IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 April 2024, accepted 2 May 2024, date of publication 22 May 2024, date of current version 30 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3402230

APPLIED RESEARCH

Open Dataset for Predicting Pilgrim Activities for Crowd Management During Hajj Using Wearable Sensors

ALI M. AL-SHAERY[1], (Member, IEEE), SOHA G. AHMED[2,3], HAMAD ALJASSMI[2,4], ABDULLAH N. AL-HAWSAWI[5], NADIA MAKSOUD[2], ABDESSAMAD TRIDANE[2,6], AND NORAH S. FAROOQI[7,8], (Member, IEEE)

[1]Department of Civil Engineering, College of Engineering and Architecture, Umm Al-Qura University, Makkah 24381, Saudi Arabia
[2]Emirates Center for Mobility Research (ECMR), United Arab Emirates University, Al Ain, United Arab Emirates
[3]Department of Information Technology, College of Information Technology, United Arab Emirates University, Al Ain, United Arab Emirates
[4]Department of Civil and Environmental Engineering, College of Engineering, United Arab Emirates University, Al Ain, United Arab Emirates
[5]Department of Information and Scientific Services, Umm Al-Qura University, Makkah 24382, Saudi Arabia
[6]Department of Mathematical Sciences, College of Science, United Arab Emirates University, Al Ain, United Arab Emirates
[7]Department of Cybersecurity, Collage of Computing, Umm Al-Qura University, Makkah 21955, Saudi Arabia
[8]Dar Al-Hekma University, Jeddah 22246, Saudi Arabia

Corresponding author: Hamad Aljassmi (h.aljasmi@uaeu.ac.ae)

This work was supported by the Deputyship for Research and Innovation, Ministry of Education, Saudi Arabia, under Grant RDO-P543.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Biomedical Research Ethics Committee from Umm Al-Qura University Approval No. HAPO-02-K-012-2023-03-1602.

ABSTRACT This study aims to create and examine a multimodal dataset to enhance crowd management during the Hajj seasons. Sixty-four participants were engaged in Hajj rituals such as Tawaf, Saai, prayer, and Doaa providing location and peripheral physiological data, collected and annotated using a custom-made smartphone application. The collected data was leveraged to conduct a comprehensive analysis, specifically focusing on the classification of the type of Hajj activity, level of fatigue, and emotional states based on peripheral physiological signals. Three deep learning classification models were developed and validated using feedforward neural networks. The models achieved satisfactory accuracy scores in classifying the type of Hajj activity (41.71%), level of fatigue (85.27%), and emotional states (82.47%). While presenting a straightforward use case, this research chiefly provides decision makers and the scientific community with a statistically significant open data set aside with a deep learning architecture capable of characterizing crowd behavior for the purpose of automating crowd management and monitoring.

INDEX TERMS Crowd control, crowd management, deep learning, Hajj, machine learning, pattern classification, physiological sensors, signal processing.

I. INTRODUCTION

Crowd behavioral management studies how people interact and behave in large groups, as well as how to glean useful information from recordings that contain large groups of people. This can be useful for spotting potentially harmful circumstances before they turn into disaster by monitoring the behavior of a large number of people and automatically analyzing the movement of people at a congested pilgrimage locations [1]. Social psychology studies have shown that an individual will behave differently when they are in a crowd. This is due to the influence that other individuals in the crowd, as well as the environment, have on cognition and action [2].

Many theoretical or empirical studies have been conducted over the past few decades to better understand the dynamics of pedestrian crowds [3], [4], [5]. In order to analyze crowd behavior, various interpretations can be made. The main goal is to understand how a group of people behave by using

The associate editor coordinating the review of this manuscript and approving it for publication was Yang Tang.

data that is mainly gathered using different data sources. Crowd safety can be ensured by gathering data on parameters like population density and size using crowd monitoring technology [6]. This information can be used to estimate the degree of crowd gathering and make decisions about crowd management and safety planning. The analysis of a number of crowd behavior factors could help in avoiding dangerous pedestrian stampedes. This involves analyzing the quantity of individuals present in a specific area and its fluctuations over time, along with the primary directions of movement within a mobile crowd. Examining these elements could aid in the detection of individuals whose motion diverges from the main flows, and in identifying the causes behind their deviation. [1]. Based on research done by [2], crowd behavioral analysis has been categorized into four main steps. The four main tasks are crowd behavior classification, crowd counting, people detection and tracking, and crowd anomaly detection, in which the identification and classification of the behavior are done in advance. Then the number of people involved in the sample is included. People detection and tracking involve following the trajectory of pedestrians in a video. Finally, detecting abnormal behaviors in a crowd in the anomaly detection approach.

Despite the valuable insights provided by previous research in crowd behavioral management and the development of numerous crowd monitoring technologies, a significant gap remains in our understanding of crowd dynamics, especially in unique situations like the Hajj pilgrimage. This is largely due to the lack of comprehensive, multimodal data that can account for the complex interplay between individual cognition, physiological states, emotional levels, and crowd behavior. Addressing this gap, the necessity and innovation of this research lie in the creation of an open dataset that links crowd demographics, physiological signals, location data, fatigue levels, and emotional statuses with the participants' performed Hajj activities. This dataset stands as an innovative addition to the field, as it facilitates a more nuanced understanding of crowd dynamics by considering physiological and emotional factors, thus offering a multifaceted view of crowd behavior.

The objective of this research is to collate a comprehensive dataset during Umrah and Hajj seasons encapsulating diverse aspects such as activities, physiological signals, and emotional states of the participants while examining the statistical significance of the data to develop various types of analyses and predictive models. Such analyses and predictive models would provide an invaluable foundation for intelligent crowd management and anomaly detection, achieved through inter-relating crowd demographics, physiological signals, location data, fatigue and emotional states with their corresponding Hajj activities. In addition, this endeavor is designed to provide the scientific community with an open dataset that could spur on more in-depth research and innovation in this field.

This paper is organized as follows: Section II provides an overview of related research. Section III details the data collection method and the considerations taken in its design. Section IV provides quantitative and statistical analysis of the data. Section V presents our experiments and the associated results. Finally, Section VI concludes the paper, emphasizing future directions for this work.

II. RELATED WORK

Different methods have been proposed in the literature for crowd management for different purposes. One such approach was developed by [7] that enables automated real-time identification of critical mass behavior using optical flow computations. The primary objective of this technique is to enhance simulations of crowd motion to identify areas that may pose potential danger and guide pedestrian traffic accordingly. To validate this approach, researchers examined microscopic pedestrian behavior in congested situations and conducted experiments on both synthetic and real datasets. In a different study, [8] used video clips annotated with labels for both crowd behavior and emotions to investigate how emotion features can aid in the recognition of crowd behavior. According to the study, complex behaviors can be accurately predicted from video by using emotional attributes as fundamental building blocks.

Moreover, a crowd management approach was developed in [9] that is based on crowd behavioral analysis and the Internet of Things (IoT). The proposed approach used video segments captured from surveillance cameras to localize the locations that require crowd management into Spatio-temporal flow blocks. This allows for the arbitrarily dense flow field to be marginalized. The same method of video-surveillance data has been used in the [10] study to build a new dataset defining 11 classes that depict how a crowd can evolve across time. The study showed that Spatio-temporal Convolutional Neural Network could be used to crowd characterization and anomalies classification.

Other studies have used network-based data in their investigations. For example, social media data usage has been used in [11] for crowd characterization. The data were recorded for different events, one held in Amsterdam and the other in the Netherlands, for different periods. A set of factors that influence the pedestrian's behavior was identified such as demographics, spatial-temporal distribution, Point of Interest (PoI) preferences, and word use, in the context of city events can be exported from social media.

Similarly, a study by [12] used a similar method to evaluate the datasets' most crucial characteristics and to list the social sensing scenarios that each dataset best captures. This information can be useful for those interested in using the datasets to evaluate and assess Mobile crowd sensing (MCS) solutions for specific mobility/social scenarios. The focus of this study's analysis has been on users' co-location traces, which record the start and finish times of user contacts (i.e., between their carried devices). This condensed representation of the dynamic encounter graph from a dataset illustrates the sociality and mobility of the participants in the experimental datasets.

In contrast, there are some studies that have used installed sensors to track and analyze pedestrian crowds in order to provide an early warning about potential violations of physical distancing rules. This approach can help identify areas where people are most likely to violate the rules and can also provide real-time information about the presence of distance offenders [13]. Meanwhile, other studies relied on mobile sensors and platforms to collect their data for the analysis [10], [14], [15].

Given these various approaches, two primary aspects emerge as crucial in the current state of the field: crowd management and public safety, and data collection infrastructure. These aspects not only frame the recent developments but also guide our understanding of existing challenges and potential improvements.

A. CROWD MANAGEMENT AND PUBLIC SAFETY

Effective crowd management plays a pivotal role in maintaining safety in public areas. It encompasses a comprehensive understanding of how people navigate different spaces and aims to optimize their movement while minimizing potential risks. In light of the COVID-19 pandemic, the significance of crowd management has become even more pronounced. The ability to comprehend and model pedestrian mobility patterns has emerged as a crucial aspect for implementing local policies and restrictions [16], [17].

By observing and analyzing the predictable patterns of people's movements, authorities can enhance their management strategies in public spaces, ensuring compliance with social distancing guidelines. Moreover, when contact tracing becomes necessary, insights into pedestrian mobility patterns can aid in identifying specific areas where individuals may have been exposed to the virus [17].

One notable study [18] has devised a real-time crowd monitoring and management system to combat the spread of COVID-19. This system employs surveillance images and videos to detect instances of social distancing violations, subsequently categorizing them accordingly. By promptly identifying and addressing such violations, the system aims to mitigate the risk of virus transmission.

Crowd dynamics can also impact individuals' emotional, behavioral, and physiological states. In 2020, a study conducted by [19] investigated the potential effects of various interventions in the tragic 2010 Love Parade incident, which resulted in 21 fatalities and numerous injuries. Using computer simulations and virtual reality (VR) experiments, the study examined simulated casualties and physiological arousal. The findings highlighted the importance of separating the inflow and outflow of a crowd, as it can enhance crowd throughput while reducing congestion.

In summary, research at the intersection of crowd management and public safety underscores the need to understand and optimize pedestrian movement, especially in the context of the COVID-19 pandemic. By leveraging insights from pedestrian mobility patterns, authorities can implement effective crowd management strategies, enforce social distancing guidelines, and mitigate the risk of virus transmission. Furthermore, studies exploring crowd dynamics contribute valuable knowledge towards improving safety measures and enhancing the overall well-being of individuals in public spaces.

B. DATA SOURCE (COLLECTION INFRASTRUCTURE)

As seen in the preceding overview, a multitude of data collection methods and tools are deployed to gather valuable crowd management information. These tools and methods can be broadly classified based on the study by [16] into two main collection infrastructures: network infrastructure and end-user devices. These data sources provide crucial insights that allow for a range of analyses and applications in crowd management. The next part of this section will further explore the nuances of these data collection infrastructures and their roles in contributing to our understanding of crowd dynamics.

