This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3603917

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Reducing Information Overload with Machine Learning in Mobile Pervasive Augmented Reality Systems

RUI PASCOAL[1,2], ANA MARIA DE ALMEIDA[2,3],(Senior Member, IEEE), and RUTE C. SOFIA[4,2], (Senior Member, IEEE)

[1]Instituto Universitário - Atlântica, Oeiras, Portugal
[2]Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, Lisboa, Portugal
[3]CISUC - Centre for Informatics and Systems of the University of Coimbra, Coimbra, Portugal
[4]fortiss - research institute of the Free State of Bavaria for software intensive systems and services, Munich, Germany

Corresponding author: Ana Maria de Almeida (e-mail: ana.almeida@iscte-iul.pt).

This article research task was partially supported by Fundação para a Ciência e a Tecnologia, I.P. (FCT) [ISTAR Projects: UIDB/04466/2020 and UIDP/04466/2020] and research grant 79759.

ABSTRACT Augmented reality systems in dynamic environments still struggle with the challenge of what information should be displayed at which time. This work focuses on the case of Mobile Pervasive Augmented Reality Systems (MPARS) and their use in dynamic environments such as outdoor sports. An open-source proof-of-concept for a machine learning-based architecture to implement an MPARS on a specific use case of outdoor usage in a sports environment is presented. The new design for the system relies on heuristics that combine technology acceptance indicators, sensing, and information volume criteria to show the user a contextually meaningful subset of information. The information to the user is displayed in close-to-real-time, and the system can adjust and customise to prevent information overload. A first set of experiments was carried out based on end-user preferences to show the feasibility of the proposed system. To provide meaningful feedback, i.e., the right information when needed or wanted, to sports users on their MPARS experience, a predictive model was trained and shown to be able to estimate when information should be displayed to the user.

INDEX TERMS Mobile Pervasive Augmented Reality System, Machine Learning, Sensing, Context-awareness, Information Modeler Learning, Adaptable system.

I. INTRODUCTION

AUGMENTED Reality (AR) is a technology that integrates images, information (e.g., situational information), and other types of digital objects. AR products are now being applied in several areas, like sports, games, health, industry, culture, tourism, and education [1] [2] [3]. Over two decades, AR systems evolved from dedicated devices or personal computers to be used on heterogeneous mobile platforms, such as desktops, tablets, smartphones, and notebooks [4]. In fact, Cao *et al.* advocate that mobile augmented reality systems (MARS) have to be adaptable to address several on-demand user interactions with various IoT devices in smart environments and for different application areas [3].

Meeting the requirements of technology adoption for AR, such as customisation, adaptability, and familiarity (i.e., end-user experience), is made possible by the intelligent data generated by appropriate devices such as smart glasses, smartphones, or other AR devices. A key aspect of ensuring that AR becomes ubiquitous is making it human-centric. This requires the underlying technology to add value in terms of *Quality of Experience (QoE)* [5] [6].

While AR holds great promise for enhancing our perceptions and helping us to see, hear, and feel our environment in new and enriching ways, there are still issues to overcome. These include a better acceptance of the technology and the ability to achieve a higher QoE. Some of the current challenges are (i) how to properly calibrate and adjust the data, (ii) how to take user preferences into account when providing information, and (iii) how to measure the amount of information to be displayed [7].

Another relevant point to be made for successful technology adoption and useful use of AR is the question of what information to display and when. This relates to the concept of *information overload*, which here means that a user is "receiving too much information" [8]. These issues become critical for the integration of MPARS in mobile devices and, in particular, for mobile outdoor activities, where the adoption of AR is expected to grow at a higher rate [3] [9].

The motivation behind this work is to provide a better *Quality of Experience (QoE)* to the user through the automated and continuous integration of user preferences into the AR system. To this end, this work focuses on *Mobile Pervasive Augmented Reality Systems (MPARS)* [10], i.e., AR systems that are carried by the user and that are capable of adapting the information to be displayed over time and space, based on situational and contextual awareness.

To achieve such a level of automation, an MPARS should consider user preferences, context awareness, and also situational awareness. This enables an MPARS to provide *meaningful feedback* to end-users in different environments, by displaying *the right information at the right time*, thus avoiding information overload. To this end, a new architecture is proposed and tested using a specific use case of outdoor sports environment usage. Previously, an MPARS was defined as an information manager for specific sports activities, with an automatic activity recognizer in outdoor environments [10] [11], but now the new MPARS Architecture has a complement and help of machine learning, capable of persistent adaptation of information for each specific user within the same sports activity. See Figure 1 with an original MPARS diagram.

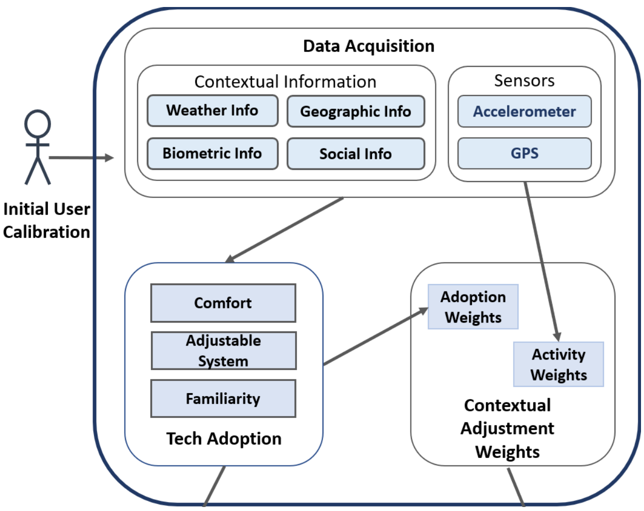

FIGURE 1: Representation of the original MPARS diagram and its subcomponents.

The main contributions of the work are two-fold:

- A functional description of a novel MPARS architecture capable of adapting its feedback to the user context in outdoor environments.
- A context assessment model based on real user preferences, which can be used as a basis to assist future work to define adaptation variables based on realistic user preferences.

To address issues relating to the increase of QoE expectations by reducing information overload in an AR for outdoor activities, this work focuses on the following research questions: (i) How can QoE be improved by automatically adjusting the information displayed/suggested to the user in real time? and (ii) What is the feasibility of using a *Machine Learning (ML)* algorithm to predict and adjust meaningful feedback?

Research into these issues led to an improved AR-adjustable MPARS architecture, which is discussed in this paper and validated through an experimental ML approach involving field tests combined with end-user surveys.

The rest of the paper is organised as follows. Section II addresses the related applicability and needs of MPARS applications. Section III presents the MPARS architecture. Section IV provides examples of scenarios where MPARS can be used. Section V presents the performance evaluation with measurements, machine learning, and an analysis of the results. Finally, Section VI presents the conclusions highlighting directions for future work.

II. RELATED WORK

On AR in sports

We start by looking at how AR is transforming fan engagement and experience in sports, highlighting AR's ubiquity, technological challenges, and applications on personal devices. The tracking and interaction in AR are analyzed, highlighting mobile sensors for activity recognition. However, information overload in AR impacts the user experience. Studies suggest that ML can filter and personalize data, but adaptation to user preferences is still limited. Sawan *et al.* provide a concise and systematic literature review, analysing how mixed reality and AR are providing a growing number of applications in the world of sport. The authors believe that the introduction of this technology in sports can implement and greatly improve fan engagement strategies and experience in the world of e-sports [12].

