IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 April 2024, accepted 18 June 2024, date of publication 28 June 2024, date of current version 17 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3420752

RESEARCH ARTICLE

Advanced IoT-Based Human Activity Recognition and Localization Using Deep Polynomial Neural Network

DANYAL KHAN[1], ABDULLAH ALSHAHRANI[2], ABRAR ALMJALLY[3], NAIF AL MUDAWI[4], ASAAD ALGARNI[5], KHALED ALNOWAISER[6], AND AHMAD JALAL[1]

[1]Department of Computer Science, Air University, Islamabad 44000, Pakistan
[2]Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Jeddah 21493, Saudi Arabia
[3]Department of Information Technology, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 13318, Saudi Arabia
[4]Department of Computer Science, College of Computer Science and Information System, Najran University, Najran 55461, Saudi Arabia
[5]Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Rafha 91911, Saudi Arabia
[6]Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia

Corresponding authors: Abrar Almjally (aamjally@imamu.edu.sa) and Ahmad Jalal (ahmadjalal@mail.au.edu.pk)

ABSTRACT Advancements in smartphone sensor technologies have significantly enriched the field of human activity recognition, facilitating a wide array of applications from health monitoring to personal navigation. This study utilized such advancements to explore human locomotion and localization recognition using data from accelerometers, microphones, gyroscopes, magnetometers, and GPS, applying Deep Polynomial Neural Networks (DPNN) and Multilayer Perceptron (MLP) across three datasets: the Continuous In-The-Wild Smart Watch Activity Dataset, the Huawei Locomotion Dataset, and the Extra Sensory Dataset. We employ two distinct approaches for activity recognition: Deep Polynomial Neural Networks (DPNN) for deep learning-based feature extraction and Multilayer Perceptron (MLP) with manual feature extraction techniques, including Linear Predictive Coding Cepstral Coefficients (LPCC), step length, signal magnitude area, spectral, and sound features. Through rigorous experimentation, we achieved remarkable accuracy in recognizing both locomotion and localization activities, with DPNN consistently outperforming MLP in terms of accuracy. Specifically, for the Continuous In-The-Wild Dataset, DPNN achieved a 93% accuracy rate for localization activities and 95% for locomotion activities, while MLP recorded 86% and 91% in the respective categories. Similarly, on the Huawei Locomotion Dataset, DPNN attained 95% accuracy for localization and 97% for locomotion, with MLP achieving 88% and 91%, respectively. Furthermore, the application of these models to the Extra Sensory Dataset yielded 92% accuracy for both localization and locomotion activities with DPNN, and 90% and 89% with MLP. In our study, we observed that in terms of accuracy, DPNN emerges as the clear winner; however, it is computationally expensive. Conversely, MLP, while being less accurate, stands out for its computational efficiency. This study not only highlights the effectiveness of incorporating advanced machine learning techniques in interpreting sensor data but also emphasizes the trade-offs between computational demands and accuracy in the domain of human activity recognition. Through our comprehensive analysis, we contribute valuable insights into the potential of smartphone sensors in enhancing activity recognition systems, paving the way for future innovations in mobile sensing technology.

INDEX TERMS Human activity recognition, deep polynomial neural network (DPNN), multi-layer perceptron (MLP), locomotion, localization.

I. INTRODUCTION

Advancements in smartphone sensor technologies have significantly enriched the field of human activity recognition

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

(HAR) [1], [3], facilitating a wide array of applications from health monitoring [4], [5], to personal navigation [6], [9] and security [10], [11]. The study in [12] created a new way to recognize what activities a person is doing based on the data from their smartphone, even if the phone isn't being carried on the person. This method, called Context-Aware Human Activity Recognition (CA-HAR), looks at various kinds of information the phone's sensors collect and combines this with two main techniques: one that uses a set of rules called ripple-down rules (RDR) and another that involves deep learning, a type of advanced computer algorithm. To tackle the problem of accurately recognizing activities when the phone's location on the body changes or when it's not on the body at all, they made special RDR rules. These rules not only consider the activity itself but also other related information that could affect the recognition process. The paper highlights several key strengths of its approach to recognizing human activities using smartphone sensors, considering the context in which these activities occur. First, it brings together data from various sensors found in a smartphone, like the accelerometer and GPS, to get a fuller picture of the user's actions. This aggregation means the system isn't relying on just one type of data but combines many to improve accuracy. Secondly, the paper uses a technique called ripple-down rules (RDR). This method helps the system better categorize activities by applying a set of rules that can adapt based on the context, ensuring that activities are identified more accurately. By incorporating context, such as whether the user is indoors or outdoors, the system can make more informed guesses about what the user is doing. This approach aims to enhance the system's ability to correctly recognize activities by considering the situation or environment in which they occur. The paper points out a significant limitation in its approach to recognizing human activities through smartphones, which revolves around real-time performance. Essentially, while the system aims to accurately identify activities by considering various contexts and using complex rules (RDR rules), this sophistication comes at a cost. Specifically, the computational effort required to build and keep up-to-date the model that combines context with activity recognition can slow down the system. This means that when trying to recognize activities as they happen, the system might lag or be less responsive due to the heavy processing needed to analyze the data and apply the rules. This delay or reduction in speed could affect the system's efficiency in real-time situations, making it a critical area for further improvement.

Another study by Gao et al. [13], introduces a system that can simultaneously identify where a smartphone is located (like in a pocket or hand) and what activity the person is doing (like walking or running) by analyzing movement data from the phone's sensors. They use a special approach called multi-task learning (MTL), which allows the system to learn about different things at the same time, making it more efficient. To deal with the issue of the phone being held in different ways, which can mess up the data, they preprocess the data using a method that adjusts the coordinates based on something called quaternions. This technique ensures that the system can understand the data correctly no matter how the phone is oriented.

The benefits of their approach include the use of the multi-task learning technique, which makes it possible to learn about the phone's location and the user's activity together, improving the overall performance. Also, by preprocessing data with quaternions, they make sure the system can accurately interpret the sensor data, even if the phone's position changes. However, the system has its drawbacks. It relies solely on the phone's motion sensors. This limitation could be problematic, especially when trying to pinpoint the exact location of the smartphone. Other types of sensors or data might be needed to improve location classification and make the system more versatile and accurate in different scenarios.

The system developed in [14], uses two wearable sensors, one on the wrist and the other on the ankle, to measure how fast and in what direction parts of the human body are moving. These measurements are then sent without wires to a computer. The computer processes this data through a couple of advanced mathematical techniques to figure out what kind of activity the person is doing, like running or jumping. The good points of this system include the use of just two sensors, which makes it less intrusive or cumbersome for the person wearing them. Plus, the fact that data is sent wirelessly to a computer means there's no need for cumbersome wires that could restrict movement. However, there are some downsides. Using only two sensors might not be enough to catch all the different ways the body can move or all the positions it can be in, which means some activities might not be recognized accurately. Additionally, relying on a wireless connection to send data to the computer could lead to problems. If the connection is weak, interrupted, or non-existent, the system won't work as it should, potentially missing out on important data or failing to monitor activities correctly.

The system designed in [15], recognizes everyday activities, like eating or driving, using a special kind of artificial intelligence called a long short-term memory (LSTM) network. The system gathers data from sensors that people wear on their wrists, ankles, and waist. It then uses a deep learning model, which is a way for computers to learn from However, the paper points out a couple of limitations. First, they only used a small amount of data, which means their system might not work as well in real life where there are many different activities. Second, their method of learning how important each sensor's data should be doesn't change over time. So, if the way activities are performed changes, the system might not be able to recognize them as well.

Another study [16], introduces a system called Marfusion, which combines a type of AI known as a convolutional neural network (CNN) with an attention mechanism to analyze human movement. This system takes data from different

kinds of sensors to get a complete picture of what the person is doing. For each sensor, it uses a CNN to pull out important features or patterns in the data. Then, it uses a special method called scaled self-attention to decide how important each sensor's data is. This means the system can focus more on the sensors that are giving the most useful information at any given time. After figuring out which features are most important, the system smartly combines them and then uses several layers of processing - including batch normalization, dropout, ReLU, and softmax - to classify the data into different activities. It guesses what activity is being performed based on the patterns it has learned from the sensor data. The system worked pretty well in tests, but it was only tried out with a limited range of movements. This means while it can accurately recognize some activities, it might not be ready for use in real-world situations where people do a wide variety of things. The researchers suggest that the system needs to be tested with more types of movement before it can be relied on in everyday settings.Another study [17] introduces a physique-based sensor-human activity recognition (S-HAR) architecture aimed at improving the performance of deep learning (DL) models in recognizing human activities. Traditional DL approaches often struggle with this task because they do not consider the physical characteristics of individuals, which can vary widely. To address this, the researchers designed an architecture that incorporates physical attributes into the DL process. They tested their system using the HARSense dataset, a publicly available collection of raw sensor data from smartphones. They trained and evaluated five different DL networks with this dataset. The results showed a significant improvement in accuracy and F1 scores when the models included physical characteristics. The limitations include, the need for additional physical data might limit the system's applicability in situations where such data isn't available. The added complexity of the system could increase computational demands, potentially hindering its use in real-time applications. This study [18] introduces a framework for position-independent human activity recognition (HAR) using deep learning on sensor data from wearable devices. The innovative Att-ResBiGRU model, evaluated on the Opportunity, PAMAP2, and REALWORLD16 datasets, combines convolutional layers for spatial feature extraction, a ResBiGRU block for temporal feature capture, and an attention mechanism to improve recognition accuracy. Despite its superior performance, the model has some limitations: it requires up to 4MB of memory, making it unsuitable for low-power devices; sensor drift over time can affect accuracy; continuous data collection can deplete battery life; and transmitting raw sensor data raises privacy concerns.

The research has been divided into the following sections: Section II discusses the problem statement and contribution and then material and methods, including noise removal, signal windowing and segmentation, feature extraction for machine learning and deep learning, and feature fusion are presented in Section III. Section IV discusses the experimental results. Section V presents the limitations and plans to resolve these issues. Finally, the research study is concluded.

II. PROBLEM STATEMENT AND CONTRIBUTION

The ubiquity of smartphones, equipped with an array of sensors such as accelerometers, microphones, gyroscopes, and GPS has opened new avenues for continuous, in-the-wild monitoring and analysis of human behaviors and their environments. Recognizing human locomotion and localization with high accuracy and efficiency is pivotal in developing applications that can provide personalized health advice, enhance personal safety, and improve the quality of life. This study takes advantage of such advancements to explore human locomotion and localization recognition using data from accelerometers, microphones, gyroscopes, magnetometers, and GPS. We apply Deep Polynomial Neural Networks (DPNN) and Multilayer Perceptron (MLP) across three distinct datasets: the Continuous In-The-Wild Smart Watch Activity Dataset, the Huawei Locomotion Dataset, and the Extra Sensory Dataset. Our approach employs two distinct methodologies for activity recognition: DPNN for deep learning-based feature extraction and MLP with manual feature extraction techniques, including Linear Predictive Coding Cepstral Coefficients (LPCC), step length, signal magnitude area, spectral, and sound features. Despite the proliferation of research in HAR, challenges persist in achieving high accuracy and computational efficiency simultaneously. Our work contributes to this evolving field by rigorously comparing the performance of DPNN and MLP models in recognizing both locomotion and localization activities. Through our experimentation, we demonstrate that DPNN consistently outperforms MLP in terms of accuracy across all datasets. However, we also highlight the computational efficiency of MLP, presenting a trade-off between accuracy and computational demand. The significance of our study lies in its comprehensive analysis and comparison of two machine learning approaches applied to a diverse set of real-world data, offering insights into the potential of smartphone sensors in enhancing HAR systems. By exploring the strengths and limitations of DPNN and MLP in the context of HAR, this study contributes valuable perspectives to the ongoing discourse on optimizing machine learning techniques for mobile sensing applications, paving the way for future innovations in the field.