The end user device data can be exported either from the device itself or from a smart technology application that is linked to the device. Recently, many smart devices such as smartphones, tablets, and laptops are used to collect and store data such as location, motion tackling, health condition, and other types of data. Moreover, devices such as smartwatches have been used also to monitor health conditions and to obtain crowd density in a crowded event. Smartphones are used to collect various types of data. For example, the location of the user can be tracked using the built-in GPS sensor. The motion of the user can be tracked using the built-in accelerometer sensor. In addition, the health condition of the user can be monitored using the built-in heart rate sensor. The use of end-user devices for crowd management can provide many benefits such as the reduction of data collection costs, the minimization of device deployment costs, and the reduction of the number of sensors used [20]. Network infrastructure data can be acquired from devices such as CCTV cameras, RFID readers, WiFi devices, and sensors. These devices are placed in a fixed area and are used to monitor the crowd and gather data. According to [16], data collected can be divided into three categories based on the type of collection:

- Sensor-Based data is collected via the sensors including GPS receivers, light sensors, accelerometers, gyroscopes, magnetometers, pedometers, oximeters, temperature sensors, heart rate sensors, and cameras. The data collected by these sensors can provide information about an individual's location, movements, and activities. Sensors are used to detect various physical quantities and convert them into electrical signals. These signals can be used to trigger other processes or to provide raw data for further analysis. Microsoft's Geolife dataset [21] is a good example of such dataset.
- Network-based data is data collected by network providers about end-user mobility. The growing popularity of mobile communications has made it possible for people to connect to networks from anywhere at any time. This has made it possible for network providers to collect information about end-user mobility through access to network devices. Examples of network connectivity

information that can be collected through mobility traces include the type of network used (WiFi, Bluetooth, or cellular), the location of the user, and the approximate number of people in a specific area. This information can be used to understand the activities and behavior of people using cellular and WiFi networks. Example of these datasets include: the upb/hyccups dataset [22], the yonsei/lifemap dataset [23] and the sigcomm2009 dataset [24].

- Other system hardware refers to the applications running on a smartphone that is capable of collecting data. For example, an app with permission to access the user's location can collect GPS data. Apps can also collect data about the user's activities, such as the apps they use, how often they use them, and for how long. Application data can also include information that is actively collected by users when they are utilizing an application. This data is usually in the form of timestamps and can be paired with details about a person's position, movement, and restrictions in the digital world. Exmaple of such datasets include Gowalla and Brightkite [25].

C. AVAILABLE PLATFORMS WITH DATASET

There are several platforms and simulation models that have been used in the past by researchers in order to experiment with applications in the real world and test their performance. The data sets used in the simulation were gathered from actual studies using built mobile applications or platforms to perceive and gather data [26].

APISENSEand SenseMyCity, MOSDEN, and ParticpAct are examples of such platforms that give resources to store and interpret data collected from a crowd, allowing researchers to use crowdsensing applications. ParticipAct and APISENSE provide resources for storing and processing data, while SenseMyCity focuses on acquiring geo-indexed data from sensors [27], [28], [29]. The SenseMyCity platform includes the engagement of data from a mobile device or participants. It can manage information gathered from a variety of sources, including GPS, accelerometer, gyroscope, and ambient sensors [29]. MOSDEN is a mobile sensor data engine that gathers and distributes sensed data among various users and dispersed applications [30].

Another available platforms are CRATER, Medusa, Prism, and MATADOR. CRATER estimates road conditions and provides APIs for data access and visualization of maps in related applications [31]. Medusa [32] is a framework for analyzing task requirements and distributing the task among users. The Prism [33] platform is designed for remote sensing using smartphones, with a focus on security and scalability. On the basis of users' context awareness, the MATADOR platform focuses on energy-efficient job allocation and execution. It seeks to effectively distribute tasks to users in accordance with a context-aware sampling method that reduces mobile devices' energy consumption [34]. Table 1 list publicly accessible datasets of pedestrian movement traces.

D. LIMITATIONS AND CHALLENGES

While the existing body of work presents a diverse range of methodologies and objectives in crowd management research, certain limitations and challenges persist in the current approaches. These limitations provide an impetus for further exploration and innovation in this field.

One of the primary limitations lies in the ability to accurately classify complex crowd behaviors and emotional states. The current methods, while sophisticated, do not always provide a comprehensive picture of crowd dynamics. For instance, while video surveillance and analysis have been widely used in crowd management studies, these methods often struggle with real-time identification of complex crowd behaviors and emotions. They can also be hindered by challenges such as occlusions, varying lighting conditions, and rapidly changing crowd densities.

Furthermore, the reliance on network infrastructure for data collection can also present certain challenges. Although network-based data can provide rich insights into crowd dynamics, these methods often suffer from issues of privacy, data security, and data integrity. Furthermore, network coverage and connectivity can significantly influence the quality and reliability of collected data, presenting challenges in areas with poor network infrastructure.

The use of end-user devices for data collection, while advantageous in terms of cost-effectiveness and scalability, also comes with its own set of challenges. These devices often rely heavily on battery life, which can limit the duration and frequency of data collection. Additionally, variations in device types, models, and operating systems can introduce inconsistencies in the data, making it difficult to standardize and compare data across different devices.

Finally, while existing platforms provide valuable resources for storing and processing crowd data, they often lack the ability to handle the high dimensional and complex nature of real-world crowd data. In addition, these platforms often have limitations in terms of scalability, interoperability, and real-time data processing capabilities.

In light of these limitations and challenges, our work aims to address these gaps by proposing a novel approach to crowd management using deep learning algorithms. This approach not only offers the potential to improve the accuracy and efficiency of crowd behavior classification but also provides a promising solution to some of the existing challenges in data collection and processing. We believe our work contributes a meaningful step forward in the ongoing efforts to improve crowd management strategies and systems.

Our work boasts several noteworthy accomplishments. Firstly, we have accumulated and made publicly available data that can be utilized by other researchers as a valuable resource. To the best of our knowledge, there is no open-source pedestrian dataset for Hajj and Umrah crowd mobility that is publicly accessible. Secondly, we illustrate the transformation of raw time series physiological data, accelerometer data, and location data into inputs that can be processed by deep learning algorithms to recognize Hajj activities. Finally, our work aims to shed light on the potential of analyzing and managing crowd behavior by analyzing wireless sensor data and we hope it will inspire further research in this field.

III. METHODOLOGY

A. BACKGROUND

The activities considered in this study, namely "Tawaf, Saai, prayer, and Doaa," are specific to the Hajj pilgrimage, which is a religious pilgrimage to Mecca, Saudi Arabia, performed by Muslims. Here's a brief explanation of each activity:

- Tawaf: Tawaf refers to the act of circumambulating the Kaaba, which is a cubic structure located in the center of the Masjid al-Haram (the Grand Mosque) in Mecca. Pilgrims walk around the Kaaba in a counterclockwise direction for 7 rounds.
- Saai: Saai involves walking between the hills of Safa and Marwa, which are located near the Kaaba. Pilgrims walk back and forth seven times, emulating the actions of Hagar, the wife of Prophet Ibrahim (Abraham), who searched for water in that area according to Islamic tradition.
- Prayer: During the Hajj pilgrimage, participants engage in various prayers at different times and locations within the holy sites of Mecca. These prayers are an essential part of the religious rituals performed during Hajj.
- Doaa: Doaa refers to supplications or prayers made by pilgrims during their Hajj journey. These prayers can be personal or collective, and individuals often make specific requests or seek blessings during this time.

In this study the location data and peripheral physiological signals of the participants were recorded as they performed these activities. This means that the participants' physical location and physiological signals (such as heart rate, respiration rate, etc.) were monitored and collected during Tawaf, Saai, prayers, and Doaa. The participants also provided annotations related to the activity lap number (if applicable), their level of fatigue, and emotional status. The collected data, including the participants' annotations and real-time physiological and location data, were streamed and fused within a smartphone application developed specifically for this purpose.

B. PARTICIPANTS

Sixty four volunteers (all males; ages: 21.95 ± 1.16 years; age range: 20 –27 years) from Saudi Arabia participated in the data collection experiment. These participants were students and staff from Umm Al-Qura University, Makkah, Saudi Arabia.

An overview of the data in terms of frequency of performing Hajj or Umrah, age distribution, height vs weight distribution, and weekly physical exercises are shown in Figure 1, Figure 2, Figure 3 and Figure 4.

The data presented in those figures provide important insights and context that are crucial for understanding the participants and their characteristics within the study. Here is a justification for the importance of each figure:

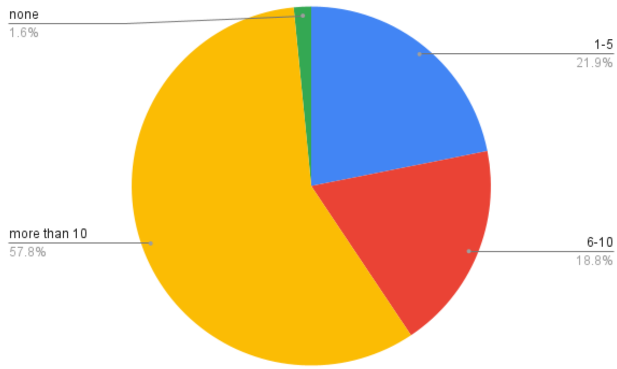

FIGURE 1. Frequency of preforming Hajj or Umrah among subjects.

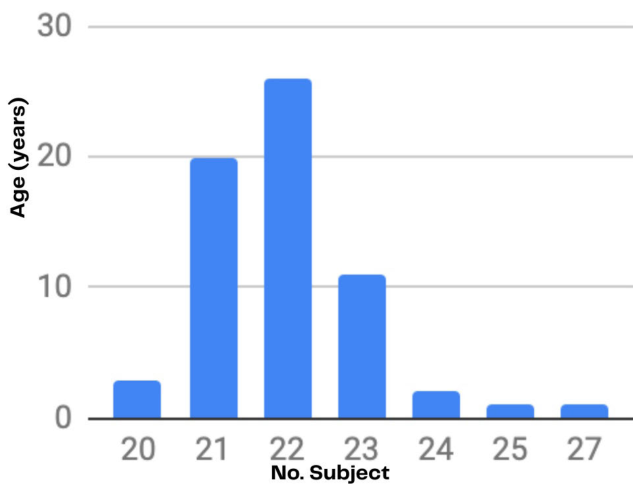

FIGURE 2. Subjects age distribution.

Figure 1: Frequency of performing Hajj or Umrah among subjects. This figure showcases the distribution of participants based on their frequency of performing Hajj or Umrah. Understanding the participants' prior experience with these religious activities is essential as it can potentially influence their familiarity with the pilgrimage environment, their level of preparedness, and their behavior during the pilgrimage. This information helps to contextualize the data analysis and interpretation, allowing researchers to consider the potential impact of prior experience on the observed outcomes.

Figure 2: Subjects' age distribution. The age distribution of the subjects provides valuable demographic information. Age can significantly influence individuals' physical capabilities, endurance, and resilience during the demanding pilgrimage activities. By examining the age distribution, researchers can identify any age-related patterns, differences, or potential correlations with other variables of interest, such as fatigue levels or emotional states. This information helps in understanding the diversity of the participant cohort and its implications for the study findings.

Figure 3: Subjects' weight vs. height distribution. The weight and height distribution of the subjects offers insights into their physical characteristics. These measurements can be associated with individuals' physical fitness levels, body

TABLE 1. Open source pedestrian mobility trace datasets.