Kim *et al.* discuss concepts of AR applications and highlight the need for technology efficiency [13]. They describe a variety of new AR applications and the issues that arise in the development of these basic technologies and applications. AR is present today on most personal devices, such as smartphones and tablets, and thus it is ubiquitous [14] [13]. A representative case of ubiquity is the continuous display of images and other types of information superimposed on the real environment [13].

On the interplay of AR and sensing

Billinghurst *et al.* explore some areas of AR, such as tracking and display, development tools, input and interaction, and social acceptance. A requirement of AR technology for virtual image superimposition on the real view is the existence of a tracking system to locate the user's point of view seamlessly blending the real and the virtual images. Mobile devices for AR, like smartphones and smart glasses, present new opportunities for hybrid tracking, as they are today capable of performing integrated sensing based on a variety of sensors such as accelerometers, gyroscopes, GPS, and wireless interfaces, thus resulting in a higher degree of accuracy in the context of activity recognition [15].

In fact, smartphones today integrate several sensors, of which the accelerometer is one of the most popular and is used to detect end-user activities. In a prior work [11], the authors investigated sensors for data acquisition in outdoor contexts, finding that good results could be achieved by using only the accelerometer and GPS for activity recognition in sports environments. The authors also investigated criteria for selecting the information to be provided to the user under specific conditions and activities to prevent information overload. The results show that potential users' interests in feedback vary with the type and level of effort of the activity being performed, and the authors concluded that information overload is directly related to user speed. Concerning functionalities/actions preferred when in an outdoor activity, users also expressed preferences varying with the activity [11].

Bayat *et al.* carried out an experiment with the smartphone's accelerometer sensor to identify human physical activities, such as walking, running, dancing, etc. The system used an ML approach for the detection of the activity being performed and tested several classifiers, each achieving good performance in recognising the activities [16].

About adaptive machine learning systems

ML techniques are important for predicting future events, such as sports. Regarding automated feedback, while most of the work focuses on automatic control loops, some works relate to adapting feedback information to the user. Schmitt *et al.* propose using static decisions based on rules and first-order logic to define situations in terms of the basic context but built with ML techniques [17]. The authors recognise that to provide self-adaptive services, it is necessary to capture contextual information from sensors and use the collected information to reason and classify situations [17].

Liu and Li present a study on an intelligent computer-based sports learning system with predictive control that can provide feedback and adjust in next to real-time based on the athlete's performance to improve training efficiency and results. The experimental results show that this intelligent learning system has great potential for application in sports training and competition, improving athletes' skill levels and performance. The system aims to provide information volume adjustments for efficient feedback to improve sports training and competition [18].

The work of Stacchio *et al.* makes an important contribution to support the need for dynamic interface adjustment in AR systems to assist users during activities. Suitable for outdoor sports, the authors present Magic AuGmentEd Workout, a dynamic AR guidance system for outdoor running that can adapt a workout scenario to a user's performance and manage a sequence of different activities: running, sprinting, bodyweight, and rest. The system follows a workout plan and then adjusts its intensity based on the user's current performance [19].

Another approach that helps support the need and opportunity for dynamic tuning of an MPARS system is the one proposed by Soltani *et al.*. The authors reviewed the current literature and found that using AR provides additional information and feedback on the learning of sporting skills can be used to encourage practice, offering supplemental advantages when compared to other technologies [20].

On Information Overload

While there is a strong focus today on using AR in personal, pervasive systems, related work regarding information overload is scarce. Existing work usually aims to improve image transmission or the decluttering of objects when information is transmitted [21] [3] [22].

Information overload is a critical aspect to handle in the context of QoE improvement and directly relates to perception. Miller describes the so-called *inelastic limit of human capacity* or *cognitive ability* [23]. If the amount of information received exceeds certain thresholds, the human ability to process information quickly degrades [23]. Sawyer *et al.* investigate the use of AR in the context of driving [5]. AR can distract drivers, as the results show that messaging using a Google Glass or a smartphone-based messaging interface impaired driving compared to driving without multitasking [5]. As explained by Bawden et al., information overload is currently a major barrier to successful MPARS adoption. The strategy for providing meaningful information can go through filtering and removing information noise, establishing balanced useful information [24]. In this regard, ML can be a useful tool to adjust the information to be received so that it is relevant and timely [25] [26].

Concluding remarks

In a nutshell, the use of sensing in AR systems concerns the integration or classification of aspects related to human behaviour in different activities. Previous related work proposed using sensing to provide additional information (context) into AR systems to improve QoE and the overall system performance regarding the system's main goals. However, none of the found works considers context awareness, not just based on information that can be sensed to understand the activity, but also integrating user preferences, i.e., employing a user-centric calibration of the system, which may bring into play the issue of information overload. Activity recognition in this setting must consider the personalisation of AR systems to improve QoE. The present work brings complementarity by (i) using user preferences to provide finer customisation, (ii) proposing an ML-based solution to improve the type and volume of information that should be provided to the user based on external conditions, the specific activity being performed, and (iii) encompassing and adjusting to the user's preferences, in order to diminish as much as possible the occurrence of information overload and providing the right information at the right time.

The gap in research lies in the lack of adaptation of AR to individual user preferences. Although there are studies on ML engagement, tracking, and information filtering, there is still little exploration of dynamic personalization of information based on external conditions and user preferences, which impacts the experience and can lead to cognitive overload.

III. MPARS ARCHITECTURE

This section presents a new proposal designed to provide *meaningful feedback*, i.e., useful and timely information to the mobile user in the outdoor activity context. As previously noted, the system is designed to be adjustable via context and user preference awareness. To best illustrate the proposal, a dynamic environment, namely an outdoor sports environment, is considered. In this environment, the information provided in the MPARS display must be adapted to the type of activity in progress and the user's surroundings, also considering the user's preferences.

Users interact with the MPARS by performing specific actions (e.g., taking photos, filming, calling, messaging, or using social networks) and by requesting information elements (e.g., weather, location, biometric counts, and social information). Therefore, the MPARS requires a set of working modules to build an adaptive AR layout, as shown in Figure 2.

In the first step (**Data Acquisition**), the system acquires (via sensing and the direct user input) information related to the user (personal information), contextual data and sensing data from the available sensors.

The second step (**Tech Adoption Variables**) concerns the data processing aspects. Namely, technology adoption metrics derived from authors' previous work [11] are used to calibrate a set of weights or technology adoption variables. Currently, the proposed proof-of-concept considers the following metric categories: convenience (contributing to *expectation*, *experience*), adjustable system to avoid *information overload*, and familiarity contributing to *expectation*, *experience*, *auto adjustment*.