III. PROPOSED SYSTEM

In our system, we crafted a robust system aimed at accurately identifying both human movements and location-specific activities. The initial step in our process involved refining the raw sensor data, for which we employed a Butterworth low-pass filter [19]. This technique effectively minimized noise, ensuring that the subsequent analysis was based on clean, reliable data. To tackle the challenge of processing extensive sequences of continuous signal data, we adopted the Windowing technique for efficient segmentation, setting

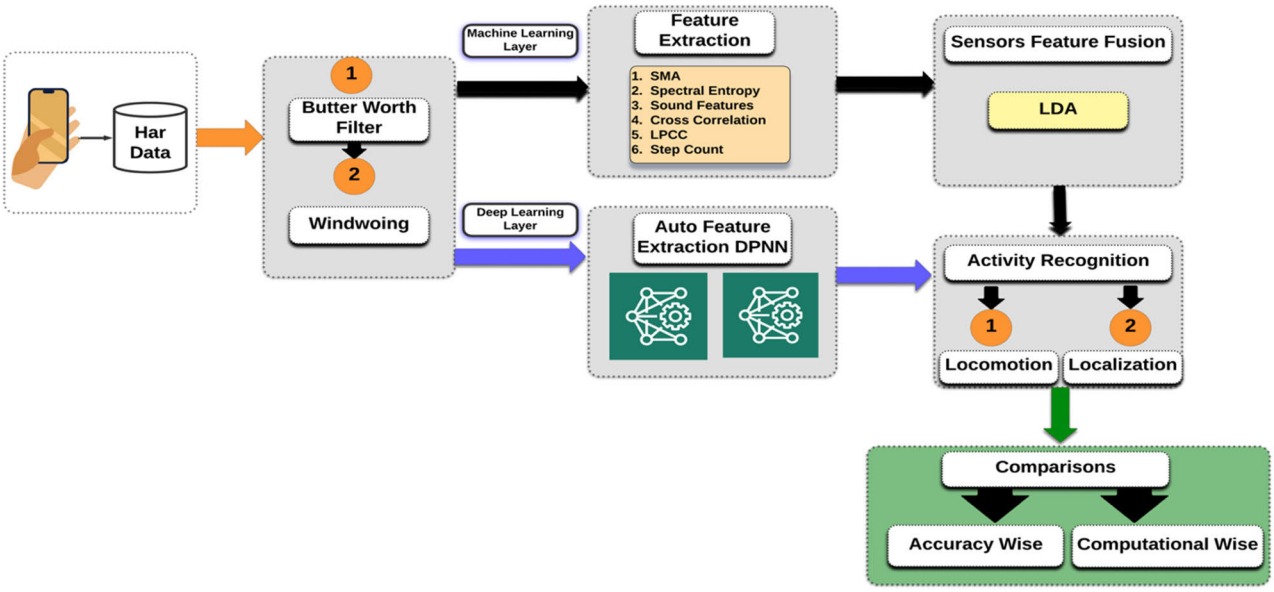


FIGURE 1. Architecture of the proposed system.

the stage for deeper analysis. Our system's architecture is distinctively divided into two principal layers: one focusing on machine learning and the other on deep learning. Within the machine learning layer, we engaged in the extraction of handpicked features, which are necessary for the recognition of distinct human activity patterns. Conversely, the deep learning layer was tasked with processing the segmented data through a Deep Polynomial Neural Network (DPNN). This choice was driven by the DPNN's capacity to adeptly handle the complex and nonlinear dynamics characteristic of human activity data. Following the preparatory stages, we then turn into the classification stage, employing both a Multilayer Perceptron (MLP) for the machine learning component and the DPNN for deep learning analysis. It was observed that the DPNN markedly exceeded the MLP in accuracy, showcasing its enhanced ability to accurately recognize complex activity patterns. However, a comparative analysis of the two methodologies in terms of accuracy unveiled that the DPNN, despite its superior performance, is significantly more demanding in computational resources. This insight into the trade-offs between accuracy and computational expenditure underscores the importance of balancing these factors in the pursuit of developing an advanced and efficient system for the recognition of human activities. The system architecture is presented in Fig. 1.

A. NOISE REDUCTION

1) BUTTER-WORTH LOW-PASS FILTER

In our study, we aimed to clear the sensor data of any noise that could interfere with our analysis. To achieve this, we decided to use a Butterworth low-pass filter [20]. This type of filter is well-regarded for its capability to smoothly remove high-frequency noise from the data while preserving the important trends and patterns that we want to analyze. The

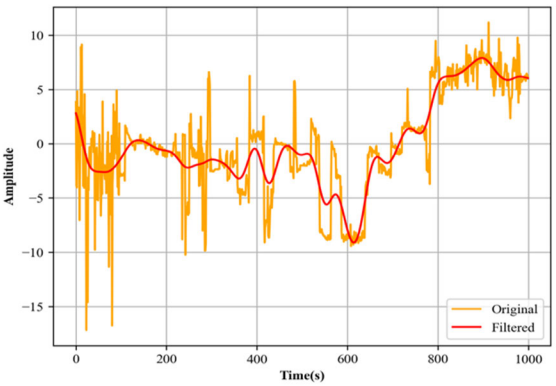


FIGURE 2. Accelerometer noisy vs filtered signal.

Butterworth filter is particularly useful because it provides a clear signal without the sharp cut-off that some other types of filters can cause. This means that the data we're left with after filtering retains its natural shape, making it easier to work with and interpret.

Mathematically, the Butterworth low-pass filter is designed to have a frequency response that is as flat as possible in the passband. It avoids ripples both in the passband and the stopband. The filter's transfer function, which tells us how it modifies the signal at different frequencies, has a specific form that depends on the order of the filter that is, how many times it applies a simple filtering process.

For a given order n, the transfer function $X(s)$ of the Butterworth low-pass filter is defined as:

$$X(s) = \frac{1}{1 + \left(\frac{s}{w_d}\right)^{2n}} \tag{1}$$

where s is the complex frequency variable. w_d is the cut-off frequency, beyond which we want attenuate the signal.

The higher the order n of the filter, the sharper the transition between the passband and the stopband. However, higher-order filters can be more complex to implement and may introduce delays in signal processing. In our implementation, we choose 2nd order that provided a good balance between effectively removing noise and maintaining the integrity of the underlying signal.

We then applied this filter to our sensor data, smoothing out the high-frequency fluctuations and preparing the data for the next stages of our analysis. Fig. 2 shows the original vs filtered signal.

B. DATA WINDOWING AND SEGMENTATION

After cleaning our sensor data with the Butterworth filter, the next step was to break down the continuous stream of data into manageable chunks. To do this, we used a technique called windowing [21], [23]. This method involves dividing the continuous data stream into fixed-size segments or 'windows,' each containing a specific number of data samples. We decided on windows that span 3 seconds each. Within each of these 3-second intervals, we collected 250 data samples from our sensors. This choice was a strategic balance, ensuring that each window was large enough to capture relevant activity patterns, but not so large as to mix different activities. Mathematically, the process of windowing can be represented as follows:

Let $x(t)$ be our continuous signal. We define $w(t)$ as our window function, which is non-zero only within the window's limits and zero otherwise. For a window size of 3 seconds, our window function can be represented as:

$$\begin{cases} 1 & \text{for } 0 \leq t < 3 \text{ seconds} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Our segmented data $x_w(t)$ is then the product of our continuous signal and our window function:

$$x_w(t) = x(t) \cdot w(t) \tag{3}$$

For each window, we take the segment of $x(t)$ from $t = n \times 3$ seconds to $t = (n+1) \times 3$ seconds, where n is the window number starting from 0. Visually the figure for segmentation can be seen in Fig. 3.

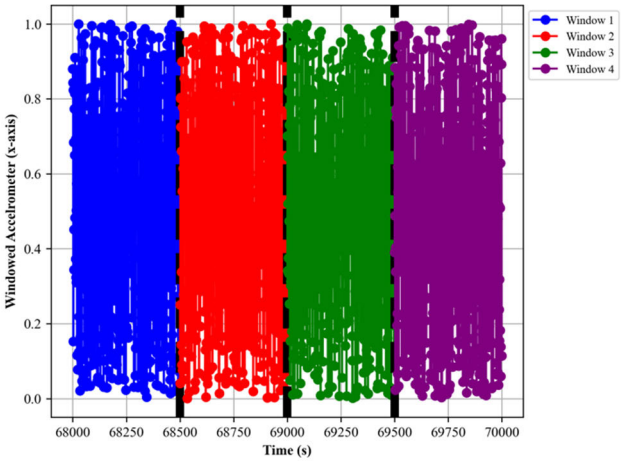


FIGURE 3. Accelerometer sensor windowed data.

C. FEATURE EXTRACTION FOR MACHINE LEARNING

The first layer dives into machine learning, where we engage in the art of feature extraction [24], [27]. This initial stage is crucial as it sets the groundwork for the algorithms to recognize patterns effectively. We carefully select features that encapsulate the essence of the sensor data, including features like Signal Magnitude Area (SMA), Linear Predictive Coding (LPC), characteristics derived from sound, Spectral Entropy, and the count of steps taken.

1) SIGNAL MAGNITUDE AREA

Signal Magnitude Area (SMA) [28], [29] is a way to sum up the activity captured by sensors, kind of like creating a summary of a story. It's particularly handy when you have a bunch of numbers coming in from sensors on different axes – like up-down, left-right, and forward-backward movements – and you want to get a single number that tells you, in general, how much movement is going on. Researchers often use SMA in fields like sports science or physical therapy. Imagine you have athletes wearing sensors, and we want to know how intense their workout was, or you're helping someone recover from an injury, and we need to track how much they're moving. SMA can give us this information in a clear, simple number.

To calculate SMA, we take the sum of the absolute values of the sensor readings across all three axes for a specific time frame, and then we average it. So, if our sensors give us readings every second for each of the three axes (x, y, and z), here's how we did it mathematically:

First, for each second, we find the absolute values of the x, y, and z readings because we're interested in the amount of movement, not the direction. Then, we add these absolute values together. This gives us the total movement for each second. So, the formula for one-second looks like this:

$$SMA_{1sec} = \frac{|x| + |y| + |z|}{3} \tag{4}$$

But we don't just want one second; we want to know the average over, say, a minute. So, we add up these one-second totals for each second in the minute and then divide them by the number of seconds in a minute (which is 60).

The formula for a minute:

$$SMA_{1min} = \frac{\sum_{i=1}^{60} SMA_{1seci}}{60} \tag{5}$$

In this way, SMA takes all that messy, complex data and turns it into something we can easily understand and use. In. Fig. 4 The SMA is presented graphically.