Dataset	Research Goal	Method of data collection	Data (images, acceleration data, physiological signal....etc)	Network (Blutooth, Wifi, RFID)	No. of samples/no of subjects
Crowd-11 [35]	Develop a highly generic and fine characterisation of crowd behaviours	manually selected and extracted from the web using keywords such as transit crowd, rush time	Video-surveillance data	Internet	11 crowd motion patterns with over 6000 video .
Brightkite [25]	understand what basic laws govern human motion and dynamics.	Brightkite network public API	online location-based social networks data, user, check-in time, latitude, longitude, location id	Internet	4491143 Checkins
Gowalla [25]	understand what basic laws govern human motion and dynamics.	Gowalla network public API	online location-based social networks data, user, check-in time, latitude, longitude, location id	Internet	6442890 Checkins
The thlab/ sig-comm2009 [24]	study opportunistic mobile social network at crowded events	Information gathered through a mobile social application called MobiClique, which operates on an opportunistic basis	Records of Bluetooth interactions, chance-based messaging, and social profiles belonging to 76 individuals who utilized the MobiClique application during SIGCOMM 2009	Bluetooth	76 users
The yonsei/lifemap [23]	a smartphone-based context inference	using LifeMap(custom-built mobility monitoring system) detailed mobility data was obtained from commercially available mobile phones	latitude and longitude, Wi-Fi fingerprints, MAC address user-defined types of places	Bluetooth, Wifi	over two months in Seoul, Korea, the system gathered this data at regular intervals of 2 to 5 minutes for the purpose of ongoing location monitoring.
The upb/hyccups [22]	a smartphone-based context inference	using custom-built mobility monitoring systemdata was collect from commercial mobile phones	mobile interaction, statistics related to usage, battery, user activity, and sensor data, participants' Facebook profiles, social connections matrix , users' interests	WiFi	63 days, with 72 participants.
GeoLife GPS Trajectories [21]	dataset for mobility pattern mining, user activity recognition, location-based: social networks, privacy, and recommendation.	GPS loggers and GPS-phones	a sequence of time-stamped points containing attitude, longitude and altitude.	Internet	182 users

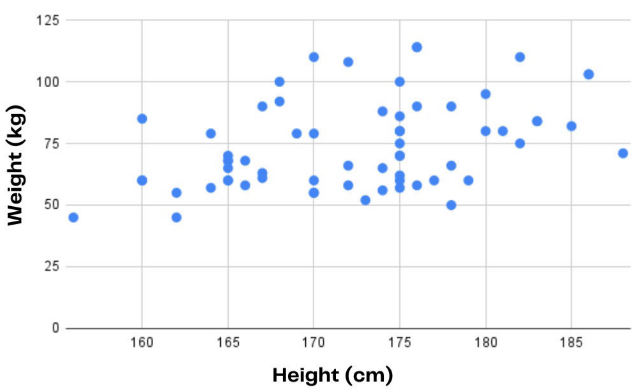

FIGURE 3. Subjects weight vs. height distribution.

mass index (BMI), and overall health status. Understanding the distribution of weight and height among participants is crucial, as it allows researchers to consider the potential impact of physical attributes on the study outcomes, such as fatigue levels or the ability to perform certain activities. This information helps in assessing the physical diversity within the participant group and its potential influence on the research findings.

Figure 4: Subjects' frequency of performing physical exercise on a weekly basis. This figure illustrates the participants' self-reported frequency of engaging in physical exercise on a weekly basis. Regular physical exercise can significantly impact individuals' fitness levels, stamina, and overall well-being, which are relevant factors in the context of a physically demanding pilgrimage. By analyzing the frequency of physical exercise among participants, researchers can explore any potential associations between exercise habits and variables of interest, such as fatigue levels or emotional states. This information helps in understanding the participants' physical preparedness for the pilgrimage activities and its potential influence on the study outcomes.

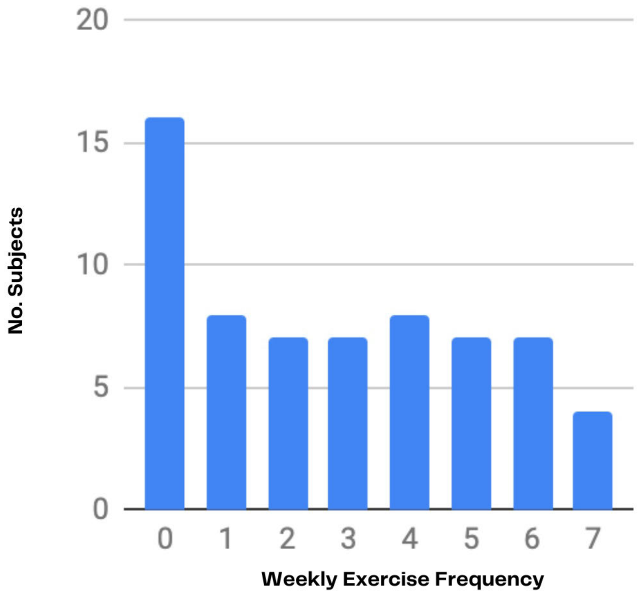

FIGURE 4. Subjects' frequency of performing physical exercise on a weekly basis.

In summary, Figures 1, 2, 3 and 4 provide essential contextual information about the participants' prior pilgrimage experience, age distribution, physical characteristics, and exercise habits. These figures help researchers consider the influence of these factors on the study findings and provide a comprehensive understanding of the participants' demographics and physical attributes within the context of the study.

C. ETHICS STATEMENT

Our study was reviewed and approved by the Biomedical Research Ethics Committee from Umm Al-Qura University Approval No. (HAPO-02-K-012-2023-03-1602).

D. MATERIALS AND SETUP

The experiments were performed in the Masjid Al-Haram in Mecca, Saudi Arabia. Location data, annotation data, subjects' demographic information and peripheral physiological signals were recorded using a custom built mobile application. For readers interested in getting more details about the collected data, the custom-built application and its development process, and sensors selection, we recommend reading our previously published work [36]. This study significantly expands upon our preliminary work [36], which primarily proposed a framework and included data from only two participants to test the feasibility of our approach. The current paper extends this by collecting a comprehensive dataset from 64 participants during the Umrah season. This dataset encompasses detailed information on activities, physiological signals, and emotional states, enabling the research community to conduct extensive statistical analyses and develop predictive models. These models are vital for understanding the interrelations among crowd demographics, physiological states, location data, and Hajj activities. Additionally, by providing an open database, this study enhances the potential for future research and applications in this field, marking a substantial expansion in data collection, analysis, and the utility of the outcomes.
The study employed two sets of data collection devices, each set comprising a Samsung Galaxy S22+ (SM-5906E/DS) smartphone with Android 13 operating system, an Empatica E4 wristband, and a Zephyr BioHarness 3.0 belt(EC38,Type 3). This configuration enabled two participants to collect data simultaneously, rather than each participant carrying two phones. This methodology was intentionally designed to optimize the efficiency of data collection and to ensure the reliability and validity of the data captured during the Hajj activities. The complete specification of Zephyr BioHarness 3.0 sensors and their values range can be found in [37]. Also, the complete technical specifications about Empatica E4 wristband sensors and values ranges can be found in [38]. The E4 watch employs an infrared thermopile sensor to accurately measure skin temperature, achieving a sensitivity of ± 0.2°C within a temperature range of 36 to 39°C. The official tourist season runs from October to March, however the best months to visit are between November and February when the weather is coldest. The experiments were conducted between 31-October-2022 to 25-January-2023. This time of the year was chosen to minimize the effect of atmospheric temperature on the subject's mood and emotion and sensors. Table 2 presents the sampling rate of the recorded data.

E. EXPERIMENT PROTOCOL

Our research protocol was approved by the Biomedical Research Ethics Committee of Umm Al-Qura University, Approval No. HAPO-02-K-012-2023-03-1602. Participants were informed that their involvement in the study was completely voluntary and that they were free to withdraw from the experiment at any point without any consequences. Before the commencement of the experiment, each participant provided their consent by signing a consent form. They were provided with a set of guidelines outlining the experimental protocol and the significance of various scales employed for self-evaluation of fatigue and emotional state. An experimenter was available to address any queries. Once the participant had a clear understanding of the instructions, the two sensors were attached to the participant, and they began operating the application, which presented demographic questions for the participant to answer (see Figure 5), also the participants performed a practice trial(unrecorded) with the help of the experimenter to familiarize themselves with the system and make sure the sensors are working. There is a setup in the app that shows if the two sensors are connected and collecting data or not. Also, there is a button to actually start collecting, recording and sending the collected data (see Figure 13 in Appendix A). After that, he/she was led to Grand Mosque in Mecca to perform real Umrah(the experiment). Next, the participant initiated the experiment by pressing the (STOPPED/COLLECTING DATA) toggle button on the app annotation UI which will start the physiological signals,

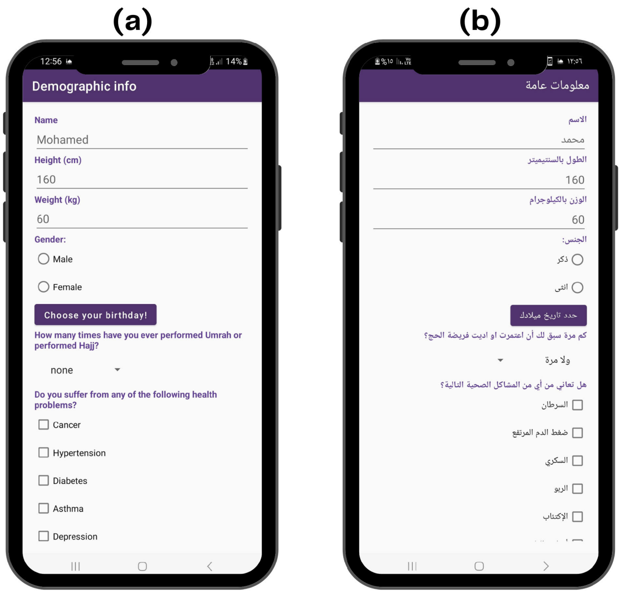

FIGURE 5. **The demographic information user interface where subject is required to fill in his/her name, height, weight, gender, date of birth, How many times they performed Hajj or Umrahh before, chronic diseases they have, nationality and how many times he/she exercise per week (a) English UI. (b) Arabic UI.**

location data, and annotation recording (see Figure 13 in Appendix A).

Each subject is asked to perform from the experiment consisting of the following steps:

1) Perform Umrah activity Tawaf for 7 rounds while annotating the data with the Tawaf round number.
2) Perform Umrah activity Saee for 7 rounds while annotating the data with the Saee round number.
3) While performing step 1 and step 2 make sure to change your fatigue level in the app UI when it change.
4) While performing step 1 and step 2 make sure to change your emotional status in the app UI when it change.
5) Save and send the data to back-end database.

Figure 6 shows a participant before the experiment. Each session last about 45 minutes.

F. DESIGN CONSIDERATIONS

Many design consideration were taken to minimize data loss and prevent app killing by the operating system. Those consideration included the following:

- The status of the connections to the two sensors is displayed in the annotation UI all the time for the participant. If any of the connection were lost or broken the participant is instructed to reconnect the two sensors;
- The app consumes huge amount of mobile battery since it uses GPS, legacy Bluetooth to connect to the Zephyer belt and Bluetooth Low Energy (BLE) to connect to the E4 watch. This cause the Android operating system (OS) to kills the app if it goes to the background. The Android OS moves apps to background if the user is not directly interacting with it. Therefore, we made the app to beep and vibrate for 1 second every 3 minutes.The sound will remind the user to interact with the app. Every time participant hears the sound he/she have to log either his/her emotional status or fatigue level, if neither have changed he/she can click a button that indicate nothing has changed. Clicking that button has no effect on the collected data, it just keeps the app in the foreground and prevent the OS from moving it to the background or killing it.
- Mobile data was used to make connection to the firebase database all the time during the experiment. Data is not saved locally in the mobile rather it is sent to the backend as soon as it is collected. This increase cost but insure no data is lost if the app gets killed by the OS at any point in time.
- Both sensors are recharged after each experiment to get optimal performance. Specially, the belt since its battery runs out faster than the E4 watch.

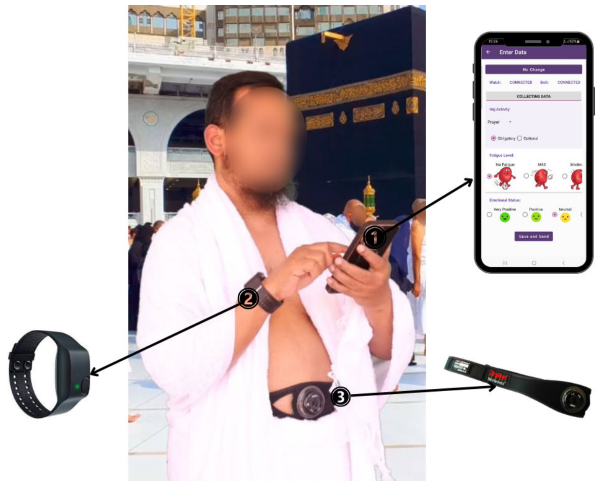

FIGURE 6. **A subject wearing the two sensors and holding the mobile before the beginning of the experiment.**

G. PARTICIPANT SELF-ASSESSMENT

During the experiment, participants performed a self assessment of their fatigue level and emotional status specifically the level of valence. Self-assessment using visual analog fatigue Scale (VAFS) [39] were used to measure participant fatigue level (see Figure 7 (a)). The fatigue scale range from No Fatigue or mild or moderate to extreme or the worst fatigue. Self-assessment smiles with descriptive language was employed to represent the emotional valence scales in a visual manner(Figure 7 (b)). The emotional valence scale ranges from very positive or positive or neutral to negative or very negative.