The third step (**Context Adaptation Weighting**) consists of calibrating the weights used to adapt the contextual information to be provided to the user, to establish a relationship between the contextual information and the user's preferences for the information to be displayed, using the acceptance weights derived from the previous step. Data relating to each type of information (weather, location, social information, or biometric signal counts) to be displayed on the MPARS screen are aggregated into *elements*. Each element is then weighted in terms of preference to provide a more fine-grained QoE and avoid information overload. For example, *Comfort* includes several weather elements such as ambient temperature, wind speed, weather outlook, relative humidity, and barometric pressure, as well as heart rate and calories burned. Data aggregation is therefore performed on each element based on a proposed weighting to reduce the information to be transmitted and displayed. The different weighting adjustments for the acceptance variables are based on the results obtained in [11] and are combined for the initial configuration of the layout shown in the experimental application, as seen in Figure 2 at the bottom left.

For instance, previously expressed preferences can be converted into percentages, giving an initial preference weight of $30,45\%$ for showing geographic information elements, $24,57\%$ for biometrical, $26,30\%$ for weather, and $18,68\%$ for social information. Geographical, demographic and personal data, as well as time of use, will therefore contribute to an initial setup that can be further enhanced through real-time dynamic tuning, such as geolocation for Points of Interest (PoI).

The fourth step (**Learning and Inference**) ensures that the system can provide meaningful feedback to the user. This step involves two main phases: (i) activity tracking detection, which identifies the sporting activity taking place at the moment, and (ii) inference, which decides what type of information is most relevant to provide to the user at each moment. The ML modules (i) and (ii) help to feed the output with meaningful information that is reflected in the final layout by training a model on the data of preferences for classes of information elements given by a dataset that can be updated with the user's latest preferences (**Final Calibration**). After using (i) to determine what activity is being performed, module (ii) infers from different classes what type of feedback is most appropriate for the user's current state: "*Alerts*", "*Advice/Suggestions*", "*Points of Interest*" or "*Encouraging Goals*".

The use of multiple sensors for activity tracking, particularly in the context of outdoor sports, is an important consideration. Today, most smart systems, gadgets, IoT devices and smartphones have internal sensors that can be used to recognise the required sports activity in a non-intrusive way, generating smart data [10] [27]. The main sensors used for these types of activities are a) accelerometer, b) GPS, c) gyroscope, and d) other sensors such as compass, microphone, camera, proximity sensor, light sensor, temperature sensor, pressure sensor [28]. Mobile sensing should be reduced to a few sensors to minimise the device's power consumption. Ideally, no more than two sensors should be used [29] [30].

The fifth and final step (**Customized Feedback**) involves feedback to the system when the user validates the active information elements received in the visible layout. In this way, customisation occurs by having the user validate the information presented in the dynamic layout according to their preferences simply by answering 'yes' or 'no'. This is done more frequently at the beginning of the user's interaction with the system and less frequently over time. This interaction helps to tailor the MPARS to each user, calibrating the weights of the information elements and tuning the ML module to approximate the ideal for each user, i.e., improving the QoE of the MPARS usage. An example of the changes in the information displayed after this adaptive AR layout process can be observed in Figure 3 (previous layout) versus Figure 4 (after incorporating the user's preferences).

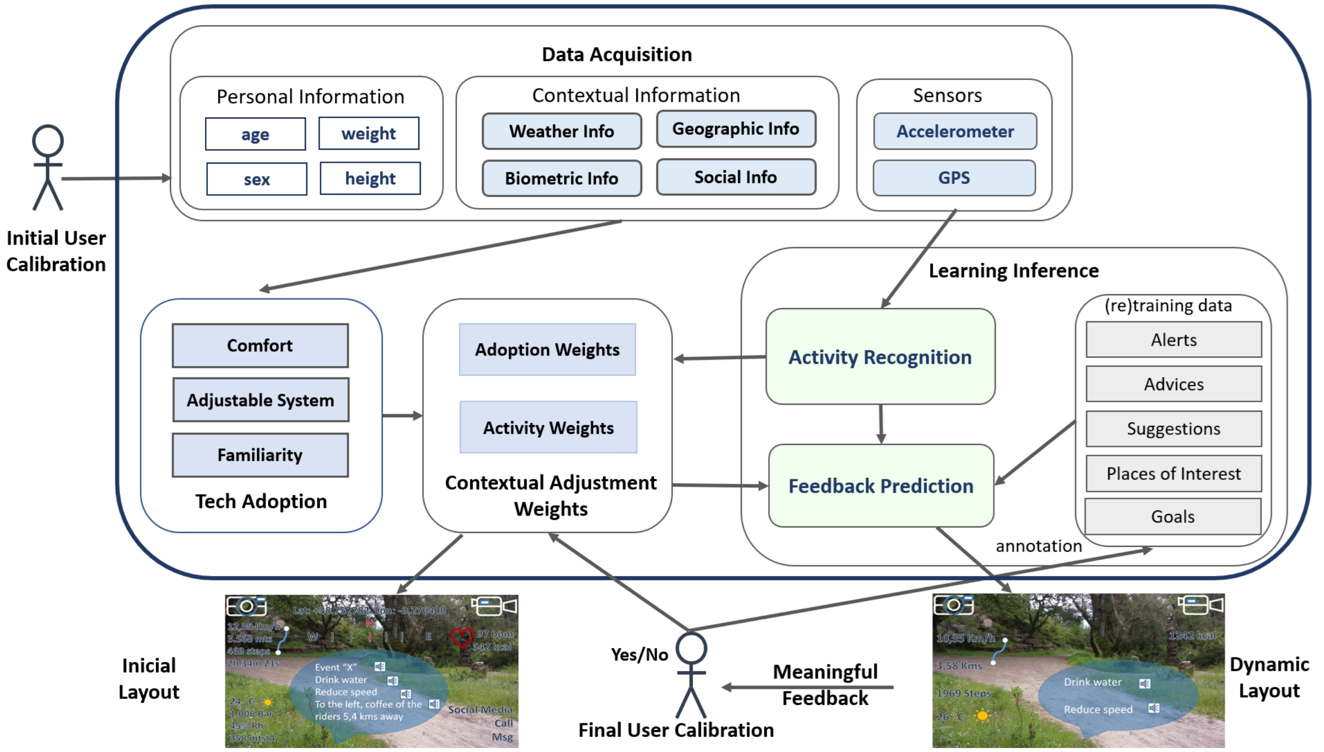

FIGURE 2: Representation of the proposed MPARS architecture and its subcomponents.

FIGURE 3: Representation of the full information in AR layout.

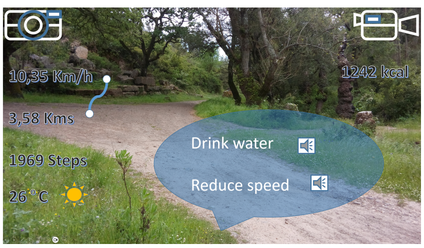

FIGURE 4: Representation in the final AR layout.

IV. MPARS ILLUSTRATIVE USE-CASES

The proposed MPARS architecture (rf. to Section III) has been devised considering the requirements of a use-case based on outdoor usage. In this scenario, a mobile user performs activities and may require continuous adjustment of information received based on the surroundings, personal status, and user preferences. It should be highlighted that the choice of scenario can be applied to different domains, like outdoor sports, smart cities, manufacturing, and health, among others. Next, the usage of the MPARS architecture proposal in the previous section is illustrated via three use cases.