2) SPECTRAL ENTROPY

Spectral Entropy [30] is a fascinating concept that tells us about the orderliness or complexity of a signal. We can think of it as a way to measure how predictable or chaotic a piece

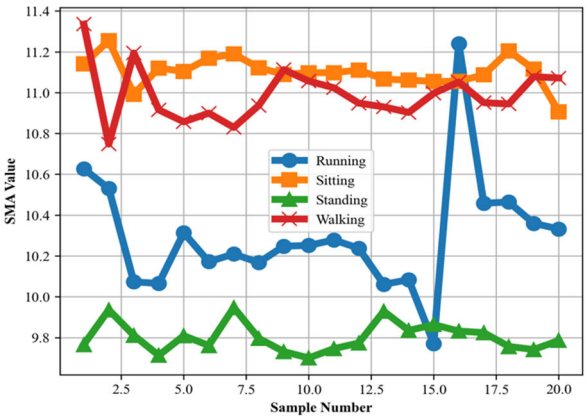


FIGURE 4. SMA calculated for accelerometer.

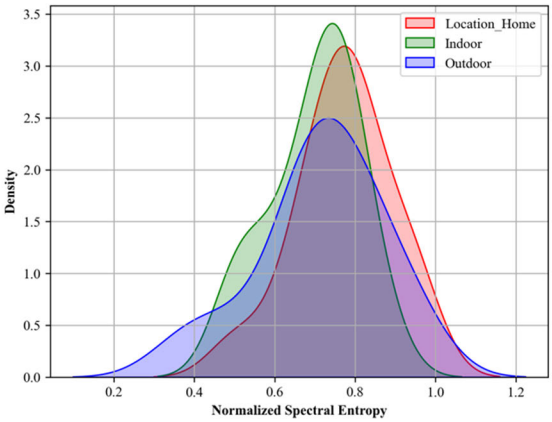


FIGURE 5. Spectral Entropy is calculated for different activities.

of music is; a pure tone is very predictable, so it has low Spectral Entropy [31], while a cacophony of random sounds is very unpredictable, giving it high Spectral Entropy [32]. It's used in various fields, like neuroscience to understand brain activity, in engineering for signal processing, and even in finance to analyze the volatility of stock markets. It's about getting a grip on how much order or disorder is present in a system. To calculate Spectral Entropy, we first get the power spectrum of the signal, which shows how the power of the signal is distributed across different frequencies. Then, we normalize the power spectrum so that it sums up to one—this makes it a probability distribution. The Spectral Entropy is then calculated using the Shannon entropy formula, which, in simplified terms, is the sum of the probability of each frequency multiplied by the logarithm of that probability, summed across all frequencies.

The basic equation for Shannon entropy:

$$K = \sum_j X\left(f_j\right) \log X\left(f_j\right) \tag{6}$$

where $X\left(f_j\right)$ is the normalized power of the i^{th} frequency component of the signal.

In the context of the below graph, this equation has been applied to the signal data from each environment, with the resulting entropy values plotted to show the overall distribution of complexity in the sound environment for each location. From the graph below, 'Home' activity tends to have a lower Spectral Entropy, suggesting a more uniform or less complex set of sounds, perhaps due to a more controlled environment. 'Outdoor' activity has higher entropy, indicating a lot of varied sounds and a higher level of unpredictability or complexity in the soundscape.

3) SOUND LEVEL EXTRACTION (DB)

We take a close look at the levels of noise in different environments by measuring the ambient sound levels [33]. This feature is represented as the 'Decibel Level'. The decibel (dB) is the unit [34] we use for measuring sound intensity. It tells us how powerful a sound is compared to a reference level. On our graph, we plotted the decibel levels for different places: outdoors, indoors, and at home. We can see that outdoor sounds tend to be louder, which makes sense given the traffic, people, and all the activity that happens there. For each location, we also calculated the average sound level, which is what we see as the 'mean' lines on the graph. These averages help us get a baseline of typical noise levels for each setting. By looking at the peaks and valleys of the lines for each location, we get an idea of the sound variations over time.

Mathematically, we can express the average decibel level over several samples (n) as:

$$\text{Average dB level} = \frac{1}{n}\sum_{j=1}^{n} dB_j \tag{7}$$

where dB_j is the sound level in decibels for the i^{th} sample. This sound level feature is very beneficial in our machine learning layer because it can help differentiate activities based on their noise profiles. For instance, the constant hum of machinery could signify industrial work, whereas intermittent, softer sounds might suggest an office setting. Fig. 6 presents the dB level for different localization activities.

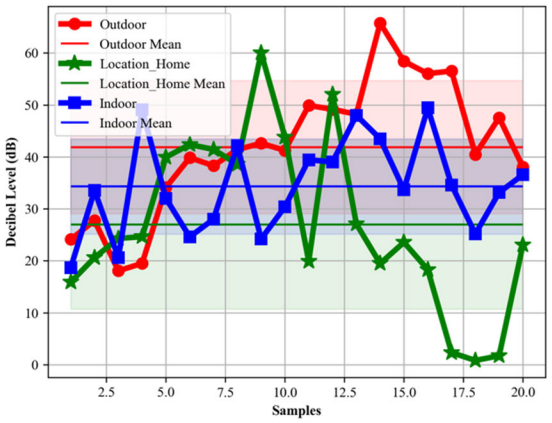


FIGURE 6. Sound Decibel level (dB) calculated for different activities.

4) CROSS-CORRELATION COEFFICIENT

To calculate the cross-correlation feature [35], [37], we select segments of data that represent steps or strides for walking

activity. We then mathematically compare two segments of the signal at different times to calculate the cross-correlation coefficient. This is done by shifting one segment (or signal) along the time axis and calculating the product of the two signals at each shift (lag). The formula used is:

$$W_{ij}(\tau) = \sum_t i(t) \cdot j(t+\tau) \tag{8}$$

Here $i(t)$ and $y(t)$ are two signals, and τ is the lag.

The graph (Figure 7) shows this relationship visually. The x-axis, labeled 'Lag,' represents the amount of shift when comparing two signals. The y-axis shows the cross-correlation coefficient [38], a number between −1 and 1 that tells us how closely the two signals match. A coefficient of 1 means they're perfectly in sync, 0 means no correlation, and −1 means they're inversely related. In the context of walking, a high correlation at certain lags indicates a consistent stride pattern, while a lower correlation may point to irregularities or asymmetry in gait. The peaks and troughs in the graph represent points of high and low similarity between the movement recorded at different points in time. This feature is particularly useful in applications like sports analytics, rehabilitation, and health monitoring [39], [40], where understanding the consistency and symmetry of movements is essential. In Fig. 7 the graph is shown.

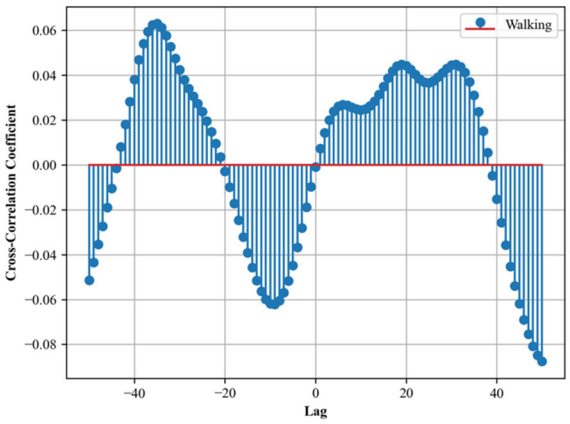


FIGURE 7. Cross-correlation Coefficient calculated for walking activities.

D. FEATURE FUSION USING LDA (LINEAR DISCRIMINANT ANALYSIS)

Incorporating Linear Discriminant Analysis (LDA) [41], [42], [43], [44], [45] into our multisensory data for feature fusion involves a series of steps, beginning with the standardization of diverse sensor data streams. This ensures that each sensor's data is given equal weight in the analysis. We label this data by the corresponding activities it represents. The mathematical core of LDA [46], [47], [48], [49], [50] is to maximize the ratio of the between-class scatter to the within-class scatter, thereby achieving maximum class separability.

Mathematically, this involves calculating two key matrices:

$$X_y = \sum_{i=1}^{c} X_i \tag{9}$$

Algorithm 1 LDA

Input: Multidimensional labeled data from various sensors, and the number of dimensions for the output.
Preprocessing: Normalize the sensor data and label them according to the corresponding activities.
Compute:
- Mean vectors for each class.
- The within-class and between-class scatter matrices.

Solve:
Calculate the eigenvectors and eigenvalues of $X_w^{-1}X_B$
Select:
Choose the top eigenvectors based on the largest eigenvalues to form a transformation matrix.
Transform:
Apply the transformation matrix to the data to obtain lower-dimensional features that are most discriminative.
Output:
The lower-dimensional data ready for the classification.

where X_i is the scatter matrix for each class i:

$$X_i = \sum_{x \in D_i} (x - m_i)(x - m_i)^T \tag{10}$$

In this equation, x represents the data samples for class i, D_i represents the set of all samples for class i, D_i and is the mean vector of the samples for class i.

The between-class scatter matrix can be calculated as:

$$X_B = \sum_{i=1}^{c} N_i (m_i - m)(m_i - m)^T \tag{11}$$

where m is the overall mean of the data, m_i is the mean vector of class i, and N_i is the number of samples for class i.

Finally, to find a suitable linear combination of features that separates the classes, we solve the eigenvalue problem:

$$X_w^{-1}X_B \underset{\upsilon}{\longrightarrow} = \lambda \underset{\upsilon}{\longrightarrow} \tag{12}$$

where $\underset{\upsilon}{\longrightarrow}$ are the eigenvectors and λ are the eigenvalues. The eigenvectors corresponding to the largest eigenvalues will form the new axes for the reduced-dimensional space. Algorithm. 1 shows the working of LDA.

By projecting our high-dimensional sensor data onto these new axes, we obtain a lower-dimensional feature space that emphasizes the differences between activities, which is important for effective classification. The LDA graph is shown in Fig. 8.

E. FEATURE EXTRACTION USING DEEP LEARNING

Deep Polynomial Neural Networks (DPNNs) [51], [52], [53], [54] are an advanced form of neural networks that have gained attention in the field of deep learning for their ability to model complex, high-order interactions between input features. Unlike traditional neural networks that learn linear combinations of inputs, DPNNs utilize polynomial functions to uncover the multifaceted relationships inherent in the data. This makes them particularly well-suited for tasks where the

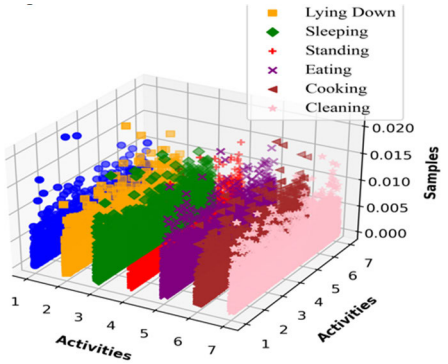


FIGURE 8. LDA was calculated for different locomotion activities.

input features interact in non-linear and complex ways, which is often the case in sensor-based human activity recognition. Applications of DPNNs span a variety of domains, from financial forecasting where non-linear patterns are common, to bioinformatics for modeling complex biological processes. In our study, we utilize a DPNN to process data from accelerometers, gyroscopes, magnetometers, microphones, and GPS units, acknowledging the rich and diverse nature of the sensor data. The network architecture is carefully crafted to handle the multi-class classification task across three distinct datasets (Huawei, Extrasensory, and CIWSA), each encompassing a variety of locomotion and localization activities.