IV. DATA QUANTITATIVE AND STATISTICAL ANALYSIS

A. DATA CONSOLIDATION AND MISSING VALUE HANDLING

In our dataset, we are consolidating data from various types of sensors, each with its own sampling frequency as depicted in Table 2. For instance, the E4 wristband consists

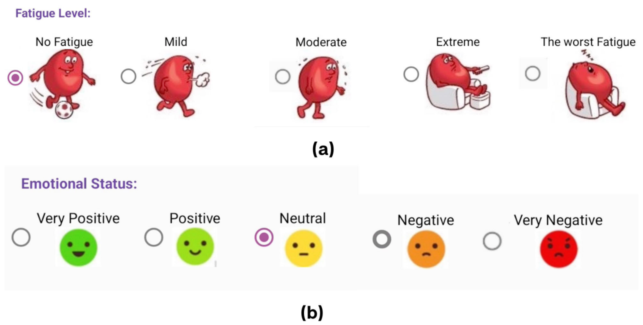

FIGURE 7. Visual scale for fatigue level and emotional status. (a) Fatigue level visual scale (b) Emotional status valence visual scale.

of four sensors, namely the Blood Volume Plus (BVP) sensor (64Hz), Temperature sensor (4Hz), Accelerometer (32Hz), and Electrodermal activity sensor (EDA) (4Hz). Meanwhile, the Zephyr Bioharness belt recorded data at 1Hz, and location data was registered at a frequency of 1Hz. Due to this discrepancy, some values are missing, prompting us to establish a protocol for filling in these gaps in the dataset. To accomplish this, we adopted the last observation carried forward (LOCF) technique, where the most recent observation replaces any missing values. However, it assumes that the missing data does not affect the outcome, which is reasonable in our current setting. To fill the initial missing values, we also used the Next Observation Carried Backward (NOCB) approach after applying LOCF to fill the missing values in-between.

TABLE 2. The type of data collected along with the sampling rate for each type.

Data Type	Data Available	Sampling Rate
subject Data	createdAt, dateOfBirth, Age, disease, firebaseUID, gender, hajVisitTimes, nationality, sportTimes, height, weight	once at the beginning of the experiment
Location Data	accuracy, altitude, bearing, createdAt, latitude, longitude, speed, time of the location	1Hz
Zephyr BioHarness	activity, breathingRateConfidence, breathingWaveAmplitude, breathingWaveNoise, ecgAmplitude, ecgNoise,heartRate, heartRateConfidence, heartRateVariability, peakAcceleration, respirationRate	1Hz
E4 wristband	accelerationTimestamp, xAcceleration, yAcceleration, zAcceleration	32Hz
	bvp, bvpTimestamp	64Hz
	gsr, gsrTimestamp	4Hz
	temp, tempTimestamp	4Hz
	battery, batteryTimestamp, createdAt, ibi, ibiTimestamp, onWristStatus, watchStatus	
Annotation Data	createdAt, emotionalMoodLevel, physicalTiredLevel,HajActivity	when it happen

B. DESCRIPTIVE STATISTICS

In this study, the original feature set comprised an extensive list of 48 variables(see Table 2), including timestamps, sensor readings, and various physiological and environmental measures. After careful consideration, we decided to exclude certain features due to redundancy or their limited relevance to our analysis. For instance, features like 'gsrTimestamp', 'timestampDay', 'timestampMilliseconds', 'timestampMonth', and 'timestampYear' were removed as they could be effectively substituted by the more comprehensive 'createdAt' timestamp. This particular timestamp played a crucial role in consolidating data from various sensors with distinct sampling frequencies, as indicated in previous section. Additionally, certain features like 'battery' and 'batteryTimestamp' were excluded from the analysis due to their limited relevance for our specific research objectives. By carefully curating the feature set, we ensured that the subsequent data analysis was focused and meaningful.

Furthermore, the term ''noise'' is used to describe two distinct types of signal interference as measured by the belt sensor: 'breathingWaveNoise' and 'ecgNoise.' 'BreathingWaveNoise' is defined as the variability in the breathing signal that arises not from the respiratory activity itself, but from ancillary sources such as sensor displacement, environmental factors, or other non-respiratory influences. 'EcgNoise' refers to similar variability in the electrocardiogram signal, which captures interference not related to cardiac function, such as muscle movements, electrical noise from nearby equipment, or issues with sensor-skin contact. These values are quantifications of the extraneous signals that are captured during the monitoring of breathing and ECG waveforms and represent the inherent variability in signal acquisition. They are not errors per se but are indicative of the multifactorial noise that can affect the quality of physiological data.

Table 3 displays the descriptive statistics of the variables collected. It includes minimum, first quartile, median, mean, third quartile, maximum, standard deviation, and mode for each variable. This table provides a comprehensive overview of the data distribution, essential for understanding the dataset's characteristics.

Overall, this table provide a detailed picture of the dataset's descriptive statistics, offering a valuable foundation for further data exploration and analysis in our research.

C. QUANTITATIVE ANALYSIS OF ACTIVITY ANNOTATIONS

This subsection presents a detailed examination of the quantity of examples generated for each subject during different Hajj or Umrah activities. The varying numbers of examples are attributed to time constraints faced by some subjects or their failure to respond to the app's prompts for annotation input during specific activities. Figure 8 showcases the quantity of records produced for each Umrah activity, with TawafOne activity, representing the first round of the 7 Tawaf rounds, having the largest number of training examples. This abundance of examples is expected to enhance the

TABLE 3. Descriptive Statistics of Variables.

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. Dev.	Mode
index	0	4033720	8067441	8067441	12101162	16134882	4657739.666	-
id	11	31	30	36	49	63	20.12279164	35
bvp	-2953.352	-23.645	0.0	-0.136	25.889	3533.993	113.0084033	0
gsr_x	0.0	0.1346	0.9573	2.5835	3.3563	37.0492	3.984031761	0
ibi	0.3125	0.4688	0.5605	0.6719	1.9688	1.7352	0.6094	-
temp	-65.29	32.85	34.89	108.25	237.11	237.11	97.15691868	237.11
x	-128.0	-48.0	-22.0	-9.27	36.0	127.0	45.32726139	-47
y	-128.0	-17.0	0.0	4.444	25.0	127.0	35.09421929	-14
z	-128.0	-37.0	-18.0	-14.73	6.0	127.0	30.8300967	-3
physicalTiredLevel	1.0	1.0	1.0	1.467	2.0	5.0	0.7599963804	1
accuracy	3.302	4.524	10.204	91.152	72.900	12907.128	428.858046	9.935
altitude	0.0	299.4	300.1	301.9	307.1	620.8	9.26943	300.1
bearing	0.0	8.931	133.170	146.765	264.866	360.0	126.688065	0
latitude	21.31	21.42	21.42	21.42	21.42	24.42	0.010607	21.42
longitude	39.50	39.83	39.83	39.83	39.83	39.95	0.01287	39.82
speed	0.0	0.06635	0.59471	0.88886	0.98585	35.01099	2.303762344	0
activity	0.01	0.27	0.46	0.4651	0.6100	7.28	0.3087296884	0.04
breathingRateConfidence	-1.0	-1.0	-1.0	-0.5143	0.0	0.0	0.4997964071	-1
breathingWaveAmplitude	-32156	280	500	821.4	900	32000	1189.508458	220
breathingWaveNoise	-32736	-239	-157	4338	7758	32612	6367.902249	-255
ecgAmplitude	0.0	0.000177	0.000777	0.002517	0.003217	0.028633	0.003864721634	0.0
ecgNoise	0.0	0.0001280	0.0005120	0.0007881	0.00064	0.017083	0.002662447421	0.000128
heartRate	0.0	90.0	109.0	103.4	123.0	213.0	33.67232332	0
heartRateConfidence	-126.0	-2.0	-2.0	-2.025	-2.0	125.0	5.019001898	-2
heartRateVariability	-5888	1792	3072	2470	3072	3327	1001.547468	3072
peakAcceleration	-327.290	-158.040	8.390	3.126	156.840	326.07	186.7972058	149.41
respirationRate	-0.1	18.0	22.0	21.94	26.0	52.0	6.045216733	23
skinTemperature	-17.9	-14.6	-0.7	-2.281	2.4	18	10.99882039	-0.7

classification accuracy for this particular activity. Additionally, it suggests that subjects tend to log their annotations more frequently at the beginning of each activity, as evidenced by SaeeOne activity, which has the second-largest number of training examples.

The count of records generated for each emotional status is presented in Figure 9. The data demonstrates that ''very positive'' emotion was the most frequently reported emotion among the subjects. Conversely, ''negative'' and ''very negative'' emotions had the fewest number of training examples, which could potentially affect the algorithm's learning of these emotional states. This outcome was foreseen by the research team during the study design phase, considering the nature of Hajj and Umrah as religious activities often associated with positive feelings, blessings, and happiness. Thus, expressing negative emotions in such a context may not be socially acceptable.

Moreover, Figure 10 presents the number of records generated for each level of physical fatigue, ''Physical Fatigue Level'' on the x-axis represents the degree of physical fatigue recorded, with each bar corresponding to the count of records for that fatigue level. The scale is as follows:

- 1.0 indicates 'No fatigue', representing a baseline or rested state.
- 2.0 signifies 'Mild fatigue', typically characterized by a slight decline in alertness and performance.
- 3.0 denotes 'Moderate fatigue', where signs of tiredness become more evident, and some decrease in physical capability may occur.
- 4.0 represents 'Considerable fatigue', associated with a substantial reduction in both mental and physical performance.
- 5.0 corresponds to 'Extreme fatigue', indicating a state where performance is significantly impaired, and rest or intervention is immediately required.

Each increment on the scale reflects an intensifying level of physical tiredness as reported by subjects during the experiments. The analysis reveals that most subjects reported feeling no physical fatigue. Possible explanations include the demographic characteristics of the participants, as all of them were male and within the age range of 20-27 years (21.95 ± 1.16 years). This finding prompts the consideration of potential variations in responses for female or elderly participants in similar situations. The subsection offers valuable insights into the quantity of data generated for different activities and emotional states, providing essential information for further analysis and modeling.