The first scenario relates to a smart city application and gaming-on-the-go. Bob, a 50-year-old user, is carrying his smartphone holding an MPARS gaming application, for which Bob gets continuous rewards based on the level of interaction with the game. The interaction requires Bob to provide regular updates on the city's historical path, e.g., historical landmarks and sightseeing. The MPARS directs Bob to perform specific tasks (like walking towards a specific area to obtain more points). Bob's surroundings (e.g., outdoors), geo-location, and even walking speed are some of the contextual aspects being considered. Based on these aspects, the MPARS adapts the information to be displayed to Bob's device layout. For instance, if Bob is sitting in a coffee shop, the MPARS may ask for more information to be sent, and the type of information (e.g., a simple click or sending a photo) may change.

A second scenario, focused on health, targets well-being awareness. Anne is on her daily run and carries an MPARS integrated into her smart glasses. The health MPARS app provides Anne with the usual fitness information, similar to several personal gadgets that exist today. The key difference is that based on situational awareness and sensed information (such as geo-location), the MPARS may suggest alternative routes to reach her goals or allow Anne to be more motivated

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3603917

to reach her goals based on her interests. The required information adaptation in this case will be strongly related to Anne's interests and also to her context - user behaviour and specific health condition.

The third example focuses on outdoor sports usage. Martin is a mountain biker equipped with an MPARS on a smartwatch. While biking, he interacts with the MPARS via voice commands. The MPARS replies with specific routes based on the current location, weather, road conditions, Martin's interests, and heart rate. The type of information provided is adjusted by Martin's measurable physical and surrounding conditions. The focus is on reducing information based on situational awareness and user preferences.

The described use cases help explain that the proposed architecture has been designed to allow its use across different usage scenarios in different outdoor applications. In the following, a use case focused on outdoor sports activities has been elected to illustrate the usage of an MPARS, the feasibility of the proposed system architecture and the power of user preferences towards information overload reduction. However, this focus does not undermine the broader scope of the proposed architecture, which can easily be adapted for other scenarios such as the ones described in this section.

V. SYSTEM'S FEASIBILITY EVALUATION

In order to establish a proof-of-concept on the viability of developing the proposed architecture and testing the integration of user preferences for costumisation and reducing information overload, a real-world experiment was conducted based on developed middleware. A simple application was designed and implemented to run on a smartphone[1] to provide the user with an initial layout that displays some AR elements and supports the user's ability to request some common functionalities, like taking a photo, making a phone call or opening a social media application while on the move.

This allowed the collection of a first set of data expressing user preferences for receiving several classes of information.

Sports activities performed with the MPARS app involve dynamic mobility, so sports in static facilities, such as aerobics and football, were excluded. Other outdoor sports (such as swimming, diving, and climbing) were also excluded since users may be unable to safely interact with a ubiquitous mobile system (in this case, a mobile phone)[2]

Figure 5 shows the layout of the middleware developed with all informational elements. The remaining elements are functionalities that allow for user interaction (taking photos, filming, calling, sending messages, and social interaction) or prototype-related elements.

A. DATA COLLECTION

Measurements using the user's smartphone sensors were recorded in a controlled outdoor environment in Lisbon,

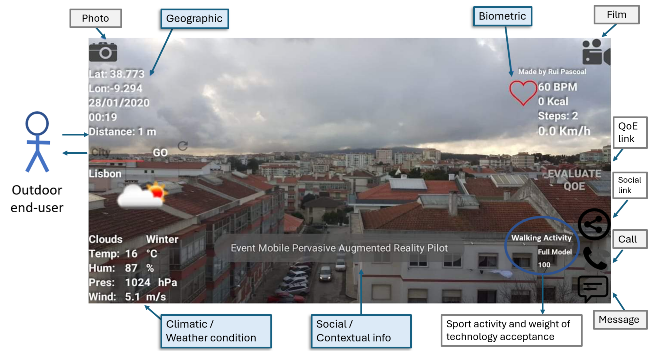

FIGURE 5: Middleware layout example. Informative elements are indicated by the blue boxes, system-related elements by white boxes, and functional elements by the grey ones.

Portugal, between February and September 2022. Outdoor sporting activities were carried out at different times of the year, covering all regular seasons in Lisbon, Portugal (winter, spring, summer, and autumn)[3] The volunteers performed four independent activities. Each activity was carried out for two minutes.

Volunteers installed the application on a smartphone with the Android operating system and at least 4GB of RAM. Then they had to enable the app's permissions and turn on the GPS. While looking at the information on the smartphone display, volunteers were also asked to perform at least one function of their choice (take a photo, make a film, make a phone call, send a message, or interact on a social network). Finally, each participant was asked to indicate their preference for receiving an alert, a suggestion (a POI or advice, for instance), or encouragement to achieve a goal. For example, if a heart rate higher than a threshold is detected, the app could provide an alert by displaying a message such as *"Reduce your speed"*.

One hundred end-users participated in the experiment using this prototype of MPARS and completed a specially designed questionnaire after the field test[4]. The questionnaire integrates eight questions that range from having experienced information overload to what users prefer about the (type of) information that an MPARS application in an outdoor sports context should convey during the activity. The questions also include demographic and gender diversity information, and the duration of the prototype test. The remaining questions aimed to understand which functionalities were chosen and to grade the level of importance of the information displayed. The volunteer should also indicate the frequency of the sporting activity. The participants are volunteers from an academic community in Lisbon, Portugal, with an average age of 36 years, an average weight of 74 kilograms, and an average

[1]MPARS Layout: https://github.com/ruilupas/MPARS/blob/master/MPARS_wireframe.pdf
[2]Demo and data at: https://github.com/ruilupas/MPARS

[3]Data set available at: https://github.com/ruilupas/MPARS/blob/master/logs_mpars_total.xlsx.
[4]Questionnaire available at: https://github.com/ruilupas/MPARS/blob/master/mpars-questionnaire.pdf

height of 1.71 meters, displaying a heterogeneous universe in terms of sexes. The participation was promoted via academic communities' mailing lists and online social networks of Iscte Instituto Universitário de Lisboa and University Lusófona, and research centers (ISTAR_Iscte[5], COPELABS[6], and CISUC[7]). These channels reach the general university population and are not targeted at any specific discipline.

B. SURVEY RESPONSE ANALYSIS

After collecting all the responses to the questionnaire, a descriptive and exploratory analysis was carried out. The percentages of positive responses regarding the prototype functionalities and of users' preference towards receiving several kinds of feedback information are shown in the following figures. In particular, there have been found obvious differences in the answers originating from volunteers less than 37 years of age and those remaining, which can be observed especially in Figures 6, 7, and 9.

By observation of the plot at the right of Figure 6 (Feedback), it is clear that the respondents are very receptive to receiving feedback in almost all existing classes. Nevertheless, they express that an alert might not be relevant, especially for the respondents over 36 years of age. In terms of functionalities (plot at the left), the receptiveness varies. While the main preference is for taking photos (86%), followed by social media (73%), the remaining functionalities are not as expressive as the former: sending messages reaches 55% of preferences, and the less preferred are filming and calling, whose expressed preferences are below 40%. Interestingly, the preferences for using functionalities during training are always lower for users older than 36 than for the youngest users.