1) THE ARCHITECTURE OF PROPOSED DPNN

a: INPUT LAYER

The first input layer is dedicated to interfacing with our multi-modal sensor data. Each sensor's output is normalized to ensure that the magnitude of the readings does not bias the network, making sure that all inputs contribute equally to the feature extraction process.

b: POLYNOMIAL EXPANSION LAYER

Next Polynomial expansion layer forms the core of our DPNN. Each layer performs a polynomial expansion of the inputs it receives, which involves creating new features by calculating non-linear combinations of the input features. For instance, if f_1 and f_2 are two features, a second-order polynomial layer would generate features such as f_1^2, f_2^2, and $f_1 \cdot f_2$, among others.

c: INTERMEDIATE PROCESSING LAYERS

In the intermediate processing layer we insert normalization and dropout layers to manage the complexity of the features generated by the polynomial expansions. These layers help to mitigate overfitting and improve the generalization of the network by selectively 'forgetting' certain neuron connections during training.

d: ACTIVATION FUNCTIONS

Despite DPNNs often forgoing the need for traditional activation functions due to their inherent non-linear nature, we incorporate activation functions such as Rectified Linear Units (ReLU) to introduce additional non-linearity and help the network's learning process. This is expressed as:

$$f(x) = \max(0, x) \text{ for each input } x \tag{13}$$

e: FUSION AND CONTEXTUAL LAYERS

To integrate the features extracted from different sensors and datasets, we use fusion layers that combine these features while taking into account the context provided by the data, such as the specific type of activity and its location.

f: OUTPUT AND CLASSIFICATION LAYERS

The final stages of our DPNN consist of fully connected layers followed by softmax classification layers. The softmax function is applied to the final polynomial features to obtain the probability distribution over the classes, with the formula given by:

$$P(x_i) = \frac{e^{zi}}{\sum_j e^{zj}} \tag{14}$$

Here z_i is the output of the last fully connected layer for class i.

g: TRAINING AND OPTIMIZATION

We trained the model using a variant of the stochastic gradient descent algorithm, specifically designed to handle the high-order polynomial parameters. We define a loss function appropriate for multi-class classification, typically the categorical cross-entropy loss, which is computed as:

$$Z = -\sum_k t_k \log(P(y_k)) \tag{15}$$

where t_k is the true label in a one-hot encoded form. In Table. 2 the hyperparameters are given while training the DPNN.

TABLE 1. Hyperparameters involved in training the DPNN.

Hyperparameter	Value
Polynomial Degree	3
Number of Dense Layers	4
Dense Layer Units	128, 128, 128, 64
Activation Function	ReLU
Output Activation Function	Softmax
Loss Function	Categorical Crossentropy
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Epochs	100

IV. PERFORMANCE EVALUATION

The performance evaluation of our system is a critical stage that ensures the robustness and accuracy of our human activity recognition model. We have conducted comprehensive testing using three benchmark datasets, each uniquely challenging and rich in sensory information. These datasets are Huawei Locomotion, ExtraSensory, and Continuous Activity

in the Wild. Each offers a diverse array of user activities, sensor modalities, and contextual information, providing a thorough testing ground for our system's capabilities. To ensure that our findings are reflective of a typical modern computational environment, the evaluation was carried out on a system equipped with 16 GB of RAM, a 3.1 GHz processor, 512 GB SSD, and running Mac OS. This setup provides a balance between computational power and efficiency, relevant to both academic research environments and industry applications.

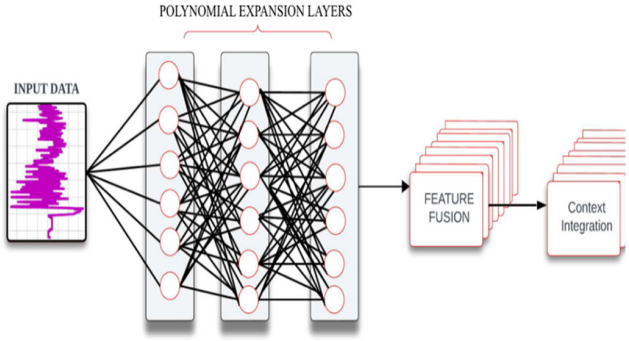


FIGURE 9. Overview of proposed deep polynomial neural network (DPNN).

A. DATASET DESCRIPTION

1) THE HUAWEI LOCOMOTION DATASET

The Huawei dataset [55] captures day-to-day life activity across various settings in the UK, spanning an extensive period of nearly seven months in 2017. It was carefully recorded by three individuals as they carried out their daily activities, which included a range of motions from complete stillness to vigorous running, from cycling to navigating the city by car or public transport. To gather a comprehensive set of data, each participant was equipped with four Huawei Mate 9 smartphones placed in different locations on their person: one in hand, another positioned at the torso area, a third-placed in the hip pocket, and the fourth in a backpack or a handbag. This strategic placement across the body allowed for a unique opportunity to investigate how the location of a sensor can influence the accuracy of tracking a person's activity. In addition to the movement data, the dataset was enriched with contextual annotations. It documented whether the participants were indoors or outdoors and described the specific activity they were engaged in at any given moment. This additional layer of information provides a multifaceted view of each participant's movements, offering valuable context for researchers to recognize activity and contextual analysis.

2) THE EXTRASENSORY DATASET

The ExtraSensory dataset [56] is an extensive collection of sensor data recorded from a group of 60 individuals. Each person was assigned a unique identifier, and their activities were captured in intervals, typically lasting around one minute. This data wasn't just sourced from a single type of device; instead, it was gathered from the participants' smartphones and smartwatches, offering a rich blend of information. Beyond the raw sensor data, which includes everything from accelerometers to location services, the participants themselves provided contextual details, presenting a clearer picture of their activities during data capture. The contributors to this dataset were from a vibrant campus community, with a balanced mix of iPhone and Android users, and a diverse gender distribution. It's worth noting that while the dataset is comprehensive, it's not without gaps some sensors weren't always available due to the device's limitations or user preferences, like turning off location services or when audio wasn't captured during phone calls.

3) THE CONTINUOUS IN-THE-WILD SMARTWATCH ACTIVITY DATASET

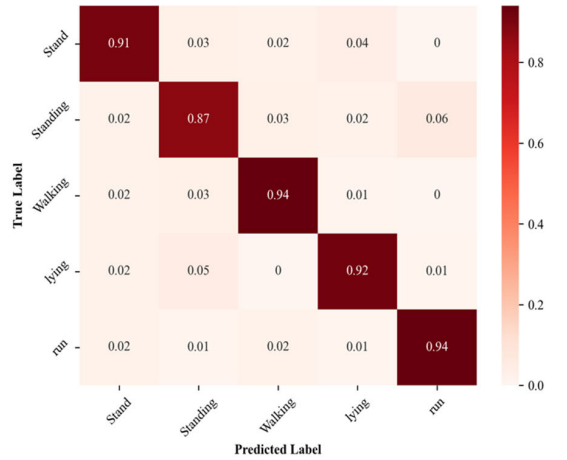
The Continuous In-the-Wild Smartwatch Activity (CIWSA) dataset [57], consists of real-world sensor and location data collected from 49 individuals. The uniqueness of this dataset lies in its continuous and naturalistic capture of daily activities through an Apple Watch. Unlike controlled studies, this data reflects real life, with participants going about their days, offering a transparent lens into human behavior and routines. The data collection was thorough, with a high-frequency capture rate, and supplemented with GPS location data. Additionally, participants provided valuable feedback by answering prompts about their current activity, enriching the dataset with self-reported information.

B. EXPERIMENT 1: CONFUSION MATRIX FOR LOCOMOTION (MULTI-LAYER PERCEPTRON MACHINE LEARNING)

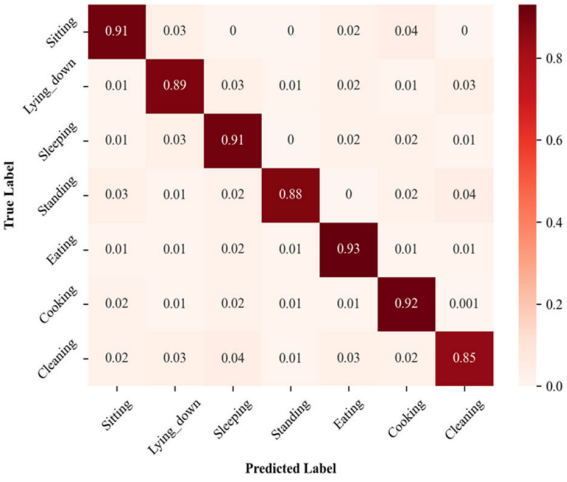
In Experiment 1, we evaluate the performance of a Multilayer Perceptron (MLP) machine learning model in classifying locomotion activities. Each confusion matrix in Fig. 10 corresponds to one of the datasets used in our study: the CIWSA Locomotion, ExtraSensory Locomotion, and Huawei Locomotion datasets. These matrices display the model's predictions across the x-axis and the true labels down the y-axis. The diagonal entries of the matrix represent correct predictions, where the predicted activity matches the true activity. Values off the diagonal indicate misclassifications. For a confusion matrix A where each element $A_{i,j}$, represents the number of instances known to be in group i but predicted to be in group j, the accuracy of the model for each activity can be calculated as:

$$\text{Accuracy for activity}i = \frac{A_{i,i}}{\sum_j A_{j,i}} \tag{16}$$

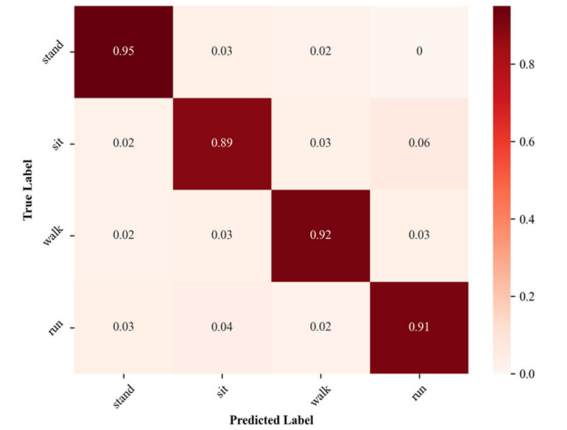
For instance, activities like 'walking' and 'running' might be easier to distinguish due to their distinct motion patterns, while 'standing' and 'sitting' could be more easily confused due to their similarity in terms of movement or lack thereof.



(a)



(b)



(c)

FIGURE 10. Confusion matrices for MLP locomotion (a) CIWSA (b) Extrasensory (c) Huawei dataset.