D. CORRELATION ANALYSIS

Figure 11 presents a heatmap visualizing the correlation matrix of sensor data collected during Hajj activities. Each cell represents the correlation coefficient between two variables, indicated on the x and y axes. The correlation coefficients range from -1 to +1, with positive correlations shown in orange and red shades, implying a direct relationship between variables. Conversely, negative correlations appear in blue shades, indicating an inverse relationship. The diagonal from

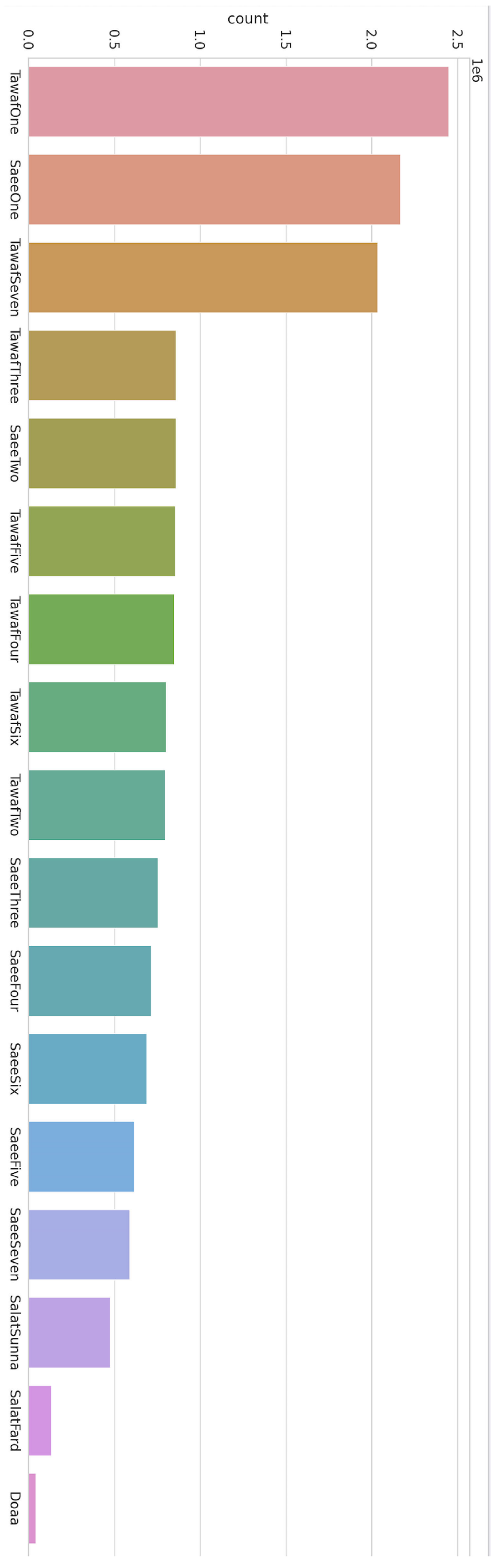

FIGURE 8. The quantity of records generated for each Hajj activity.

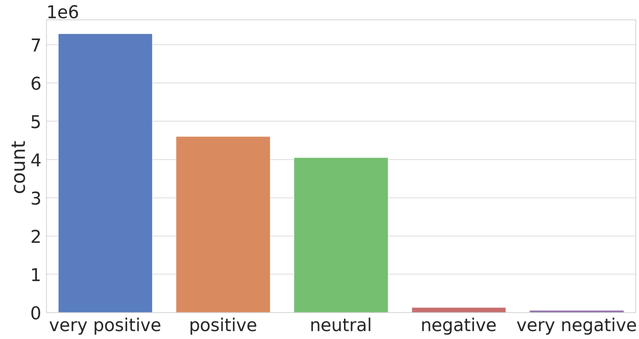

FIGURE 9. The number of records generated for each emotional status.

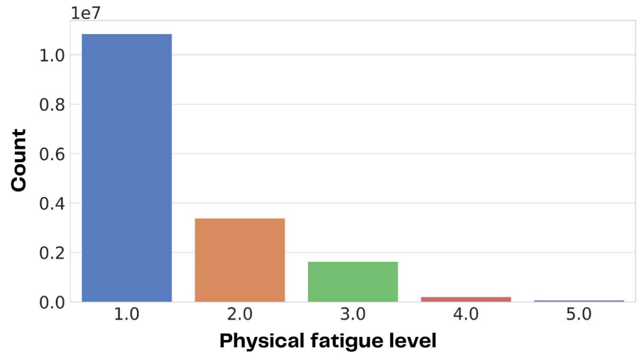

FIGURE 10. The number of records generated for each physical fatigue level.

the top-left to the bottom-right, colored in dark red, represents the perfect correlation of each variable with itself.

Analysis of this heatmap offers insights into the interrelationships among different physiological parameters and their influence on participants' emotional states, fatigue levels, and Hajj activities. Identifying these correlations is crucial for developing predictive models tailored to the Hajj experience.

Based on the matrix, it appears there are several variables with strong correlations, both positive and negative.

1) Strong Negative Correlation between Longitude and Latitude: This might indicate a geographical pattern or trend within the dataset. In the context of the dataset, perhaps the data points are from a region where changes in longitude and latitude are inversely related due to the specific routes or movements being monitored. It could be worth exploring if this pattern holds across different segments of the data or if it's influenced by specific conditions or events.
2) Moderate Negative Correlation between BreathingRateConfidence and BreathingWaveAmplitude:
   A moderate negative correlation suggests that higher confidence in the breathing rate may be associated with lower amplitude in the breathing wave. This could imply that a more stable or consistent breathing pattern results in a higher confidence value for the breathing rate estimation.
3) Strong Positive Correlation between BreathingRateConfidence and BreathingWaveNoise:
   This is somewhat counterintuitive, as one might expect higher noise levels to lead to lower confidence in the breathing rate. It's possible that in this context, 'noise' refers to a feature that, when high, actually indicates a

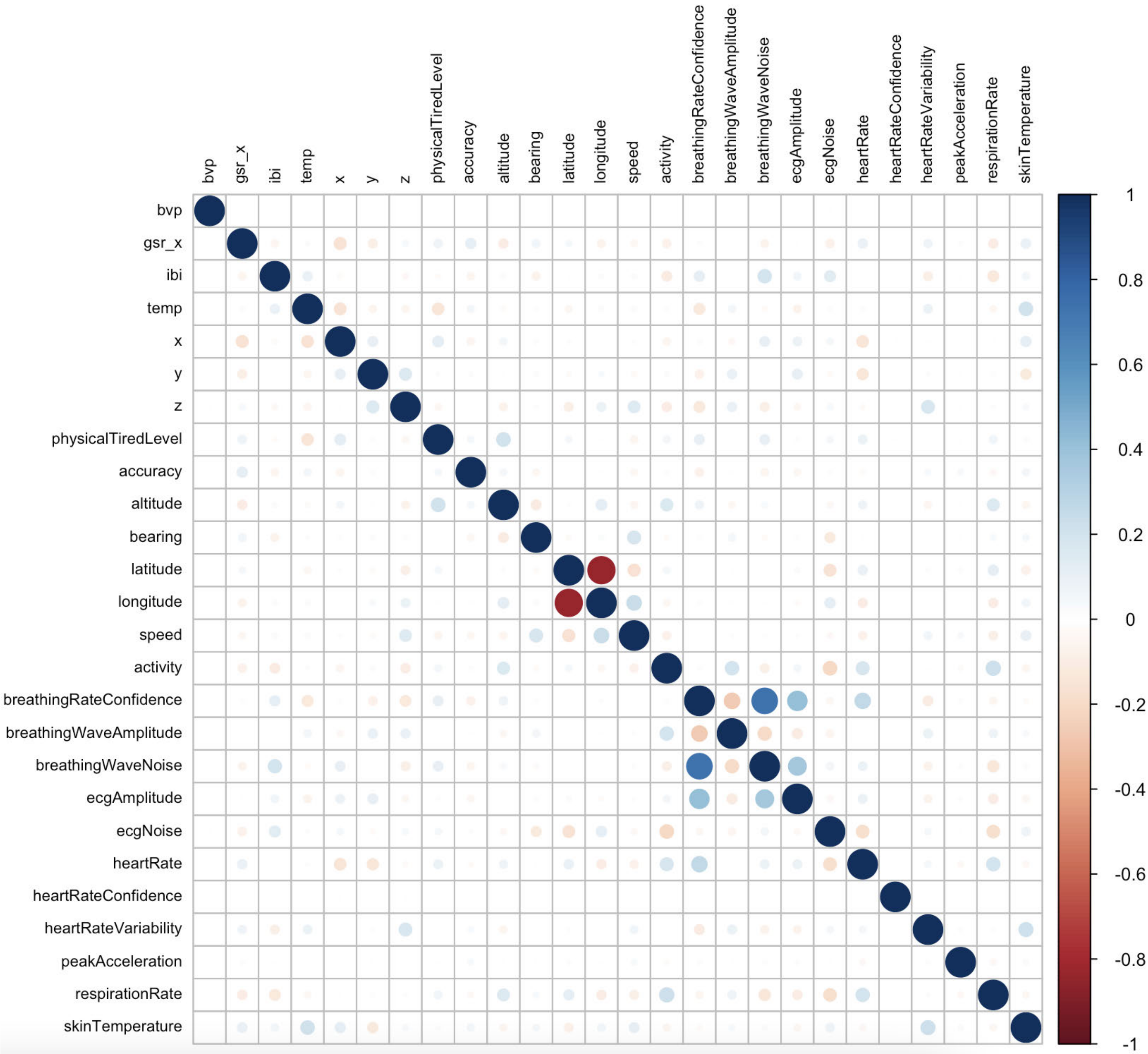

FIGURE 11. Correlation Heatmap of features.

clearer signal of the breathing pattern, thus increasing confidence in the measurement.

4) Moderate Positive Correlation between BreathingRate-Confidence and ECG Amplitude:
This indicates that as the amplitude of the ECG signal increases, so does the confidence in the breathing rate. A stronger ECG signal might be associated with clearer physiological signals overall, which could explain the increased confidence.

5) Moderate Positive Correlation between Breathing-WaveNoise and ECG Amplitude:
Similar to the previous point, this suggests that when the ECG amplitude is higher (indicating a strong heart signal), the 'noise' in the breathing wave is also higher. Again, 'noise' might not be detrimental in this context but could be related to the strength or clarity of breathing signals.

6) Weak Positive Correlation between Temp (Temperature from Wristband) and SkinTemperature (Temperature from Belt):
A weak positive correlation indicates that there is a slight tendency for these two temperature readings to move in the same direction, but the relationship is not strong. This could be due to differences in how the devices measure temperature since the belt use a formula to extract the temperature reading while the wristband has a physical sensor to measure the temperature.

7) Weak Negative Correlation between Temp (Temperature from Wristband) and PhysicalTiredLevel:
This suggests that as the wristband temperature increases, there's a slight tendency for the physical tiredness level to decrease, although the relationship is not strongly pronounced. The physiological implications might be subtle and would require further investigation to understand the causation, if any.
8) Weak Positive Correlation between Altitude and PhysicalTiredLevel:
The slight positive correlation here could indicate that higher altitudes might be associated with a higher level of physical tiredness. This makes sense as altitude can affect oxygen levels and thus fatigue levels, but the weak correlation suggests other factors are also at play.
9) Weak Positive Correlation between Activity and (BreathingWaveAmplitude, HeartRate, and RespirationRate):
These weak positive correlations are logical, as one would expect that with increased activity, there might be increases in respiratory and heart rate measures. The weakness of the correlations may indicate that the relationship is not straightforward and that individual differences or other variables also influence these measures.

V. USAGE NOTES

The dataset is saved in .csv and .json formats, while the metadata is in .xls format. It can be used for various purposes such as testing hypotheses for fatigue detection, emotional classification, and activity recognition, as well as developing real-time algorithms for detecting Hajj activities. The dataset is particularly relevant for crowd management engineers who are responsible for designing and implementing systems for managing the safety and comfort of large crowds. Additionally, scientists from different backgrounds, such as psychologists, health data scientists, and biomedical engineers interested in wearable solutions for fatigue detection, emotional classification, and activity recognition, will also find this dataset useful.

VI. EXPERIMENT AND RESULTS

A. EXPERIMENTAL SETUP AND COMPUTING INFRASTRUCTURE

The data preprocessing and experimental tasks were conducted on Google Colab Pro+, which was deemed necessary due to the limitations of the RAM memory offered by the standard and Pro versions of Colab. All programming was performed in Python 3, leveraging the capabilities of Keras (version 2.9.0) and TensorFlow (version 2.9.2) libraries.

The computational environment was equipped with an Intel(R) Xeon(R) CPU, operating at 2.20 GHz. This central processing unit featured seven cores, each with a cache size of 56320 KB. The system boasted a substantial memory capacity, with total system RAM reaching 83.5 GB.