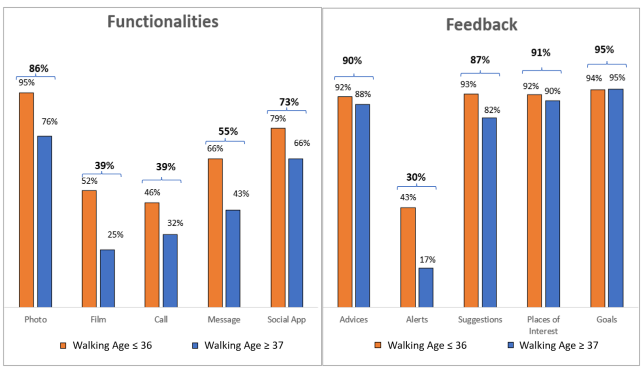

FIGURE 6: Walking: preferred functionalities and feedback information.

However, the walking scenario responses contrast with the results from other activities results. While at the different activities, the respondents still prefer to receive information about personal goals (92% for biking and 95% for running and race-walking), the remaining feedback classes preferences vary between the different activities. For example, when cycling (Figure 7), places of interest (89%) and warnings are considered to be very important (88%). The least preferred feedback information is suggestions (55%) and advice (42%), which are basically in the same category of informational content. In terms of functionalities, calling and recording are the most preferred, with 80% and 69% respectively, while the remaining functionalities only reach about 10% preferences each.

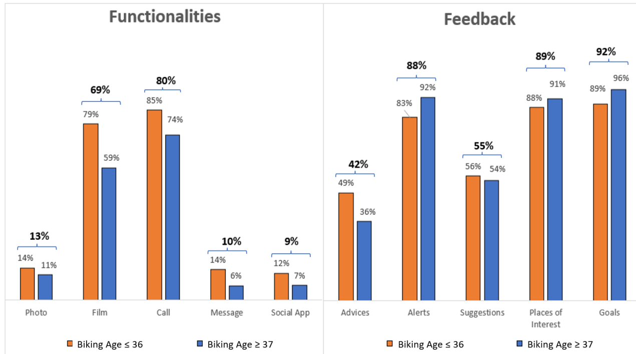

FIGURE 7: Biking: preferred functionalities and feedback information.

For the race-walking activity, respondents ignore all functionalities except for the call functionality. Even then, they only reach 53% (Figure 8). Users also express that they are most interested in receiving information about goals, alerts, and suggestions with 95%, 89%, and 88% respectively. Running presents a preference scenario very similar to race-walking, but with a significant decrease in receiving suggestions, reaching only 55% of user preferences (Figure 9), while alerts are perceived as more relevant (92%).

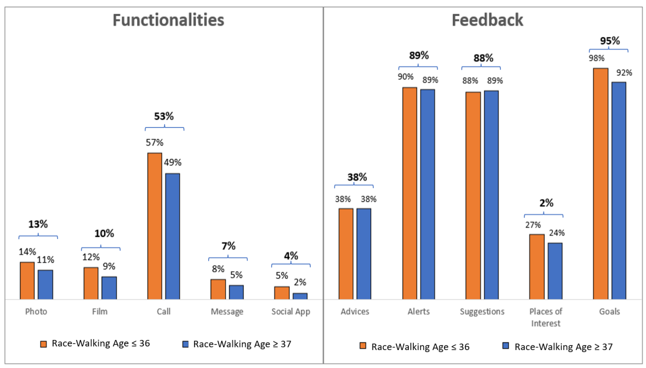

FIGURE 8: Race-walking: preferred functionalities and feedback information.

The higher the intensity of the physical effort, the fewer users care about receiving optional information. In this case, users prefer to receive information about targets (almost 95% of preferences) and warnings (about 90%). It is also clear that age influences the preferences expressed for the different information to be received for each activity, although more so for some sports (such as cycling) than others.

[5] http://istar.iscte-iul.pt/
[6] http://copelabs.ulusofona.pt/
[7] https://www.cisuc.uc.pt/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3603917

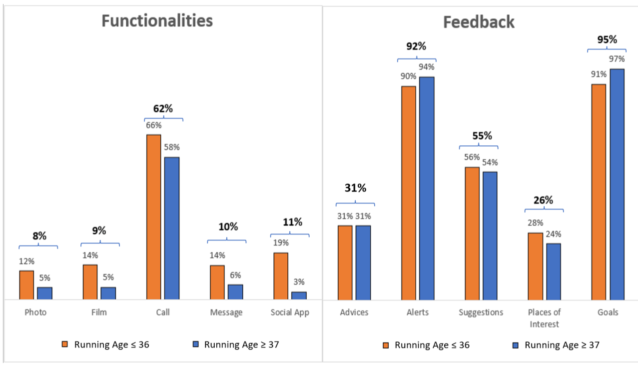

FIGURE 9: Running: preferred functionalities and feedback information.

C. ADAPTIVE FEEDBACK RESPONSE MODULE EVALUATION

The answers to the survey show that preferences for meaningful feedback vary with the activity being performed and are influenced by the age of the practitioner. Furthermore, the expressiveness of preferences may also depend on the location where the activity is performed or other contextual factors. It is therefore important to understand whether the proposed MPARS system, fed with contextual information and end-user expressed preferences, can successfully adapt the information in the layout. The idea is to provide the user with some type of information — suggestion, advice, alert — in a way that is adjusted in time and space, to the situation and context of the user. As such, a proof-of-concept has been developed for testing the viability of the proposed MPARS architecture adaptive feedback module where, once the activity being performed is detected, the opportunity for giving a certain type of feedback to the user is decided by an ML module trained on the feedback preferences data set.

Using the data collected with the questionnaire, selected classification methods have been tested with the aim to understand their accuracy in terms of determining which feedback information should be used in a given context and at a given time. Simple tests of the predictive power of the system were carried out. This was done using the data collected in the field test described in Subsection V-A on preferred information feedback during each activity.

However, even with 100 participants testing the system, the data obtained resulted in very fine-grained data and small data sets. The original data collected consisted of $2,401$ logs for each outdoor sports activity, with 24 logs per user collected every 5 seconds, for a total of $9,604$ logs. To better understand the behaviour, user logs were concatenated into 30-second observations: based on 4 logs per user, the previous 5-second logs were logged into a 30-second log to get the accumulated trend of logs of user interactions. This gives a total of 772 logs of real data observations (a reduction of the original $9,604$ logs).

Based on the original 772 logs, a second synthetic data set containing in total of 1.632 logs has been generated. Hence, the synthetic data set consisted of 2.404 logs, that is, 601 logs per sporting activity. This means that, for each one of the real data observations, four new synthetic instances have been created [8].

Synthetic data was created using random functions, such as the Microsoft Excel *RANDBETWEEN()* function, and data analysis techniques to create systematic samples. For example, when creating a new instance from an existing one, a new data age value was created by imputing a random number generated in the range $[y-5, y+5]$, where y is the real data instance value. Similarly, for temperature, a random value was generated in the range ± 5 degrees Celsius, and for wind speed, the new value was drawn from ± 10 km/h. Relative humidity and air pressure depend on the weather forecast (*clear sky, clouds, rain*). Thus, the last two were created conditionally according to the weather outlook previously drawn (e.g. for a clear sky, lower relative humidity and higher air pressure should be drawn).