C. EXPERIMENT 2: CONFUSION MATRIX FOR LOCALIZATION (MULTI-LAYER PERCEPTRON MACHINE LEARNING)

In Experiment 2, we focus on evaluating the efficacy of a Multilayer Perceptron (MLP) machine learning model in the domain of localization - determining a user's location based on sensor data. The evaluation employs confusion matrices, which are pivotal in understanding how well the model distinguishes between different locations, such as indoors, outdoors, in a car, or various urban environments.

For instance, distinguishing between being 'in a car' and 'outdoors' may present a challenge if the environmental sensor signatures are similar. However, a high accuracy in differentiating 'indoors' from 'at the beach' indicates the model's strength in recognizing distinct environmental characteristics. The confusion matrices for localization activities are shown in Fig. 11.

D. EXPERIMENT 3: CONFUSION MATRIX FOR LOCOMOTION (DEEP POLYNOMIAL NEURAL NETWORK)

In Experiment 3, we start by assessing the performance of the Deep Polynomial Neural Network (DPNN) for the classification of various locomotion activities. Utilizing confusion matrices, we were able to visually dissect the accuracy of the DPNN in recognizing distinct motions such as sitting, standing, walking, lying down, and running. In. Fig. 12 the confusion matrices for locomotion activities can be seen.

E. EXPERIMENT 4: CONFUSION MATRIX FOR LOCALIZATION (DEEP POLYNOMIAL NEURAL NETWORK)

In Experiment 4, we turn our attention to evaluating the Deep Polynomial Neural Network (DPNN) in the context of localization tasks. This involves interpreting confusion matrices generated from the DPNN's predictions to assess its capability to accurately identify a user's location. By analyzing these matrices for each dataset, we gain insights into how well the DPNN distinguishes between different environments such as being indoors, outdoors, in a car, or transit. In Fig. 13 the confusion matrices for localization are presented.

F. EXPERIMENT 5: COMPARISONS IN TERMS OF ACCURACY

1) CIWSA LOCOMOTION AND LOCALIZATION

In Fig. 14, we observe the DPNNs outperforming MLPs, suggesting that the complex interactions within the ciwsa data are better captured by the polynomial architectures of DPNNs.

2) EXTRASENSORY LOCOMOTION AND LOCALIZATION

The performance is competitive, but again DPNNs showcase a slight edge over MLPs. This could be attributed to the richness and diversity of sensor data in the ExtraSensory dataset, which may benefit from the higher-order feature interactions modeled by DPNNs.

3) HUAWEI LOCOMOTION AND LOCALIZATION

The trend continues with DPNNs maintaining higher accuracy. The complex patterns within the Huawei dataset, possibly due to varying sensor positions and the context data, seem to be more effectively decoded by the DPNN model. The comparisons between both models can easily be interpreted in Fig. 14.

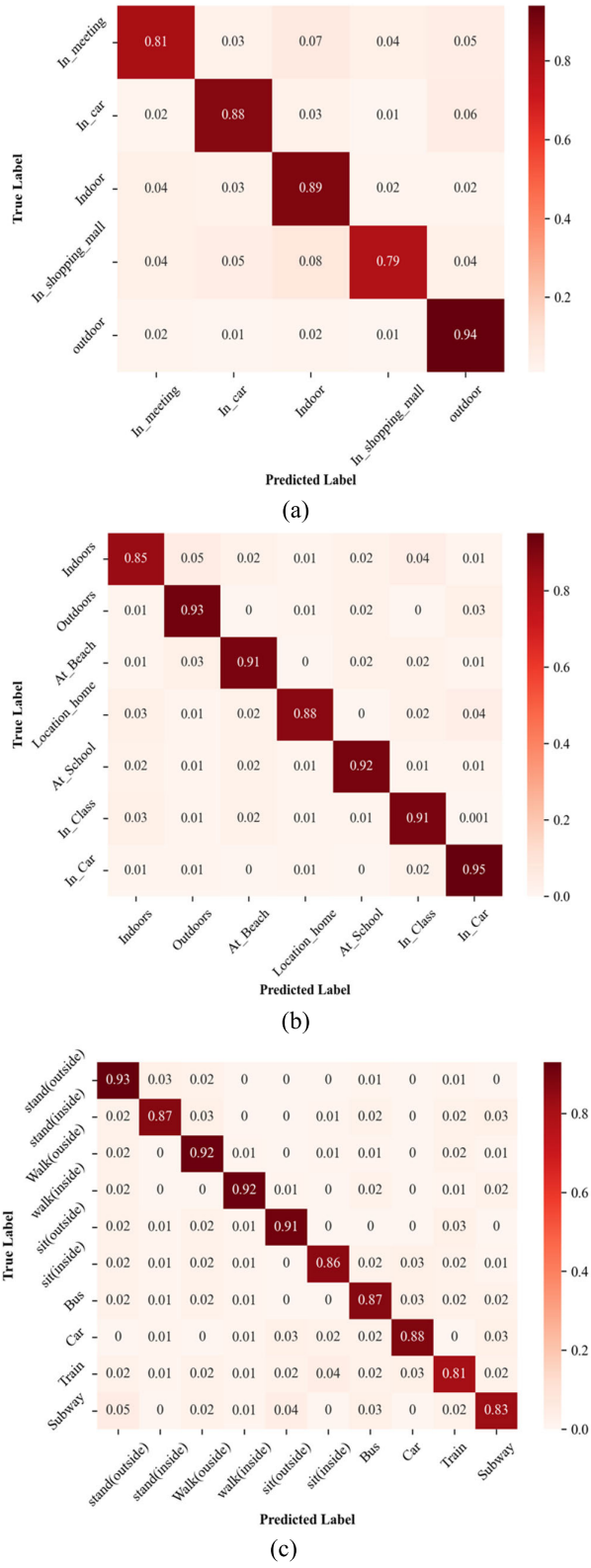


FIGURE 11. Confusion matrices for MLP localization (a) CIWSA (b) Extrasensory (c) Huawei dataset.

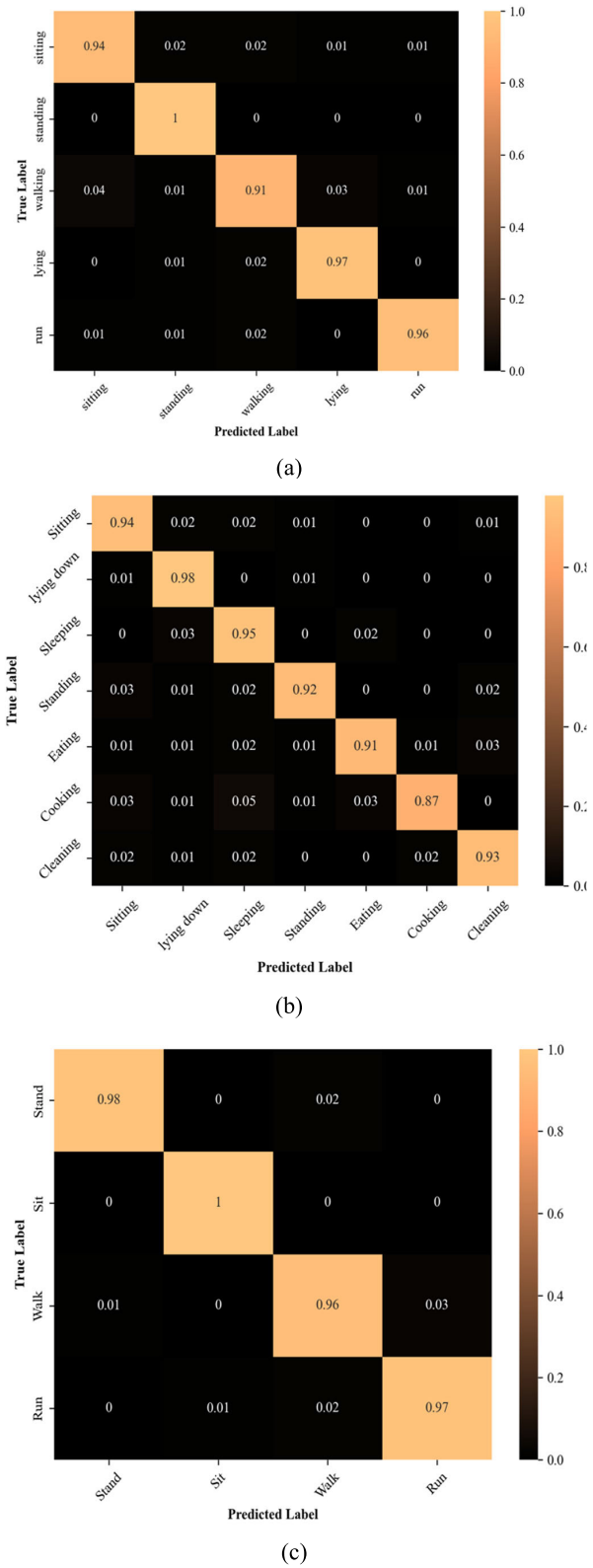


FIGURE 12. Confusion matrices for DPNN locomotion (a) CIWSA (b) Extrasensory (c) Huawei dataset.

G. EXPERIMENT 6: CROSS-VALIDATION

In our evaluation, we've implemented 5-fold cross-validation to rigorously test the robustness of our machine learning models, both Multilayer Perceptron (MLP) and Deep Polynomial Neural Network (DPNN), across the different datasets. Cross-validation is a technique used to assess how the results

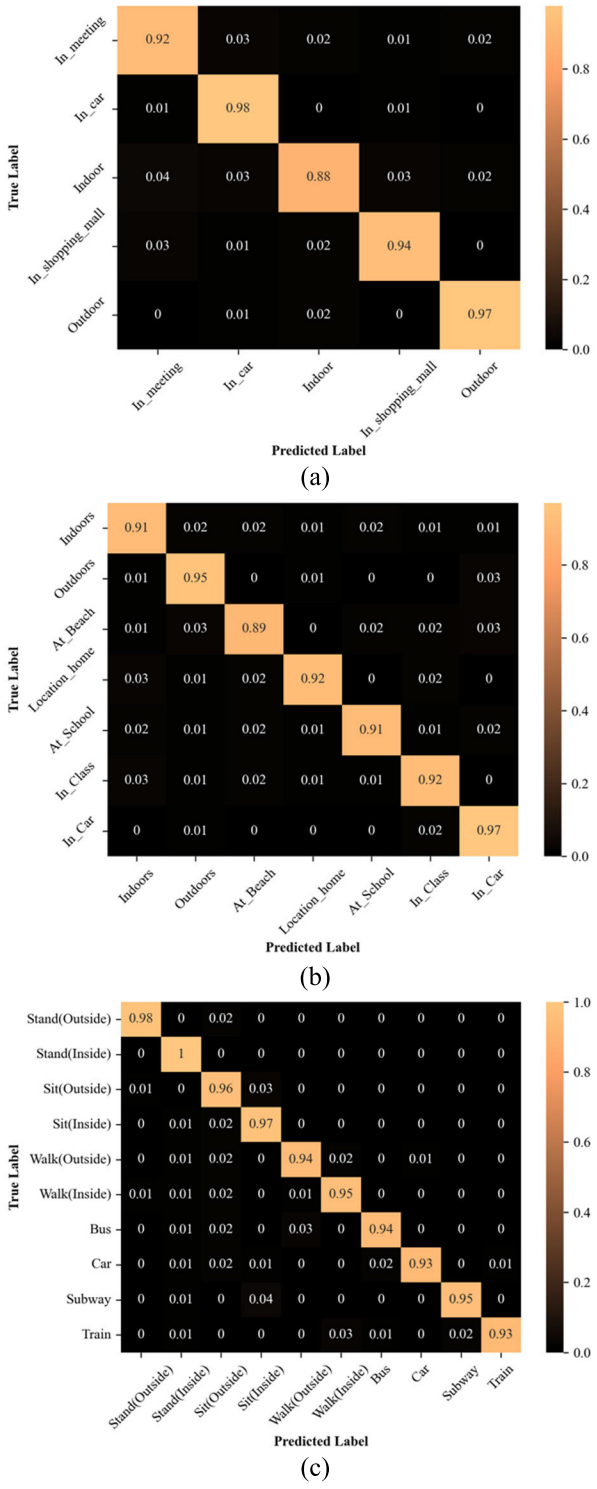


FIGURE 13. Confusion matrices for DPNN localization (a) CIWSA (b) Extrasensory (c) Huawei dataset.

