LoRa Networks," IEEE Internet of Things Magazine, vol. 3, (4), pp. 53-59, 2020.DOI: 10.1109/IOTM.0001.2000019.

[11] O. P. Babalola and V. Balyan, "WiFi fingerprinting indoor localization based on dynamic mode decomposition feature selection with hidden Markov model," Sensors, vol. 21, (20), pp. 6778, 2021.

[12] F. Zafari, A. Gkelias and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," IEEE Communications Surveys & Tutorials, vol. 21, (3), pp. 2568-2599, 2019. DOI: 10.1109/COMST.2019.2911558.

[13] E. L. Gomes et al, "Clustering and Hierarchical Classification for High-Precision RFID Indoor Location Systems," IEEE Sensors Journal, vol. 22, (6), pp. 5141-5149, 2022. DOI: 10.1109/JSEN.2021.3103043.

[14] N. Fet, M. Handte and P. J. Marrón, "An approach for autonomous recalibration of fingerprinting-based indoor localization systems," in - 2016 12th International Conference on Intelligent Environments (IE), 2016,DOI: 10.1109/IE.2016.13.

[15] E. Hatem et al, "Performance, accuracy and generalization capability of RFID tags' constellation for indoor localization," Sensors, vol. 20, (15), pp. 4100, 2020.

[16] S. G. Obreja and A. Vulpe, "Evaluation of an indoor localization solution based on bluetooth low energy beacons," in 2020 13th International Conference on Communications (COMM), 2020.

[17] D. Bouchabou, S. M. Nguyen, C. Lohr, B. LeDuc, and I. Kanellos, "A survey of human activity recognition in smart homes based on IOT sensors algorithms: Taxonomies, challenges, and opportunities with Deep Learning," Sensors, vol. 21, no. 18, p. 6037, 2021. DOI:10.3390/s21186037.

[18] L. S. Ambati and O. El-Gayar, "Human activity recognition: a comparison of machine learning approaches," Journal of the Midwest Association for Information Systems (JMWAIS), vol. 2021, (1), pp. 4, 2021.

[19] J. K. Solano Meza, D. Orjuela Yepes, J. Rodrigo-Ilarri, and E. Cassiraga, "Predictive analysis of Urban Waste Generation for the city of Bogotá, Colombia, through the implementation of decision trees-based machine learning, support vector machines and artificial neural networks," Heliyon, vol. 5, no. 11, 2019. doi:10.1016/j.heliyon.2019.e02810.

[20] P. Probst, M. N. Wright and A. Boulesteix, "Hyperparameters and tuning strategies for random forest," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 9, (3), pp. e1301, 2019.

[21] [1] S. Radovanovic, A. Petrovic, B. Delibasic, and M. Suknovic, "Enforcing fairness in logistic regression algorithm," 2020 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), 2020. DOI :10.1109/inista49547.2020.9194676.

[22] S. Chen, G. I. Webb, L. Liu, and X. Ma, "A novel selective naïve Bayes algorithm," Knowledge-Based Systems, vol. 192, p. 105361, 2020. DOI:10.1016/j.knosys.2019.105361.

[23] S. M. A. Benedict, "Distinguishing between multiple Users in an IntelligentEnvironment and their Activities while gaining insights intotheir Behaviour: A Data-driven Approach". Middlesex University, 29-Sep-2023, doi: 10.22023/mdx.24216723.v1.

[24] Chenshu Wu, Zheng Yang, Y. Liu, and Wei Xi, "Will: Wireless Indoor Localization Without Site Survey," 2012 Proceedings IEEE INFOCOM, 2012. DOI:10.1109/infcom.2012.6195809.