Table 1 shows the real and mixed (real + synthetic) data distributions of users' preferences for features and feedback information. Note that the final preference distributions are slightly different from the real data distributions because the synthetic data was drawn uniformly using a predefined randomisation function.

To understand the predictive power of the data to decide on meaningful feedback, several experiments were conducted with models trained on a dataset consisting of the concatenation of real and synthetic data for all the different outdoor sports activities, using the Orange Data Mining tool[9]

It is assumed that a specific activity detection technique has already established the activity being performed. For the final adjustment, the system will rely on trained models to decide if any type of meaningful information should be displayed in the AR layout. As a proof-of-concept, the results obtained by fixing *Biking* as the activity being performed and using a predictive model trained on a binary feedback option target, i.e., one of the expressions of preferences for feedback: either *Advices*, *Alerts*, *Suggestions*, *Interesting Places*, or *Goals* are considered.

The mixed dataset (real data plus synthetic data) consisting of 601 (mixed) data instances corresponding to the cycling activity was uploaded into an Orange workflow. The dataset consists of 20 features with information on personal attributes (age, sex, height, and weight), biometric attributes representing measurements during the activity (such as heart rate, calorie consumption, steps), speed (km/h), distance traveled (m), and weather conditions (ambient temperature (hPa), wind speed (km/h), as well as weather outlook (clear, some clouds, cloudy, rain, relative humidity), and binary expressions of preferences (true or false) for *Advice*, *Alerts*, *Suggestions*, *Interesting Places* and *Goals*. Each data point is also marked as real or synthetic.

[8]Data available at: https://iscteiul365-my.sharepoint.com/:x:/g/personal/rmspl_iscte-iul_pt/EbOk59QRka1Ln_qTNv4Z61ABIn--iC2K3ufVgU6HKudbmw?e=DzBIMT

[9] https://orangedatamining.com/

TABLE 1: Comparison of distributions for preferred feedback between real data and mixed data (real + synthetic) by age.

Activity	Age	Advices		Alerts		Suggestions		POI		Goals	
		Real	Mixed	Real	Mixed	Real	Mixed	Real	Mixed	Real	Mixed
Walking	≤ 36	0.92	0.82	0.43	0.41	0.93	0.88	0.92	0.95	0.94	1.00
	≥ 37	0.88	0.89	0.17	0.17	0.82	0.70	0.90	0.89	0.95	0.99
Race-Walking	≤ 36	0.38	0.25	0.90	0.85	0.88	0.86	0.27	0.27	0.98	1.00
	≥ 37	0.38	0.29	0.89	0.86	0.89	0.84	0.24	0.19	0.92	0.97
Running	≤ 36	0.31	0.22	0.90	0.86	0.56	0.38	0.28	0.29	0.91	0.99
	≥ 37	0.31	0.20	0.94	0.97	0.54	0.35	0.24	0.38	0.97	1.00
Biking	≤ 36	0.49	0.30	0.83	0.79	0.56	0.48	0.88	0.86	0.89	0.84
	≥ 37	0.36	0.23	0.92	0.77	0.54	0.47	0.91	1.00	0.96	1.00
Average preferences		0.50	0.40	0.75	0.71	0.72	0.62	0.58	0.60	0.94	0.97

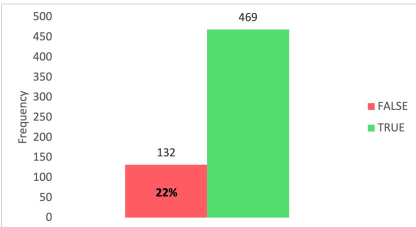

FIGURE 10: Biking activity dataset: distribution of the preferences for the binary target *Alerts*.

The first experiment involved predicting whether an alert would be issued (true) or not (false) for a new data instance, so the target feature will be *Alerts*. Therefore, a hold-out set was randomly sampled from the previously described dataset by sampling a small portion of each target class (approximately 10%). It should be noted that this is an unbalanced classification problem since the distribution of preferences in the examples for the feature *Alerts* consists of 132 negative observations (false labels) and 469 true labels (Figure 10). Consequently, the hold-out sample consists of 61 instances, of which 46 are true labels and 15 are not. A more balanced dataset has been selected for training. The true label class was randomly sampled and contributed 208 observations, which were concatenated with the 117 false label observations to create a training dataset containing 325 instances.

In a recent work [31], the authors' findings suggest that personalisation is most effective when applied with traditional ML techniques rather than deep learning ones, which supports our proposal to use an ML approach for personalisation. Thus, several ML algorithms were used to train models using this more balanced dataset and most of the features described above. Unused features were: the binary indication of whether it was a real or synthetic data point, *Outdoor Activity*, the remaining feedback-giving options (*Goals*, *Interesting Places*, *Advice*, and *Suggestions*), and the attributes *Sex*, *Heart Rate*, and *Footsteps*. The latter is due to the fact that the activity analysed is cycling, and, of course, the number of steps is usually zero. On the other hand, the sex of the user was not used because the distribution of the target values is very similar between the sexes. Finally, heart rate was not considered because it is consistent throughout the set.

The models have been tested using the hold-out instances.

TABLE 2: Models' evaluation using usual metrics averaged over both classes: AUC (area under the curve), CA (classification accuracy), F1 (F1-score), Prec (precision), Recall, and MCC (Matthews correlation coefficient).

Model	AUC	CA	F1	Prec	Recall	MCC
Gradient Boosting	0.981	0.93	0.94	0.94	0.93	0.83
kNN	0.93	0.85	0.86	0.89	0.85	0.68
Tree	0.83	0.82	0.83	0.85	0.82	0.57
Random Forest	0.84	0.79	0.79	0.80	0.79	0.47
Logistic Regression	0.62	0.69	0.70	0.72	0.69	0.25
Naive Bayes	0.65	0.66	0.68	0.74	0.66	0.28

Overall, they have achieved good performance, as can be seen in Table 2. The best performance is achieved by a *scikit-learn* [32] *Gradient Boosting* [10] model, which took 0.105 seconds to train its model and 0.004 seconds to test. The confusion matrix in Figure 11 shows that the performance is good, not only in terms of accuracy but especially because of a very interesting balance between precision and recall, both above 93%.

		Predicted		
		false	true	∑
Actual	false	93.3 %	6.7 %	15
	true	6.5 %	93.5 %	46
	∑	17	44	61

FIGURE 11: Confusion matrix resulting from the test of the model *Gradient Boosting*.

Notably, the second best model is the 5-*kNN* model, whose confusion matrix is shown in Figure 12. Compared with *Gradient Boosting*, the *kNN* model loses sensitivity, increasing the number of false negatives. Nevertheless, the model correctly predicts the true label in more than 82% of the cases.