In addition to the robust CPU capabilities, the environment was further enhanced with substantial GPU RAM of 40.0 GB, aiding in the efficient execution of deep learning computations. The system also provided a generous disk space of 166.8 GB, accommodating the storage needs for the data-intensive operations. The amalgamation of these resources facilitated the successful execution and evaluation of the deep learning models in our study.

B. FEATURE ENGINEERING

In order to format the data for deep learning, a sliding window approach was employed on the output from each sensor. Classification algorithms cannot be directly applied to raw time-series such as physiological data, accelerometer data and location data. Instead, the initial step involves converting the raw time series data into a series of consecutive, consistent windows of fixed size and fixed overlap. This was achieved by dividing the data into 10-second intervals, each comprising 640 readings from every feature. We have 640 reading because the E4 watch collects 64 reading for BVP every second. Now, since we have used two techniques to handle null/missing value resulting from the features merge namely we used:Last Observation Carried Forward (LOCF) followed by Next Observation Carried Backward (NOCB). Therefore, now we have 64 readings from each feature every second. Experiments were not conducted to determine the ideal segment duration value (window-size). Instead, a 10-second duration was chosen for each segment (window) based on a heuristic approach. We believe this duration provided sufficient time to capture patterns and repetitions involved while performing Hajj activities. The overlap duration was also choose heuristically to be 2-seconds (128 readings).

In our study, we further enhance the data's structure and quality by employing advanced feature engineering techniques, which involve creating new variables from the existing ones to uncover additional insights. We chose to apply interaction and polynomial features to our dataset, both of which have proven beneficial in multiple data analysis scenarios.

We generate interaction features through combining 'x', 'y', and 'z' accelerometer readings using polynomial features of degree 2. These readings, initially recorded separately, are combined to capture the potential combined effect of these variables on the Hajj participants' experience. The PolynomialFeatures function from sklearn.preprocessing is utilized to carry out this process [40], resulting in a new set of features representing all the polynomial combinations of 'x', 'y', and 'z' up to the specified degree.

Additionally, we handle the categorical variables in our dataset by using one-hot encoding, a common technique for transforming categorical data into a format that can be used by machine learning algorithms. In our study, we apply one-hot encoding to the 'rukun' variable, which represents the type of Hajj activity being performed. The OneHotEncoder [41] function from sklearn.preprocessing is used to create binary, or "dummy", variables for each category in the 'rukun' feature.

Following these transformations, we concatenate the original DataFrame with the newly created polynomial and

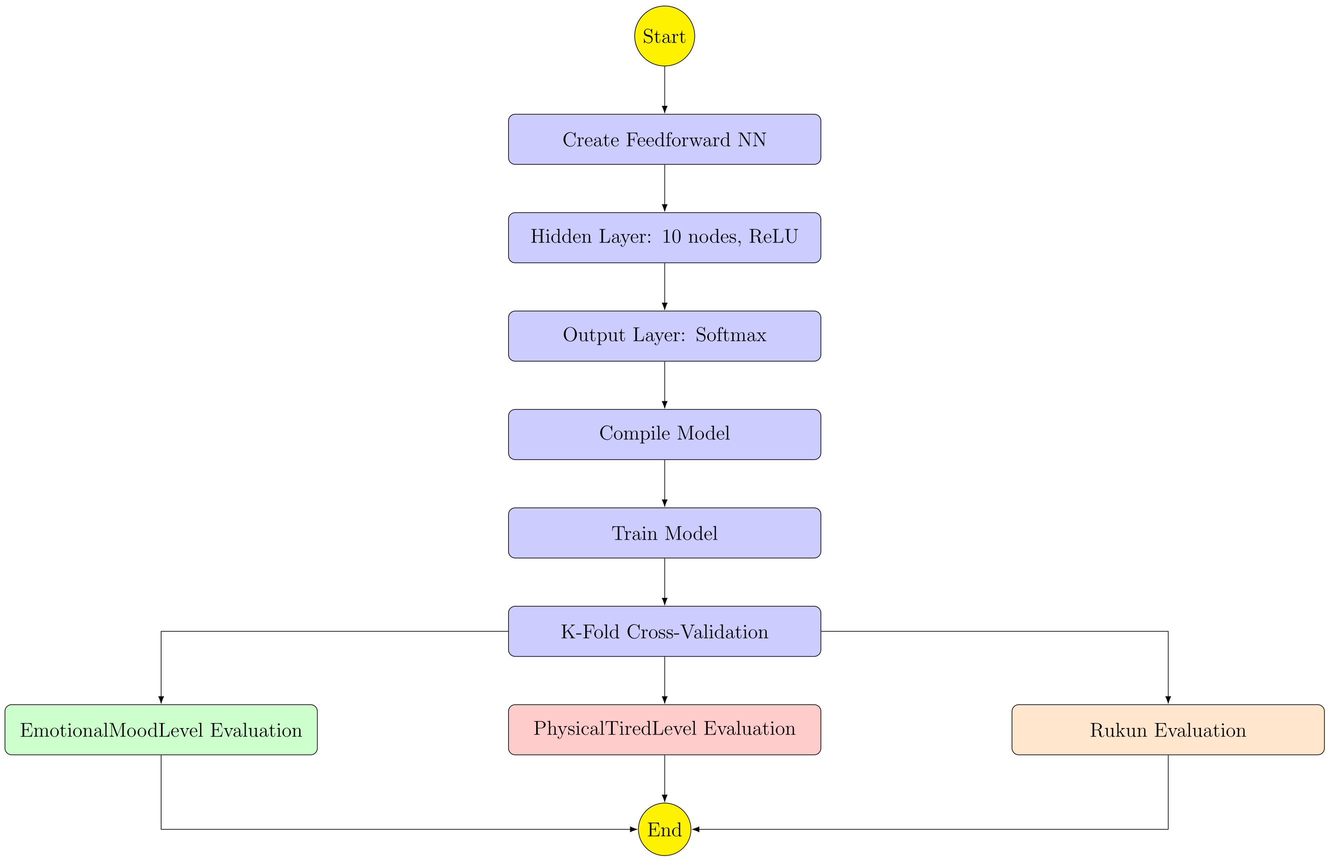

FIGURE 12. Deep learning model construction and performance evaluation.

one-hot encoded features. This results in a comprehensive DataFrame, ready for deep learning analysis. By employing these feature engineering techniques, we hope to uncover complex patterns and relationships in the data that could significantly enhance our deep learning model's performance.

The benefit of these transformations and the quality of the newly generated features will be further evaluated during the model training and testing phases. If these engineered features prove beneficial for the model performance, they will be included in the final model; otherwise, they will be excluded to avoid overfitting and unnecessary computational complexity.

C. DEEP LEARNING MODEL CONSTRUCTION AND PERFORMANCE EVALUATION

Figure 12 represent the deep learning model construction and performance evaluation process.The construction of our deep learning model involved the creation of a simple feedforward neural network with one hidden layer and an output layer. The hidden layer comprised 10 nodes and utilized the Rectified Linear Unit (ReLU) activation function, a popular choice for deep learning models due to its ability to reduce the likelihood of vanishing gradients and speed up the training process.

The output layer used the softmax activation function, which is appropriate for multi-class classification problems as it produces a probability distribution over the output classes, allowing us to interpret the outputs as the model's confidence in each possible class.

The model was compiled using categorical cross-entropy as the loss function [42], a suitable choice for multi-class classification problems as it measures the dissimilarity between the predicted and true distributions. The Adam optimizer was employed due to its efficiency and little requirement for hyperparameter tuning [43].

In the training phase, we employed a batch size of 128 and trained the model for one epoch in each fold of the cross-validation process. The 'verbose' parameter was set to 0 to limit the output during the training process, providing a cleaner output that highlights the final validation accuracy of each fold.

We performed k-fold cross-validation with stratified sampling for a more robust evaluation of our model's performance [44]. The dataset was split into ten folds, and in each iteration, the model was trained on nine folds and validated on the remaining fold.

For the target variable 'physicalTiredLevel', the model achieved an average cross-validation accuracy of 85.59% with a standard deviation of 0.78%. The model's accuracy on the holdout set was 85.27%, suggesting that the model generalizes well to unseen data.

TABLE 4. Performance comparison of our proposed model with other leading methods for emotion classification from physiological signals and fatigue detection.

Classification Task	Study	ML model	Accuracy
Fatigue Level Classification	Zhang et al. [45]	Deep Convolutional Autoencoding Memory network (CAE-M)	82.9%
		One-class SVM (OCSVM)	53.9%
		One-class single-layer LSTM	71.85%
	Our model	Feedforward neural network	85.27%
Emotional Status (valence) classification	Nakisa et al [46]	ConvNet LSTM	71.61%
	Xing et al. [47]	LSTM	81.1%
	Alhagry et al. [48]	LSTM RNN	72.06%
	Our model	Feedforward neural network	82.47%

The process was then repeated for 'emotionalMoodLevel', where the model achieved a cross-validation accuracy of 82.19% with a standard deviation of 0.78%, and a holdout set accuracy of 82.47%.

Finally, for the 'rukun' target, despite a warning about the least populated class having fewer members than the number of splits, the model achieved a cross-validation accuracy of 41.71% with a standard deviation of 0.78%, and a holdout set accuracy of 42.03%. The relatively lower performance on 'rukun' might be due to the imbalanced distribution of the classes.

Overall, our model showed acceptable performance on all the target variables, albeit with varying degrees of accuracy. The model's performance on 'physicalTiredLevel' and 'emotionalMoodLevel' was particularly noteworthy, indicating its potential for reliable predictions in these areas. However, the performance on 'rukun' suggests that there may be room for improvement, potentially through techniques for handling class imbalance or gathering more data for under-represented classes.

To the best of our knowledge, no previous research has investigated the application of physiological sensors for the task of classifying Hajj ritual activities. In forthcoming work, we plan to collect more data using this framework to construct a comprehensive model capable of recognizing and classifying Hajj rituals and activities. Despite the novelty of our approach, we present Table 4 as a way to contextualize our results in relation to prior studies focused on fatigue and emotional recognition utilizing physiological signals.

This table provides a comparison between the performance of our model and other state-of-the-art methodologies for emotion classification based on physiological signals and fatigue detection.

VII. DISCUSSION

In this study, we applied a range of advanced statistical methods and analytical techniques to the data collected, with the aim of gaining a deeper understanding of the dynamics of the Hajj season. Although the results might not display statistical significance in the conventional sense, they do provide a wealth of insights into the interplay between Hajj activities and the physical fatigue and emotional states of the participants. For example, the inclusion of latitude and longitude is essential for accurately determining the participants' locations without manual annotations. These coordinates allow us to distinguish whether a participant is engaged in Tawaf or Saai, activities specific to different areas of the Hajj. This automated geographical tagging significantly enhances the precision of activity classification in our dataset.

It is important to recognize that certain aspects of our research, such as participants' emotional states and levels of fatigue, inherently require subjective input to capture personal experiences accurately during the Hajj. These subjective annotations are crucial for the detailed insights they provide into individual participant experiences. However, we also collected a substantial amount of objective data through devices like the Samsung Galaxy S22+ smartphones, Empatica E4 wristbands, and Zephyr BioHarness belts. While our current predictive models are primarily based on these subjective annotations, the objective data we've gathered offers substantial potential for future research avenues. For instance, we could develop unsupervised models for fatigue and emotion detection relying solely on physiological data collected, which would enhance the objectivity of our findings. Additionally, using only the location data from the mobile devices and timestamps, we can infer critical aspects such as which specific ritual (Rukun) a participant is performing. This capability underscores the valuable and multifaceted nature of the dataset we have compiled, supporting both our current analyses and future explorations that could further minimize subjective biases and maximize the use of objective data.