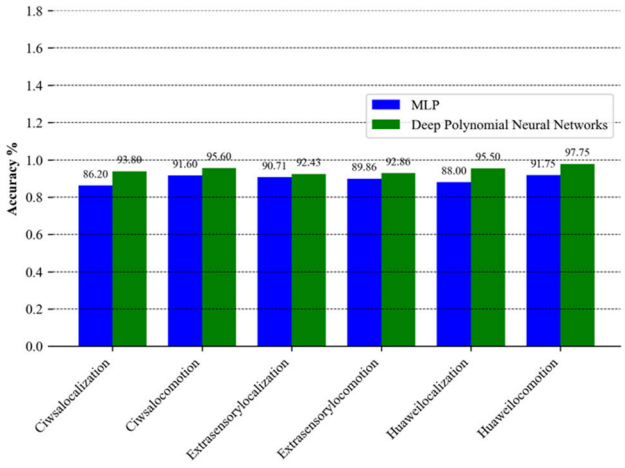


FIGURE 14. Comparisons between MLP and DPNN in terms of accuracy.

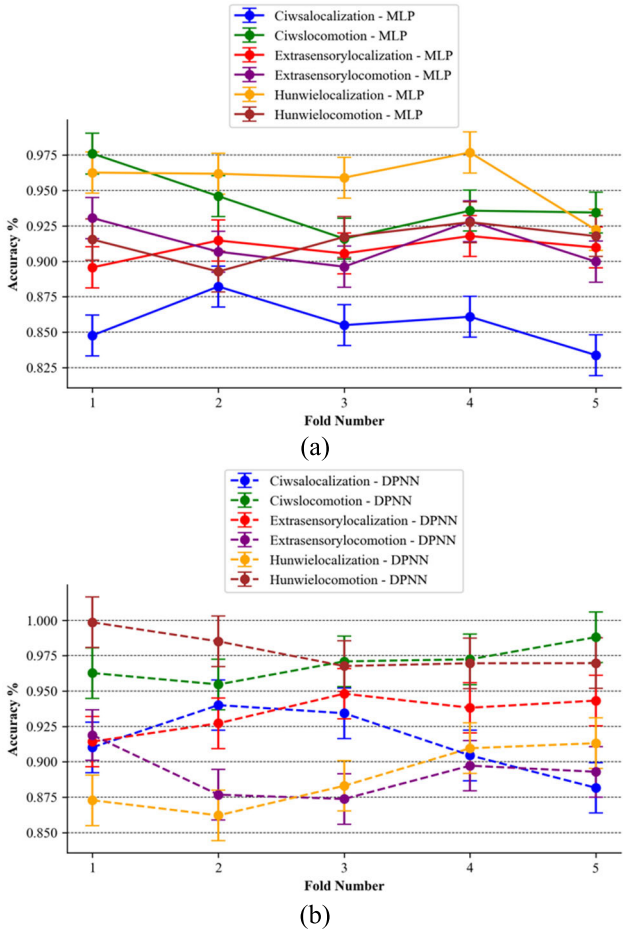


FIGURE 15. Cross-validation (a) MLP (b) DPNN model.

of a statistical analysis will generalize to an independent dataset. It's particularly useful in scenarios where the goal is to predict the outcome of a new data point not present in the analysis. In k-fold cross-validation, the original sample is randomly partitioned into *k*equal-sized subsamples. Of the *k* subsamples, a single subsample is retained as the validation data for testing the model, and the remaining *k-1* subsamples are used as training data. This process is repeated k times (the folds), with each of the k subsamples used exactly once as the validation data. By setting *k* to 5, we ensure that each fold serves as the validation set once while also being part of the training set four times. The results from the

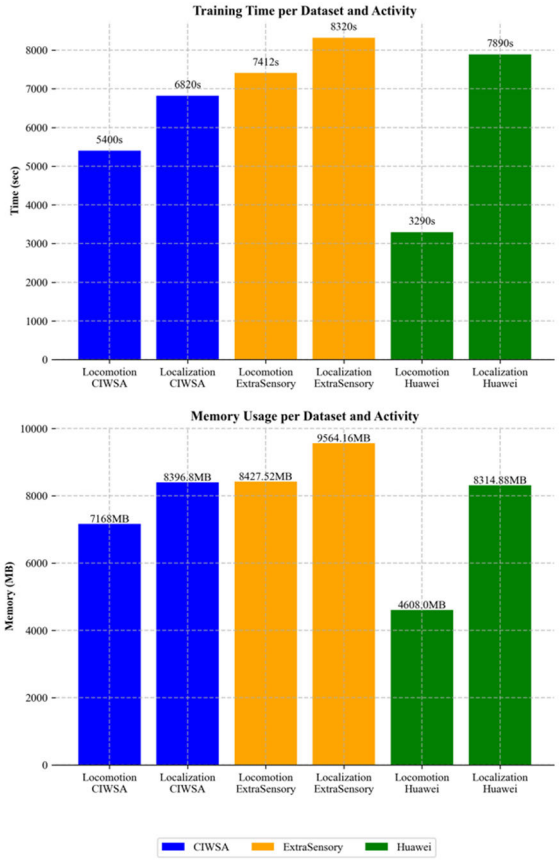


FIGURE 16. Computational requirements for DPNN model.

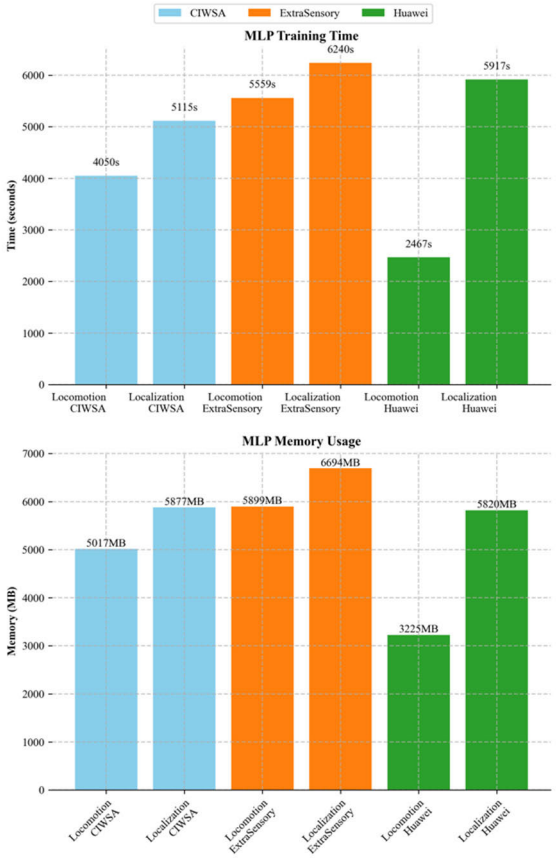


FIGURE 17. Computational requirements for MLP model.

folds can then be averaged to produce a single estimation. The advantage of this method is that all observations are used for both training and validation, and each observation is used for validation exactly once. The two charts represent the accuracy percentages across all five folds for each dataset and model combination. The variation in the accuracy across different folds indicates the model's sensitivity to the specific subset of the data it's trained on. By analyzing the spread of these results, we can infer the consistency of the model's performance. A model that performs consistently well across all folds is typically considered robust and well-generalized.

H. EXPERIMENT 7: COMPUTATIONAL COMPARISONS

1) TIME AND MEMORY CALCULATION FOR DPNN

We conducted a detailed comparison of the computational requirements for training Deep Polynomial Neural Network (DPNN) models across various datasets and corresponding activities. For the CIWSA dataset, the training time for locomotion activities consumed approximately 5400 seconds, with a memory usage of 7168 MB, whereas localization activities necessitated slightly more time at 6820 seconds and marginally increased memory consumption at 8396.8 MB. Moving to the ExtraSensory dataset, we notice an uptick in resource demands; locomotion activities required 7412 seconds of training time and 8427.52 MB of memory, with localization activities reaching 8320 seconds and 9564.16 MB respectively. The Huawei dataset exhibited contrasting characteristics, with locomotion activities completing in a brisk 3290 seconds, utilizing 4608 MB of memory, though localization activities showed a significant increase in both time and memory usage, taking 7890 seconds and consuming 8314.88 MB. In Fig. 16 the comparisons can be seen very clearly.

2) TIME AND MEMORY CALCULATION FOR MLP

Upon examining the Multilayer Perceptron (MLP) model, we observe a reduction in both training time and memory usage across the CIWSA, ExtraSensory, and Huawei datasets when compared to the Deep Polynomial Neural Network (DPNN) model. This is consistent with expectations, as MLP architectures are generally less complex than DPNNs. For the CIWSA dataset, the estimated training time for locomotion activities with MLP is approximately 3,240 seconds (54 minutes), and for localization activities, it's around 3,420 seconds (57 minutes). This suggests a more streamlined training process, which can be particularly advantageous when working with constrained computational resources or under time-sensitive conditions. In terms of memory usage, the MLP model is estimated to utilize around 4,224 MB for locomotion and 4,320 MB for localization activities within the CIWSA dataset. This reduced memory footprint can lead

TABLE 2. Comparison of recognition accuracy for the proposed system with other state-of-the-art systems over the continuous in-the-wild smartwatch activity, extrasensory, and huawei locomotion dataset.

Existing Systems	Extra-sensory (%)	SHL (%)	CIWSA (%)
Vaizman et al [58]	0.83	-	-
Asim et al [59]	0.87	-	-
Abdullah et al[60]	0.87	-	-
Sharma et al [61]	-	0.92	-
Akbari et al [62]	-	0.92	-
Brimacombe et al [63]	-	0.79	-
Alazeb et al [64]	-	-	90
Proposed system accuracy	0.92	0.95	0.93

to more cost-effective scaling, especially when deploying models in cloud environments or on devices with limited memory capacity. For the ExtraSensory dataset, the MLP's training time for locomotion activities is estimated at 6,247 seconds (approximately 104 minutes), while for localization activities, it's around 5,472 seconds (approximately 91 minutes). Memory usage estimates are 4,938 MB and 5,604 MB, respectively. The Huawei dataset shows the MLP's training time for locomotion at 2,574 seconds (about 43 minutes) and for localization at 4,734 seconds (about 79 minutes), with memory usage at 1,920 MB and 4,872 MB, respectively. In Fig. 17 the interpretation of the graph can be seen.