The good performance of *KNN* was an interesting result. Since this technique is very light in terms of retraining the model with new labelled examples, it is suitable for the MPARS implementation, where user feedback is used for system adaptation and tuning. The MPARS proposed here relies on the computational power of the underlying device

[10] https://orangedatamining.com/widget-catalog/model/gradientboosting/

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3603917

		Predicted		
		false	**true**	**∑**
Actual	**false**	93.3 %	6.7 %	**15**
	true	17.4 %	82.6 %	**46**
	∑	**22**	**39**	**61**

FIGURE 12: Confusion matrix resulting from the test of the model 5-*kNN*.

for the continuous adaptation of the AR layout, so the easier the retraining of the ML model, the better. In this sense, *Naïve Bayes* would be preferable, but its performance in this test is quite inferior.

These are very simple experiments where no special preparation was used to train the models, and they demonstrate the feasibility of the MPARS system proposed in Section III, where models trained on data without any special preparation and preprocessing show good performance in terms of reducing information overload. In fact, the results of the models *Gradient Boosting* and *KNN* (Figures 11 and 12, respectively) are quite accurate for both true positive and true negative hits but are particularly robust for true negative results (over 93%). This suggests the feasibility of a working MPARS prototype capable of real-time adaptation based on expressed user preferences and ML techniques.

VI. CONCLUSIONS

This work proposes a first architectural framework for the implementation of a mobile pervasive augmented reality system or MPARS specifically designed for dynamic outdoor environments involving sports activities, taking into account technology acceptance indicators derived from end-users. The work also provides a proof-of-concept for the feasibility of the proposed MPARS prototype architecture, which is capable of real-time adaptation based on a collected sample of end-user preferences and volume of information criteria indicators to allow adaptation of the system and avoid information overload in the augmented reality display. The indicators considered in this work are based on geographical, meteorological, biometric and social information. Together with personal and biometric data, these elements work in a real-time adaptation system to adjust the volume of information and reduce information overload to provide the right information at the right time, thus improving end-user QoE in AR sports environments.

From the analysis of the questionnaire obtained after a field test, it is clear that aspects such as age or activity impact the interest in getting specific information. Speed, integrally linked with the activity, is, naturally, an additional factor of impact in information preferences, which is directly related to the user's speed, as users feel more overloaded when performing dynamic activities such as running and cycling than when performing less dynamic activities such as walking. The higher the intensity of the physical effort, the lower the intention to receive optional information. Furthermore, on average, the expressed preferences tend towards receiving information on personal *Goals* (94%), *Suggestions* (close to 81%) and *Alerts* such as weather or route conditions (close to 75% of preferences), indicating that this information should be prioritised.

The MPARS framework proposed here is that of a context-aware system both in terms of its user and in terms of the surrounding environment and activity. All variables that determine the information elements of the layout are to be adapted both from data obtained via the end user's device (e.g., accelerometer) and from other online sources (e.g., weather data). Calibration of the user with personal data (such as age and height) is also essential to better adapt the system to the user. The final calibration of the AR layout is determined by an intelligent module based on the device's sensors, the determination of the activity being performed, the end user's personal calibration, and trained models for controlling the information elements to be displayed at any given moment.

The findings of this work open opportunities for future work in the field of real user preferences, serving as a basis for realistically approximating preferences for outdoor sports activities for a greater diversity of users. It will also be relevant to better explore the possibility of *encouraging goals* by introducing a *gamification* component, to help users fulfill sporting objectives, especially when competing with other users.

As usual, this work has its limitations. In order to implement a functional prototype, a more operational application needs to be developed and distributed to a larger cohort of volunteers to be used over a wider range and duration of sporting activity types and to allow a more in-depth investigation of the system's performance. Furthermore, comparative investigations should be implemented to validate the usability of the proposed MPARS *vs.* a solution that does not support adaptation and collection of user feedback. This research should also address the use of device resources in terms of energy and computational load. The implementation of ML modules presents its challenges. Unlike activity detection, where a large number of approaches already exist, the same cannot be said for automated adaptation systems. Especially in the present case, where the feedback of information concerns various possibilities for providing informational elements concurrently, a deeper and comprehensive investigation is needed, integrating all possible feedback elements and a measure for the degree of information overload experienced.

REFERENCES

[1] F. Zhou, H. B.-L. Duh, and M. Billinghurst, "Trends in augmented reality tracking, interaction and display: A review of ten years of ismar," in Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality. IEEE Computer Society, 2008, pp. 193–202.

[2] Y. Siriwardhana, P. Porambage, M. Liyanage, and M. Ylianttila, "A survey on mobile augmented reality with 5g mobile edge computing: Architectures, applications, and technical aspects," IEEE Communications Surveys & Tutorials, vol. 23, no. 2, pp. 1160–1192, 2021.