Participants were engaged in actual Umrah rituals, not a simulation, which typically lasts about 45 minutes. This duration reflects the genuine time it takes to perform Umrah, ensuring that our study mirrors real-world conditions. We observed that participants initially logged data more frequently, with the frequency diminishing towards the end of the ritual. To counteract this, we introduced auditory alarms as reminders to ensure consistent data logging throughout the Umrah process. Despite these measures, we recognize the need for further enhancements to sustain participant engagement for the duration of the ritual. We are exploring additional methods such as more frequent reminders or an enhanced interactive interface for the data logging app to improve consistency and reliability of the data collection. These adjustments will help ensure that our dataset comprehensively captures the entire Umrah experience.

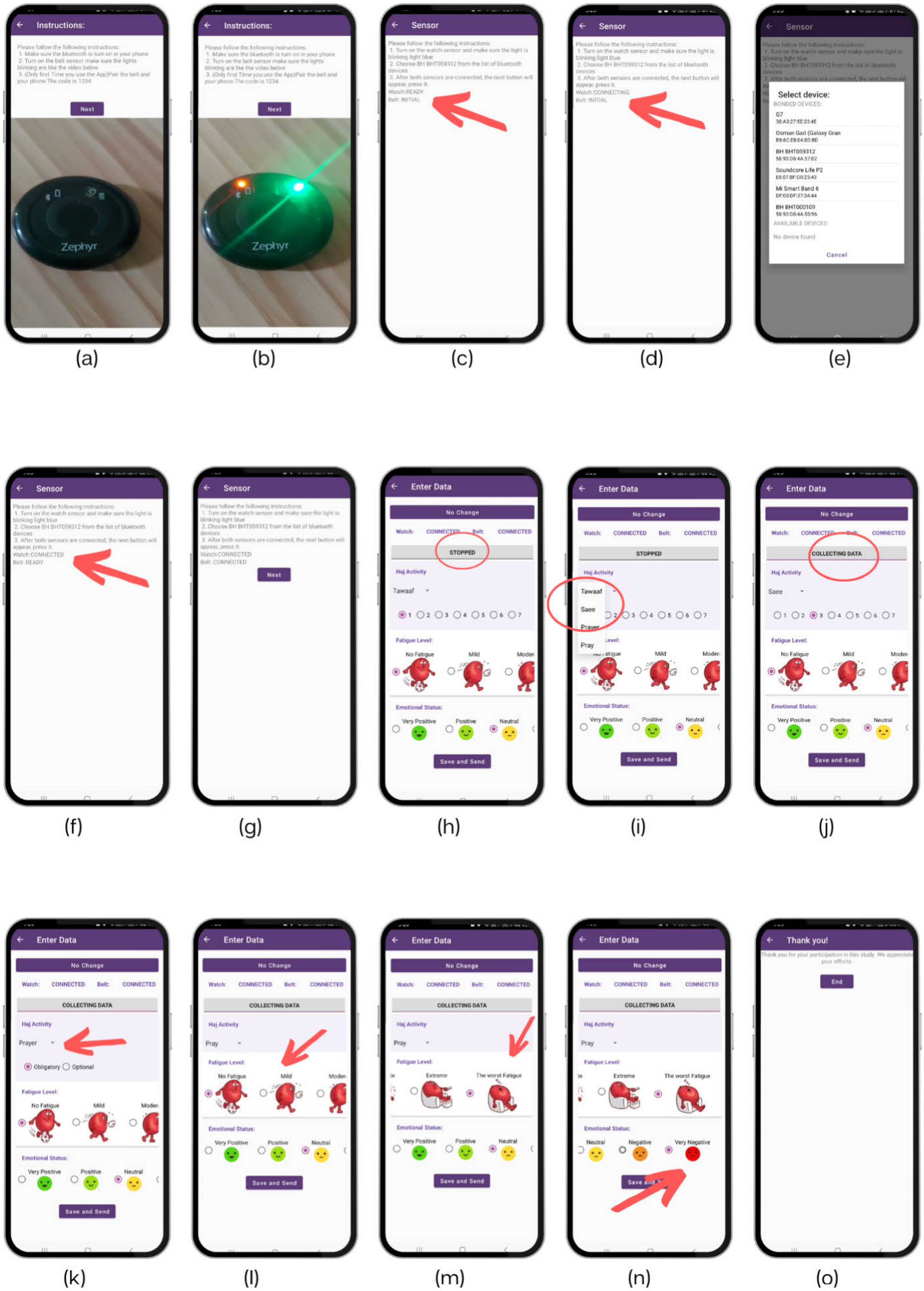

FIGURE 13. The custom made Android application different user interfaces is presented here. For example (a) Instruction UI. (c) Sensors UI showing watch and belt bluetooth connection status.

We acknowledge that the sample size of our study is relatively small when juxtaposed with the vast scale of the Hajj phenomenon. A larger sample size would undoubtedly provide greater generalizability to our findings, as well as lead to more robust conclusions. However, the unique context of the Hajj season, with its specific constraints and challenges, especially in terms of data collection and participant recruitment, makes a larger sample size challenging to achieve.

In spite of these limitations, we made every effort to ensure a diverse array of experiences within our participant group, with individuals engaged in various Hajj activities such as Tawaf, Saai, prayer, and Doaa. Although our findings might be specific to the sample group, they nonetheless provide an initial exploration into the complex relationship between crowd management and activity prediction within the Hajj context.

Moreover, the study employed tools specifically designed for research purposes, with the developers rigorously selecting and validating the metrics used in the study. These metrics were chosen based on their established relevance in the field and their alignment with our research objectives. Even with a smaller sample size, our study yields insightful results, contributing valuable knowledge to the field and paving the way for future research.

Additionally, the significance of our research extends beyond the statistical findings. The study highlights the potential of deep learning algorithms in managing crowds during large-scale events, such as Hajj. By using physiological signals to predict fatigue levels, emotional states, and types of activity, our research introduces an innovative approach to crowd management, offering a promising solution to existing challenges.

In conclusion, while our study may have its limitations, its value lies in the unique insights it provides into the dynamics of crowd management during Hajj, as well as the innovative approach it introduces in using deep learning algorithms for crowd management. We believe that our work will inspire further research in this field, ultimately leading to more efficient and effective crowd management strategies during large-scale events.

VIII. CONCLUSION

This research endeavor has led to the successful collection of a comprehensive dataset, capturing a wide array of activities, physiological signals, and emotional annotations of participants during Hajj and Umrah. As a rich source of information on crowd dynamics and behavior in the unique context of pilgrimage, this dataset is not only valuable for the insights it already provides but also for its vast potential for future applications.

The dataset lends itself naturally to the development and evaluation of predictive models aimed at optimizing crowd management during pilgrimage events. The wealth of information captured, including activity patterns, physiological signals, and emotional annotations, presents an exciting opportunity for researchers to apply advanced machine learning and statistical techniques to anticipate crowd behaviors, identify potential congestion points, and enhance crowd management strategies.

Furthermore, the dataset serves as a robust benchmark for the assessment of existing crowd management algorithms and methodologies. It offers researchers a means to test their proposed techniques against real-world data, thus promoting the evolution of crowd management systems that are specifically designed for pilgrimage contexts.

Our dataset also facilitates further exploration of the intricate relationships between pilgrimage activities, participant well-being, and crowd dynamics. By examining correlations between specific activities, fatigue levels, and emotional states, researchers can glean insights into the factors influencing crowd behavior and the impact of various activities on the overall pilgrimage experience.

We stress that our study's main goal was not to use the data for prediction purposes. Instead, we aimed to provide the research community with a unique dataset to dissect and characterize Hajj and Umrah crowd behaviors. While we have illustrated one potential use-case of the dataset in our study, we eagerly encourage the research community to explore the full depth and breadth of its potential.

In making this dataset publicly available, we hope to stimulate collaborations and catalyze further investigations into crowd analysis, management, and optimization during Hajj and Umrah. We are confident that sharing this dataset contributes to pushing the frontiers of the field forward, spurring researchers to unearth new patterns and insights that can significantly enhance crowd management practices during large-scale events like Hajj and Umrah.

APPENDIX A
EXTRA FIGURES

See Figure 13.

REFERENCES

[1] F. L. Sánchez, I. Hupont, S. Tabik, and F. Herrera, "Revisiting crowd behaviour analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects," *Inf. Fusion*, vol. 64, pp. 318–335, Dec. 2020.

[2] N. Bellomo, D. Clarke, L. Gibelli, P. Townsend, and B. J. Vreugdenhil, "Human behaviours in evacuation crowd dynamics: From modelling to," *Phys. life Rev.*, vol. 18, pp. 1–21, Sep. 2016.

[3] A. Gorrini, S. Bandini, M. Sarvi, C. Dias, and N. Shiwakoti, "An empirical study of crowd and pedestrian dynamics: The impact of different angle paths and grouping," *Transp. Res. Rec.*, vol. 41, p. 42, Aug. 2013.

[4] S. Bandini, A. Gorrini, L. Manenti, and G. Vizzari, "Crowd and pedestrian dynamics: Empirical investigation and simulation," *Proc. Measuring Behav.*, vol. 8, pp. 308–311, Aug. 2012.

[5] D. Helbing and A. Johansson, "Pedestrian, crowd, and evacuation dynamics," 2013, *arXiv:1309.1609*.

[6] X. Ding, F. He, Z. Lin, Y. Wang, H. Guo, and Y. Huang, "Crowd density estimation using fusion of multi-layer features," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 8, pp. 4776–4787, Aug. 2021.

[7] B. Krausz and C. Bauckhage, "Analyzing pedestrian behavior in crowds for automatic detection of congestions," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Nov. 2011, pp. 144–149.

[8] H. Rabiee, J. Haddadnia, and H. Mousavi, "Crowd behavior representation: An attribute-based approach," *SpringerPlus*, vol. 5, no. 1, pp. 1–17, Dec. 2016.