I. COMPARISONS WITH STATE-OF-THE-ART

The authors in [58], tackle the challenge of recognizing behavioral context using sensor data collected in the wild, which often includes unbalanced and incomplete data. They propose a multiple-layer perceptron (MLP) model that simultaneously predicts multiple context labels from multi-modal sensors. The MLP handles unbalanced data with instance-weighting and remains robust to missing sensors through sensor dropout. Evaluated on the ExtraSensory Dataset, the MLP outperforms previous models by sharing hidden representations among all labels, enhancing recognition accuracy even with fewer parameters. In [60], the authors introduce HAR-GCCN, a deep graph convolutional neural network (CNN) model designed to leverage correlations between chronologically adjacent sensor measurements to predict activity labels. The model uses a training strategy that enforces the prediction of missing activity labels by utilizing known ones. HAR-GCCN demonstrates superior performance compared to baseline methods, achieving an improvement in classification accuracy by approximately 25% to 68% across different datasets. The paper in [61] presents a method for early detection of transportation modes, balancing accuracy and timeliness using partially observed sensory time series data. The authors develop a hybrid deep learning classifier that combines convolutional neural networks, recurrent neural networks, and deep neural networks. A decision policy on top of the classifier predicts transportation modes efficiently. The proposed model, evaluated on two public datasets, shows good performance in accuracy. The article in [62] presents a hierarchical search algorithm that quickly and accurately identifies transitions in activity recognition models. It works with various window lengths and uses a new 2D signal input structure for 2D CNNs to capture inter-axes correlations better. The method improves the F1-score by 28% over fixed-size windowing and is 20 times faster than a basic search. The system in [64] uses a deep neural decision forest classifier for human activity recognition and localization. It starts by denoising the input signal with a Butterworth filter and segments it using a Hamming window. Features are extracted through activity recognition and localization, followed by recursive feature elimination and data augmentation using a genetic algorithm. Tested on the ExtraSensory and Sussex-Huawei Locomotion datasets, the system achieves good accuracy.

V. CONCLUSION

In this study, we have presented a comprehensive system designed to recognize human locomotion and location-based activities, utilizing data from multiple sensors. Our system fuses machine learning techniques, specifically Multilayer Perceptron (MLP), with advanced Deep Polynomial Neural Network (DPNN) methodologies to achieve high accuracy in activity recognition. Through rigorous computational analysis, we've demonstrated that while the MLP model excels in environments with limited computational resources, offering a quicker training process and reduced memory usage, the DPNN model surpasses in accuracy, making it suitable for complex pattern recognition tasks. The practical application of our system spans a wide range of fields, from healthcare, where it could be used for patient monitoring, to smart homes, enhancing the interactivity and responsiveness of home automation systems. Moreover, the system's adaptability makes it excellent for integration into wearable devices, providing real-time analytics for fitness tracking and potentially augmenting virtual reality experiences.

However, our system is not without limitations. The DPNN model, while accurate, requires significant computational resources, which may not be readily available in all settings. This limitation becomes particularly pronounced in real-time applications where rapid data processing is crucial. Moreover, the models' performance is heavily dependent on the quality and diversity of the training data, which could present challenges in scenarios where data collection is constrained by privacy concerns or technical limitations.

REFERENCES

[1] M. Batool, A. Jalal, and K. Kim, "Telemonitoring of daily activity using accelerometer and gyroscope in smart home environments," *J. Electr. Eng. Technol.*, vol. 15, no. 6, pp. 2801–2809, Nov. 2020.

[2] X. Xu, Z. Lin, X. Li, C. Shang, and Q. Shen, "Multi-objective robust optimisation model for MDVRPLS in refined oil distribution," *Int. J. Prod. Res.*, vol. 60, no. 22, pp. 6772–6792, Nov. 2022, doi: 10.1080/00207543.2021.1887534.

[3] N. Wang, J. Chen, W. Chen, Z. Shi, H. Yang, P. Liu, X. Wei, X. Dong, C. Wang, L. Mao, and X. Li, "The effectiveness of case management for cancer patients: An umbrella review," *BMC Health Services Res.*, vol. 22, no. 1, Oct. 2022, Art. no. 1247.

[4] U. Azmat and A. Jalal, "Smartphone inertial sensors for human locomotion activity recognition based on template matching and codebook generation," in *Proc. Int. Conf. Commun. Technol. (ComTech)*, Rawalpindi, Pakistan, Sep. 2021, pp. 109–114.

[5] C. Zhang, H. Ge, S. Zhang, D. Liu, Z. Jiang, C. Lan, L. Li, H. Feng, and R. Hu, "Hematoma evacuation via image-guided para-corticospinal tract approach in patients with spontaneous intracerebral hemorrhage," *Neurol. Therapy*, vol. 10, no. 2, pp. 1001–1013, Dec. 2021, doi: 10.1007/s40120-021-00279-8.

[6] Y. Di, R. Li, H. Tian, J. Guo, B. Shi, Z. Wang, K. Yan, and Y. Liu, "A maneuvering target tracking based on fastIMM-extended Viterbi algorithm," *Neural Comput. Appl.*, pp. 1–10, Oct. 2023, doi: 10.1007/s00521-023-09039-1.

[7] R. Mafrur, I. G. D. Nugraha, and D. Choi, "Modeling and discovering human behavior from smartphone sensing life-log data for identification purpose," *Hum.-Centric Comput. Inf. Sci.*, vol. 5, no. 1, p. 31, Dec. 2015, doi: 10.1186/s13673-015-0049-7.

[8] Y. Chen and C. Shen, "Performance analysis of smartphone-sensor behavior for human activity recognition," *IEEE Access*, vol. 5, pp. 3095–3110, 2017, doi: 10.1109/ACCESS.2017.2676168.

[9] S. Zhao, W. Liang, K. Wang, L. Ren, Z. Qian, G. Chen, X. Lu, D. Zhao, X. Wang, and L. Ren, "A multiaxial bionic ankle based on series elastic actuation with a parallel spring," *IEEE Trans. Ind. Electron.*, vol. 71, no. 7, pp. 7498–7510, Jul. 2024, doi: 10.1109/TIE.2023.3310041.

[10] K. Wang, A. Boonpratatong, W. Chen, L. Ren, G. Wei, Z. Qian, X. Lu, and D. Zhao, "The fundamental property of human leg during walking: Linearity and nonlinearity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 4871–4881, 2023, doi: 10.1109/TNSRE.2023.3339801.

[11] X. Hou, L. Zhang, Y. Su, G. Gao, Y. Liu, Z. Na, Q. Xu, T. Ding, L. Xiao, L. Li, and T. Chen, "A space crawling robotic bio-paw (SCRBP) enabled by triboelectric sensors for surface identification," *Nano Energy*, vol. 105, Jan. 2023, Art. no. 108013, doi: 10.1016/j.nanoen.2022.108013.

[12] L. Fan, D. P. Haghighi, Y. Zhang, A. R. M. Forkan, and P. P. Jayaraman, "Context-aware human activity recognition (CA-HAR) using smartphone built-in sensors," in *Proc. Int. Conf. Adv. Mobile Comput. Multimedia Intell.*, vol. 13634. Berlin, Germany: Springer, 2022, pp. 108–121.

[13] Z. Gao, D. Liu, K. Huang, and Y. Huang, "Context-aware human activity and smartphone position-mining with motion sensors," *Remote Sens.*, vol. 11, no. 21, p. 2531, Oct. 2019.

[14] Y.-L. Hsu, S.-C. Yang, H.-C. Chang, and H.-C. Lai, "Human daily and sport activity recognition using a wearable inertial sensor network," *IEEE Access*, vol. 6, pp. 31715–31728, 2018.

[15] S. Chung, J. Lim, K. J. Noh, G. Kim, and H. Jeong, "Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning," *Sensors*, vol. 19, no. 7, p. 1716, Apr. 2019.

[16] Y. Zhao, S. Guo, Z. Chen, Q. Shen, Z. Meng, and H. Xu, "Marfusion: An attention-based multimodal fusion model for human activity recognition in real-world scenarios," *Appl. Sci.*, vol. 12, no. 11, p. 5408, May 2022.

[17] Q. Liu, H. Yuan, R. Hamzaoui, H. Su, J. Hou, and H. Yang, "Reduced reference perceptual quality model with application to rate control for video-based point cloud compression," *IEEE Trans. Image Process.*, vol. 30, pp. 6623–6636, 2021, doi: 10.1109/TIP.2021.3096060.

[18] Z. Gu, T. He, Z. Wang, and Y. Xu, "Device-free human activity recognition based on dual-channel transformer using WiFi signals," *Wireless Commun. Mobile Comput.*, vol. 2022, pp. 1–14, Jun. 2022, doi: 10.1155/2022/4598460.

[19] T. Guo, H. Yuan, L. Wang, and T. Wang, "Rate-distortion optimized quantization for geometry-based point cloud compression," *J. Electron. Imag.*, vol. 32, no. 1, p. 13047, Feb. 2023, doi: 10.1117/1.jei.32.1.013047.

[20] X. Zhang and S. Jiang, "Application of Fourier transform and Butterworth filter in signal denoising," in *Proc. 6th Int. Conf. Intell. Comput. Signal Process. (ICSP)*, Xi'an, China, Apr. 2021, pp. 1277–1281, doi: 10.1109/ICSP51882.2021.9408933.

[21] S. Ma, Y. Chen, S. Yang, S. Liu, L. Tang, B. Li, and Y. Li, "The autonomous pipeline navigation of a cockroach bio-robot with enhanced walking stimuli," *Cyborg Bionic Syst.*, vol. 4, Jan. 2023, Art. no. 67, doi: 10.34133/cbsystems.0067.

[22] P. L. Chithra and R. Aparna, "Performance analysis of windowing techniques in automatic speech signal segmentation," *Indian J. Sci. Technol.*, vol. 8, no. 29, pp. 1–7, Nov. 2015.

[23] J. Luo, G. Wang, G. Li, and G. Pesce, "Transport infrastructure connectivity and conflict resolution: A machine learning analysis," *Neural Comput. Appl.*, vol. 34, no. 9, pp. 6585–6601, May 2022, doi: 10.1007/s00521-021-06015-5.

[24] S. Wei and Z. Wu, "The application of wearable sensors and machine learning algorithms in rehabilitation training: A systematic review," *Sensors*, vol. 23, no. 18, p. 7667, Sep. 2023, doi: 10.3390/s23187667.

[25] X. Shen, H. Jiang, D. Liu, K. Yang, F. Deng, J. C. S. Lui, J. Liu, S. Dustdar, and J. Luo, "PupilRec: Leveraging pupil morphology for recommending on smartphones," *IEEE Internet Things J.*, vol. 9, no. 17, pp. 15538–15553, Sep. 2022, doi: 10.1109/JIOT.2022.3181607.

[26] H. Jiang, S. Chen, Z. Xiao, J. Hu, J. Liu, and S. Dustdar, "Pa-count: Passenger counting in vehicles using Wi-Fi signals," *IEEE Trans. Mobile Comput.*, vol. 23, no. 4, pp. 2684–2697, Mar. 2023, doi: 10.1109/TMC.2023.3263229.