[3] J. Cao, K.-Y. Lam, L.-H. Lee, X. Liu, P. Hui, and X. Su, "Mobile augmented reality: User interfaces, frameworks, and intelligence," ACM Computing Surveys, vol. 55, no. 9, 2023, cited by: 14; All Open Access, Bronze Open Access, Green Open Access.
[4] I. Irwanto, R. Dianawati, and I. Lukman, "Trends of augmented reality applications in science education: A systematic review from 2007 to 2022," International Journal of Emerging Technologies in Learning (iJET), vol. 17, no. 13, pp. 157–175, 2022.
[5] B. D. Sawyer, V. S. Finomore, A. A. Calvo, and P. A. Hancock, "Google glass: A driver distraction cause or cure?" Human factors, vol. 56, no. 7, pp. 1307–1321, 2014.
[6] R. M. Pascoal and S. L. Guerreiro, "Information overload in augmented reality: The outdoor sports environments," in Information and Communication Overload in the Digital Age. IGI Global, 2017, pp. 271–301.
[7] D. Van Krevelen and R. Poelman, "A survey of augmented reality technologies, applications and limitations," International Journal of Virtual Reality, vol. 9, no. 2, p. 1, 2010.
[8] M. J. Eppler and J. Mengis, "The concept of information overload: A review of literature from organization science, accounting, marketing, mis, and related disciplines," The Information Society, vol. 20, no. 5, pp. 325–344, 2004. [Online]. Available: https://doi.org/10.1080/01972240490507974
[9] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies, "The case for vm-based cloudlets in mobile computing," IEEE pervasive Computing, vol. 8, no. 4, 2009.
[10] R. M. Pascoal, A. de Almeida, and R. C. Sofia, "Activity recognition in outdoor sports environments: smart data for end-users involving mobile pervasive augmented reality systems," in Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers. ACM, 2019, pp. 446–453.
[11] R. Pascoal, A. D. Almeida, and R. C. Sofia, "Mobile pervasive augmented reality systems—mpars: The role of user preferences in the perceived quality of experience in outdoor applications," ACM Transactions on Internet Technology (TOIT), vol. 20, no. 1, pp. 1–17, 2020.
[12] N. Sawan, A. Eltweri, C. De Lucia, L. Pio Leonardo Cavaliere, A. Faccia, and N. Roxana Moşteanu, "Mixed and augmented reality applications in the sport industry," in Proceedings of the 2020 2nd International Conference on E-Business and E-commerce Engineering, 2020, pp. 55–59.
[13] S. K. Kim, S.-J. Kang, Y.-J. Choi, M.-H. Choi, and M. Hong, "Augmented-reality survey: from concept to application." KSII Transactions on Internet & Information Systems, vol. 11, no. 2, pp. 982–1004, 2017.
[14] Z. Lv, A. Halawani, S. Feng, S. ur Réhman, and H. Li, "Touch-less interactive augmented reality game on vision-based wearable device," Personal and Ubiquitous Computing, vol. 19, no. 3, pp. 551–567, 2015. [Online]. Available: https://doi.org/10.1007/s00779-015-0844-1
[15] M. Billinghurst, A. Clark, G. Lee et al., "A survey of augmented reality," Foundations and Trends® in Human–Computer Interaction, vol. 8, no. 2-3, pp. 73–272, 2015.
[16] A. Bayat, M. Pomplun, and D. A. Tran, "A study on human activity recognition using accelerometer data from smartphones," Procedia Computer Science, vol. 34, pp. 450–457, 2014.
[17] J. Schmitt, M. Hollick, C. Roos, and R. Steinmetz, "Adapting the user context in realtime: Tailoring online machine learning algorithms to ambient computing," Mobile Networks and Applications, vol. 13, pp. 583–598, 2008.
[18] Y. Liu and L. Liu, "Analysis of auxiliary modes for sports intelligence training system based on nonlinear model optimization and improved algorithms," Soft Computing, pp. 1–11, 2023.
[19] L. Stacchio, V. Armandi, S. Hajahmadi, L. Donatiello, and G. Marfia, "M-agew: Empowering outdoor workouts with data-driven augmented reality assistance," in 2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR). IEEE, 2024, pp. 301–305.
[20] P. Soltani and A. H. Morice, "Augmented reality tools for sports education and training," Computers & Education, vol. 155, p. 103923, 2020.
[21] S. Zollmann, T. Langlotz, R. Grasset, W. H. Lo, S. Mori, and H. Regenbrecht, "Visualization techniques in augmented reality: A taxonomy, methods and patterns," IEEE Transactions on Visualization and Computer Graphics, vol. 27, no. 9, pp. 3808–3825, 2021.
[22] M. Tatzgern, V. Orso, D. Kalkofen, G. Jacucci, L. Gamberini, and D. Schmalstieg, "Adaptive information density for augmented reality displays," in 2016 IEEE Virtual Reality (VR), 2016, pp. 83–92.
[23] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information." Psychological review, vol. 63, no. 2, p. 81, 1956.
[24] D. Bawden and L. Robinson, "Information overload: An introduction," in Oxford research encyclopedia of politics, 2020.
[25] B. Furlow, "Information overload and unsustainable workloads in the era of electronic health records," The Lancet Respiratory Medicine, vol. 8, no. 3, pp. 243–244, 2020.
[26] D. K. K. Reddy, H. S. Behera, J. Nayak, A. R. Routray, P. S. Kumar, and U. Ghosh, A Fog-Based Intelligent Secured IoMT Framework for Early Diabetes Prediction. Cham: Springer International Publishing, 2022, pp. 199–218. [Online]. Available: https://doi.org/10.1007/978-3-030-81473-1_10
[27] R. Sofia, S. Firdose, L. A. Lopes, W. Moreira, and P. Mendes, "Nsense: A people-centric, non-intrusive opportunistic sensing tool for contextualizing nearness," in 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), 2016, pp. 1–6.
[28] A. Kos, S. Tomažič, and A. Umek, "Evaluation of smartphone inertial sensor performance for cross-platform mobile applications," Sensors, vol. 16, no. 4, p. 477, 2016.
[29] M. Youssef, M. A. Yosef, and M. El-Derini, "Gac: energy-efficient hybrid gps-accelerometer-compass gsm localization," in 2010 IEEE Global Telecommunications Conference GLOBECOM 2010. IEEE, 2010, pp. 1–5.
[30] R. M. Pascoal, "Battery efficiency in outdoor sports environments for mobile pervasive augmented reality systems," in Handbook of Research on Sustainable Development Goals, Climate Change, and Digitalization. IGI Global, 2022, pp. 520–536.
[31] A. Ferrari, D. Micucci, M. Mobilio, and P. Napoletano, "Deep learning and model personalization in sensor-based human activity recognition," Journal of Reliable Intelligent Environments, vol. 9, pp. 27–39, 2023.
[32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

RUI MIGUEL PASCOAL (PhD student, MSc, BSc Computer Engineering) is a student and Research Assistant at ISTAR-IUL - Information Sciences, Technologies and Architecture Research Center. He is an Assistant Professor at ISTEC - Higher Institute of Advanced Technologies, and at Atlantica – University Institute, where he teaches several curricular units such as Software Engineering, IT Project Management, Systems Administration, and Structured Programming, among others. Previously, Rui pursued a career as a technician at HP (1998-2000) focusing on configuring UNIX and NT servers; at IBM (2004-2014) operating Z/OS Mainframes in Critical Recovery, software distribution and biometric identity management systems; at CTT Correios of Portugal (2015-2018) focusing on Management Planning and Control, Business Intelligence, Executive Information System and SAP Business Warehouse. His current research interests are information overload, activity recognition, mobile augmented reality technology, speech recognition, and battery efficiency.

[] Ana Maria de Almeida (PhD 04, IEEE Senior Member, ACM member) is an Associate Professor at Information Science and Technologies Department of Iscte Instituto Universitário de Lisboa, and Researcher at ISTAR-IUL - Information Sciences, Technologies and Architecture Research Center, where she is the Coordinates the Software Systems Engineering research group, and at the Cognitive and Media Systems research group of CISUC – Center for Informatics and Systems of the University of Coimbra. Her current research interests lie within the areas of Algorithmics, Complexity, ML and Pattern Recognition, Data Science, Evolutionary Computation, and Ethics for AI and Research. She also has a particular interest in the development of evolutionary strategies for tackling multicriteria combinatorial problems, as well as in the construction of self-adjusting predictive and reactive models for real applications.

[] Rute C. Sofia (PhD 04, IEEE Senior Member) is Head of Industrial IoT at the Fortiss Research Institute and Invited Associate Professor at the University Lusófona de Humanidades e Tecnologias; Associate Researcher at ISTAR, Instituto Universitário de Lisboa. Rute's research background has been developed in industry (Grupo Forum, Lisbon; Siemens AG, Nokia Networks, Munich) and academia (FCCN, Lisbon; INESC TEC, Porto; ULHT, Lisbon; Universität der Bundeswehr, Munich). She was the scientific director of the Portuguese research unit COPELABS (2013-2017) and co-founder of COPELABS and the Portuguese startup Senception Lda (2013-2019).

Her current research interests are: network architectures and protocols; IoT; edge computing; edge AI; in-network computing; 6G. Rute has over 80 peer-reviewed publications and 9 patents in her areas of interest. She is the IEEE ComSoc WICE Deputy Liaison for Industry; and an ACM Europe Councilor. She leads the 6G CONASENSE platform. She was co-chair of the IEEE ComSoc Awards (20-21). She is an Associate Editor for several venues including IEEE Access, IEEE Network.

• • •