[9] T. H. Noor, "Behavior analysis-based IoT services for crowd management," *Comput. J.*, vol. 66, no. 9, pp. 2208–2219, Sep. 2023.
[10] P. Barsocchi, A. Crivello, D. La Rosa, and F. Palumbo, "A multisource and multivariate dataset for indoor localization methods based on WLAN and geo-magnetic field fingerprinting," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Oct. 2016, pp. 1–8.
[11] V. X. Gong, W. Daamen, A. Bozzon, and S. P. Hoogendoorn, "Crowd characterization for crowd management using social media data in city events," *Travel Behaviour Soc.*, vol. 20, pp. 192–212, Jul. 2020.
[12] S. Chessa, M. Girolami, L. Foschini, R. Ianniello, A. Corradi, and P. Bellavista, "Mobile crowd sensing management with the participact living lab," *Pervas. Mobile Comput.*, vol. 38, pp. 200–214, Jul. 2017.
[13] C. A. S. Pouw, F. Toschi, F. van Schadewijk, and A. Corbetta, "Monitoring physical distancing for crowd management: Real-time trajectory and group analysis," *PLoS ONE*, vol. 15, no. 10, Oct. 2020, Art. no. e0240963.
[14] V. F. S. Mota, T. H. Silva, D. F. Macedo, Y. Ghamri-Doudane, and J. M. S. Nogueira, "Towards scalable mobile crowdsensing through device-to-device communication," *J. Netw. Comput. Appl.*, vol. 122, pp. 99–106, Nov. 2018.
[15] T. N. Nguyen and S. Zeadally, "Mobile crowd-sensing applications: Data redundancies, challenges, and solutions," *ACM Trans. Internet Technol.*, vol. 22, no. 2, pp. 1–15, May 2022.
[16] S. Mukhopadhyay King, F. Nawab, and K. Obraczka, "A survey of open source user activity traces with applications to user mobility characterization and modeling," 2021, *arXiv:2110.06382*.
[17] L. Durán-Polanco and M. Siller, "Crowd management COVID-19," *Annu. Rev. Control*, vol. 52, pp. 465–478, Aug. 2021.
[18] S. Yadav, P. Gulia, N. S. Gill, and J. M. Chatterjee, "A real-time crowd monitoring and management system for social distance classification and healthcare using deep learning," *J. Healthcare Eng.*, vol. 2022, pp. 1–11, Apr. 2022.
[19] H. Zhao, T. Thrash, M. Kapadia, K. Wolff, C. Hölscher, D. Helbing, and V. R. Schinazi, "Assessing crowd management strategies for the 2010 love parade disaster using computer simulations and virtual reality," *J. Roy. Soc. Interface*, vol. 17, no. 167, Jun. 2020, Art. no. 20200116.
[20] J. R. Santana, L. Sánchez, P. Sotres, J. Lanza, T. Llorente, and L. Muñoz, "A privacy-aware crowd management system for smart cities and smart buildings," *IEEE Access*, vol. 8, pp. 135394–135405, 2020.
[21] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, "Understanding mobility based on GPS data," in *Proc. 10th Int. Conf. Ubiquitous Comput.*, Sep. 2008, pp. 312–321.
[22] R. I. C. and C. Dobre. (2016). *CRAWDAD Dataset UPB/Hyccups*. [Online]. Available: https://crawdad.org/upb/hyccups/20161017
[23] Y. Chon, E. Talipov, H. Shin, and H. Cha. (2012). *CRAWDAD Dataset Yonsei/Lifemap*. [Online]. Available: https://crawdad.org/yonsei/lifemap/20120103
[24] A.-K. Pietilainen and C. Diot. (2012). *CRAWDAD Dataset Thlab/Sigcomm2009*. [Online]. Available: https://crawdad.org/thlab/sigcomm2009/20120715
[25] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2011, pp. 1082–1090.
[26] A. Capponi, C. Fiandrino, B. Kantarci, L. Foschini, D. Kliazovich, and P. Bouvry, "A survey on mobile crowdsensing systems: Challenges, solutions, and opportunities," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2419–2465, 3rd Quart., 2019.
[27] G. Cardone, A. Corradi, L. Foschini, and R. Ianniello, "ParticipAct: A large-scale crowdsensing platform," *IEEE Trans. Emerg. Topics Comput.*, vol. 4, no. 1, pp. 21–32, Jan. 2016.
[28] R. Rouvoy, "Apisense: Mobile crowd-sensing made easy!" 2016. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01332594
[29] J. G. P. Rodrigues, A. Aguiar, and J. Barros, "SenseMyCity: Crowdsourcing an urban sensor," 2014, *arXiv:1412.2070*.
[30] P. P. Jayaraman, C. Perera, D. Georgakopoulos, and A. Zaslavsky, "Efficient opportunistic sensing using mobile collaborative platform MOSDEN," in *Proc. 9th IEEE Int. Conf. Collaborative Comput., Netw., Appl. Worksharing*, Oct. 2013, pp. 77–86.
[31] F. Kalim, J. P. Jeong, and M. U. Ilyas, "CRATER: A crowd sensing application to estimate road conditions," *IEEE Access*, vol. 4, pp. 8317–8326, 2016.
[32] M.-R. Ra, B. Liu, T. F. La Porta, and R. Govindan, "Medusa: A programming framework for crowd-sensing applications," in *Proc. 10th Int. Conf. Mobile Syst. Appl. Services*, Jun. 2012, pp. 337–350.
[33] T. Das, P. Mohan, V. N. Padmanabhan, R. Ramjee, and A. Sharma, "PRISM: Platform for remote sensing using smartphones," in *Proc. 8th Int. Conf. Mobile Syst., Appl., Services*, Jun. 2010, pp. 63–76.
[34] I. Carreras, D. Miorandi, A. Tamilin, E. R. Ssebaggala, and N. Conci, "Matador: Mobile task detector for context-aware crowd-sensing campaigns," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, Mar. 2013, pp. 212–217.
[35] C. Dupont, L. Tobías, and B. Luvison, "Crowd-11: A dataset for fine grained crowd behaviour analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 2184–2191.
[36] A. M. Al-Shaery, H. Aljassmi, S. G. Ahmed, N. S. Farooqi, A. N. Al-Hawsawi, M. Moussa, A. Tridane, and M. D. Alam, "Real-Time pilgrims management using wearable physiological sensors, mobile technology and artificial intelligence," *IEEE Access*, vol. 10, pp. 120891–120900, 2022.
[37] Zephyer-Technology. (2012). *BioHarness 3.0 User Manual*. Accessed: Nov. 21, 2022. [Online]. Available: https://www.zephyranywhere.com/media/download/bioharness3-user-manual.pdf
[38] Empatica. (2020). *E4 Wristband Technical Specifications*. Accessed: Nov. 21, 2022. [Online]. Available: https://support.empatica.com/hc/en-us/articles/202581999
[39] Z. Jiang, "Role of methylphenidate in the treatment of fatigue in advanced pancreatic cancer population," *Ann. Gastroenterology*, p. 536, 2016.
[40] Scikit-learn developers. (2023). *PolynomialFeatures—Scikit-learn 1.4.1 Documentation*. Accessed: Feb. 24, 2024. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.PolynomialFeatures.html
[41] Scikit-learn developers. (2023). *OneHotEncoder—Scikit-Learn 1.4.1 Documentation*. Accessed: Feb. 2, 2024. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html
[42] *Tf.Keras.Losses.CategoricalCrossentropy|TensorFlow V2.15.0.Post1*. Accessed: Feb. 2, 2024. [Online]. Available: https://www.tensorflow.org/versions/r2.15/api_docs/python/tf/keras/losses/CategoricalCrossentropy
[43] *Optimizers—Keras*. Accessed: Feb. 24, 2024. [Online]. Available: https://keras.io/api/optimizers/adam/
[44] *Scikit-Learn 1.4.1 Documentation*. Accessed: Feb. 24, 2024. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html
[45] Y. Zhang, Y. Chen, and Z. Pan, "A deep temporal model for mental fatigue detection," in *Proc. IEEE Int. Conf. Syst. Man, Cybern. (SMC)*, Oct. 2018, pp. 1879–1884.
[46] B. Nakisa, M. N. Rastgoo, A. Rakotonirainy, F. Maire, and V. Chandran, "Automatic emotion recognition using temporal multimodal deep learning," *IEEE Access*, vol. 8, pp. 225463–225474, 2020.
[47] X. Xing, Z. Li, T. Xu, L. Shu, B. Hu, and X. Xu, "SAE+LSTM: A new framework for emotion recognition from multi-channel EEG," *Frontiers Neurorobotics*, vol. 13, p. 37, Jun. 2019.
[48] S. Alhagry, A. Aly, and R. A., "Emotion recognition based on EEG using LSTM recurrent neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 10, pp. 1–18, 2017.

ALI M. AL-SHAERY (Member, IEEE) received the B.Sc. degree in civil engineering from Umm Al-Qura University, Makkah, Saudi Arabia, in 2003, the M.Sc. degree in surveying from University College London, U.K., in 2006, and the Ph.D. degree in geospatial engineering from the University of New South Wales, Australia, in 2013. He was the Dean of the Institute of Consulting Research and Studies and the Deputy Director of the Transportation and Crowd Management Center of Research Excellence (TCMCORE). He is currently an Associate Professor with the Department of Civil Engineering, Umm Al-Qura University, the Vice-President for Innovation and Entrepreneurship (VPIE), and the Vice-President for Development and Community Service (VPDCS). His research interests include global navigation satellite systems (GNSS) real-time kinematic positioning RTK, geographic information systems, location-based applications, smart and sustainable cities, and crowd management. He is the Chair of the Governing Board of the Center of Safety, Risk, and Crisis Management.

SOHA G. AHMED received the B.Sc. degree in computer science (First Class Honors with a GPA of 3.98 out of 4) from Abu Dhabi University, UAE, in 2008, and the M.Sc. degree (Hons.) in computer engineering from the American University of Sharjah, UAE, in 2014. She is currently pursuing the dual-Ph.D. degree with UAE University (UAEU), UAE, and KU Leuven, Belgium. Throughout her bachelor's and graduate studies, she received merit scholarships from both Abu Dhabi University and the American University of Sharjah. In addition to her academic pursuits, Soha has played a pivotal role in the tech industry as the software development lead for various startup projects, including Fleet Telematics, a car rental marketplace, and an educational application. Her professional experience also extends to several research centers and universities across the UAE. Her research interests are diverse and interdisciplinary, spanning the fields of natural language processing (NLP), text mining, computational intelligence, digital signal processing, human-computer interaction, and biomedical sensors and systems. Her work in these areas is driven by a commitment to innovation and excellence in both academic and practical applications. She is a recipient of the UAEU Full Package Fellowship.

HAMAD ALJASSMI received the Ph.D. degree from the School of Civil and Environmental Engineering, University of New South Wales (UNSW). He is currently an Associate Professor of Civil Engineering and the Director of Emirates Center for Mobility Research at UAE University. He published more than 80 peer-reviewed articles, most which appeared in top ranked journals. He has more than 15 years industry experience in which he acted as Project Manager, Planner or Researcher at numerous high-profile consultancy projects related to life-cycle assessment of surface transport projects, mobility policy making, transportation planning and traffic safety. His research interests include expert systems and machine learning for the buildings and infrastructure sectors.

ABDULLAH N. AL-HAWSAWI received the B.Sc. and M.Sc. degrees in computer science from Oklahoma State University, Stillwater, OK, USA, in 2008 and 2010, respectively, and the Ph.D. degree in crowd modeling and simulation from The University of Melbourne, Australia, in 2021. He is currently an Assistant Professor with the Department of Information and Scientific Services, Custodian of the Two Holy Mosques Institute for Hajj and Umrah Research, Umm Al-Qura University, the Director of the SAFE Center for Safety, Risks and Crisis Management, and the Head of the Crowd Unit, Custodian of the Two Holy Mosques Institute for Hajj and Umrah Research. He was the Deputy Director of the Transportation and Crowd Management Center of Research Excellence (TCMCORE). His research interests include crowd dynamics and interactions with the physical environment, crowd modeling and simulation, AI and its applications in crowd control, VR, and AR. He has published several works about these. He is one of the winners of Makkah Award from His Royal Highness Prince Khalid Al Faisal, for The Makkah Digital Gateway for Research and Innovation Reward, in 2022.

NADIA MAKSOUD received the B.S. degree in civil engineering from Abu Dhabi University, Abu Dhabi, United Arab Emirates, in 2019, and the M.S. degree in civil engineering from United Arab Emirates University, Al Ain, United Arab Emirates, in 2023. Previously, she was a Research Assistant at the Emirates Centre for Mobility Research, UAEU. Her research interests include urban planning, transportation engineering, and data analysis.

ABDESSAMAD TRIDANE received the Ph.D. degree in control systems theory. He is a Full Professor with the Department of Mathematical Sciences, United Arab Emirates University, Al Ain, UAE. He has authored many papers in the areas of control systems, mathematical epidemiology, and mathematical medicine. His main research interests include dynamical systems, control systems, mathematical epidemiology, and mathematical medicine.

NORAH S. FAROOQI (Member, IEEE) received the B.Sc. degree (Hons.) in computer science from Umm Al-Qura University, Saudi Arabia, in 2007, and the M.Sc. degree (Hons.) in information systems and the Ph.D. degree in computer science from The University of Sheffield, U.K., in 2010 and 2013, respectively. She was the Vice Dean of the Institute of Consulting Research and Studies and the Deputy Director of the Engineering Science Research Center. She is currently an Associate Professor of computer science with the Information Science Department, College of Computer and Information Systems, Umm Al-Qura University. She is also the General Supervisor of the Data Management Office. She had many publications related to databases and security. Her research interests include systems development, information systems, crowd management, and the IoT. She is a reviewer in a number of academic journals related to computer topics.