[27] S. A. Rizwan, A. Jalal, M. Gochoo, and K. Kim, "Robust active shape model via hierarchical feature extraction with SFS-optimized convolution neural network for invariant human age classification," *Electronics*, vol. 10, no. 4, p. 465, Feb. 2021.

[28] H. Jiang, M. Wang, P. Zhao, Z. Xiao, and S. Dustdar, "A utility-aware general framework with quantifiable privacy preservation for destination prediction in LBSs," *IEEE/ACM Trans. Netw.*, vol. 29, no. 5, pp. 2228–2241, Oct. 2021, doi: 10.1109/TNET.2021.3084251.

[29] J. Zhu, R. San-Segundo, and J. M. Pardo, "Feature extraction for robust physical activity recognition," *Hum.-Centric Comput. Inf. Sci.*, vol. 7, no. 1, p. 16, Dec. 2017, doi: 10.1186/s13673-017-0097-2.

[30] W. Zheng, L. Lin, X. Wu, and X. Chen, "An empirical study on correlations between deep neural network fairness and neuron coverage criteria," *IEEE Trans. Softw. Eng.*, vol. 50, no. 3, pp. 391–412, Mar. 2024, doi: 10.1109/TSE.2023.3349001.

[31] W. Wu, H. Zhu, S. Yu, and J. Shi, "Stereo matching with fusing adaptive support weights," *IEEE Access*, vol. 7, pp. 61960–61974, 2019, doi: 10.1109/ACCESS.2019.2916035.

[32] A. Jalal, S. Kamal, and D. Kim, "A depth video-based human detection and activity recognition using multi-features and embedded hidden Markov models for health care monitoring systems," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 4, no. 4, p. 54, 2017.

[33] B. Wang, W. Zheng, R. Wang, S. Lu, L. Yin, L. Wang, Z. Yin, and X. Chen, "Stacked noise reduction auto encoder–OCEAN: A novel personalized recommendation model enhanced," *Systems*, vol. 12, no. 6, p. 188, May 2024, doi: 10.3390/systems12060188.

[34] F. Song, Y. Liu, D. Shen, L. Li, and J. Tan, "Learning control for motion coordination in wafer scanners: Toward gain adaptation," *IEEE Trans. Ind. Electron.*, vol. 69, no. 12, pp. 13428–13438, Dec. 2022, doi: 10.1109/tie.2022.3142428.

[35] R. Cui, G. Hua, A. Zhu, J. Wu, and H. Liu, "Hard sample mining and learning for skeleton-based human action recognition and identification," *IEEE Access*, vol. 7, pp. 8245–8257, 2019.

[36] Z. Wu, H. Zhu, L. He, Q. Zhao, J. Shi, and W. Wu, "Real-time stereo matching with high accuracy via spatial attention-guided upsampling," *Appl. Intell.*, vol. 53, no. 20, pp. 24253–24274, Oct. 2023, doi: 10.1007/s10489-023-04646-w.

[37] J. J. Peng, X. G. Chen, X. K. Wang, J. Q. Wang, Q. Q. Long, and L. J. Yin, "Picture fuzzy decision-making theories and methodologies: A systematic review," *Int. J. Syst. Sci.*, vol. 54, no. 13, pp. 2663–2675, Oct. 2023, doi: 10.1080/00207721.2023.2241961.

[38] Y. Shi, Y. Tian, Y. Wang, and T. Huang, "Sequential deep trajectory descriptor for action recognition with three-stream CNN," *IEEE Trans. Multimedia*, vol. 19, no. 7, pp. 1510–1520, Jul. 2017.

[39] H. He, X. Li, P. Chen, J. Chen, M. Liu, and L. Wu, "Efficiently localizing system anomalies for cloud infrastructures: A novel dynamic graph transformer based parallel framework," *J. Cloud Comput.*, vol. 13, no. 1, p. 115, Jun. 2024, doi: 10.1186/s13677-024-00677-x.

[40] Y. Sun, Z. Peng, J. Hu, and B. K. Ghosh, "Event-triggered critic learning impedance control of lower limb exoskeleton robots in interactive environments," *Neurocomputing*, vol. 564, Jan. 2024, Art. no. 126963, doi: 10.1016/j.neucom.2023.126963.

[41] S. Ali, M. Hassan, J. Y. Kim, M. I. Farid, M. Sanaullah, and H. Mufti, "FF-PCA-LDA: Intelligent feature fusion based PCA-LDA classification system for plant leaf diseases," *Appl. Sci.*, vol. 12, no. 7, p. 3514, Mar. 2022, doi: 10.3390/app12073514.

[42] Z.-Y. Wang, "A novel face recognition method based on feature fusion of global and local structure," in *Proc. Int. Conf. Comput. Eng. Artif. Intell. (ICCEAI)*, Shijiazhuang, China, Jul. 2022, pp. 533–538, doi: 10.1109/ICCEAI55464.2022.00116.

[43] P. Zhou, J. Qi, A. Duan, S. Huo, Z. Wu, and D. Navarro-Alarcon, "Imitating tool-based garment folding from a single visual observation using hand-object graph dynamics," *IEEE Trans. Ind. Informat.*, vol. 20, no. 4, pp. 6245–6256, Apr. 2024, doi: 10.1109/TII.2023.3342895.

[44] J. Chen, Y. Sun, and S. Sun, "Improving human activity recognition performance by data fusion and feature engineering," *Sensors*, vol. 21, no. 3, p. 692, Jan. 2021, doi: 10.3390/s21030692.

[45] X. Bai, Y. He, and M. Xu, "Low-thrust reconfiguration strategy and optimization for formation flying using Jordan normal form," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 5, pp. 3279–3295, Oct. 2021, doi: 10.1109/TAES.2021.3074204.

[46] P. P. Ariza-Colpas, E. Vicario, A. I. Oviedo-Carrascal, S. B. Aziz, M. A. Piñeres-Melo, A. Quintero-Linero, and F. Patara, "Human activity recognition data analysis: History, evolutions, and new trends," *Sensors*, vol. 22, no. 9, p. 3401, Apr. 2022, doi: 10.3390/s22093401.

[47] X. Wang, R. Zhang, Y. Miao, S. Wang, and Y. Zhang, "PI^2-based adaptive impedance control for gait adaption of lower limb exoskeleton," *IEEE/ASME Trans. Mechatronics*, vol. 29, no. 2, pp. 1–11, Mar. 2024, doi: 10.1109/TMECH.2024.3370954.

[48] X. Ruan, S. Wang, and S. Liu, "Multi-structure feature fusion for face recognition based on multi-resolution exaction," in *Proc. 9th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Datong, China, Oct. 2016, pp. 291–296, doi: 10.1109/CISP-BMEI.2016.7852724.

[49] Y. Chen, N. Li, D. Zhu, C. C. Zhou, Z. Hu, Y. Bai, and J. Yan, "BEVSOC: Self-supervised contrastive learning for calibration-free BEV 3-D object detection," *IEEE Internet Things J.*, vol. 11, no. 12, pp. 22167–22182, Jun. 2024, doi: 10.1109/JIOT.2024.3379471.

[50] D. Cai, R. Li, Z. Hu, J. Lu, S. Li, and Y. Zhao, "A comprehensive overview of core modules in visual SLAM framework," *Neurocomputing*, vol. 590, no. 5, 2024, Art. no. 127760, doi: 10.1016/j.neucom.2024.127760.

[51] G. G. Chrysos, S. Moschoglou, G. Bouritsas, J. Deng, Y. Panagakis, and S. Zafeiriou, "Deep polynomial neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4021–4034, Aug. 2022, doi: 10.1109/TPAMI.2021.3058891.

[52] S. Yang, Y. Jin, J. Lei, and S. Zhang, "Multi-directional guidance network for fine-grained visual classification," *Vis. Comput.*, pp. 1–12, Jan. 2024, doi: 10.1007/s00371-023-03226-w.

[53] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 4396–4405, doi: 10.1109/CVPR.2019.00453.

[54] C. Liu, K. Xie, T. Wu, C. Ma, and T. Ma, "Distributed neural tensor completion for network monitoring data recovery," *Inf. Sci.*, vol. 662, Mar. 2024, Art. no. 120259, doi: 10.1016/j.ins.2024.120259.

[55] L. Wang, H. Gjoreski, M. Ciliberto, P. Lago, K. Murao, T. Okita, and D. Roggen, "Three-year review of the 2018–2020 SHL challenge on transportation and locomotion mode recognition from mobile sensors," *Frontiers Comput. Sci.*, vol. 3, Sep. 2021, Art. no. 713719, doi: 10.3389/fcomp.2021.713719.

[56] Y. Vaizman, K. Ellis, G. Lanckriet, and N. Weibel, "ExtraSensory app: Data collection in-the-wild with rich user interface to self-report behavior," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, New York, NY, USA, Apr. 2018, pp. 1–12, doi: 10.1145/3173574.3174128.

[57] Y. Vaizman, K. Ellis, and G. Lanckriet, "Recognizing detailed human context in the wild from smartphones and smartwatches," *IEEE Pervasive Comput.*, vol. 16, no. 4, pp. 62–74, Oct. 2017, doi: 10.1109/MPRV.2017.3971131.

[58] Y. Vaizman, N. Weibel, and G. Lanckriet, "Context recognition in-the-wild: Unified model for multi-modal sensors and multi-label classification," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 4, p. 168, 2018.

[59] Y. Asim, M. A. Azam, M. Ehatisham-ul-Haq, U. Naeem, and A. Khalid, "Context-aware human activity recognition (CAHAR) in-the-wild using smartphone accelerometer," *IEEE Sensors J.*, vol. 20, no. 8, pp. 4361–4371, Apr. 2020.

[60] A. Mohamed, F. Lejarza, S. Cahail, C. Claudel, and E. Thomaz, "HAR-GCNN: Deep graph CNNs for human activity recognition from highly unlabeled mobile sensor data," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops Affiliated Events (PerCom Workshops)*, Mar. 2022, pp. 124–126.

[61] A. Sharma, S. K. Singh, S. S. Udmale, A. K. Singh, and R. Singh, "Early transportation mode detection using smartphone sensing data," *IEEE Sensors J.*, vol. 21, no. 14, pp. 15651–15659, Jul. 2021, doi: 10.1109/JSEN.2020.3009312.

[62] A. Akbari and R. Jafari, "Transition-aware detection of modes of locomotion and transportation through hierarchical segmentation," *IEEE Sensors J.*, vol. 21, no. 3, pp. 3301–3313, Feb. 2021.

[63] O. Brimacombe, L. C. González, and J. Wahlström, "Smartphone-based CO_2e emission estimation using transportation mode classification," *IEEE Access*, vol. 11, pp. 54782–54794, 2023.

[64] A. Alazeb, U. Azmat, N. Al Mudawi, A. Alshahrani, S. S. Alotaibi, N. A. Almujally, and A. Jalal, "Intelligent localization and deep human activity recognition through IoT devices," *Sensors*, vol. 23, no. 17, p. 7363, Aug. 2023, doi: 10.3390/s23177363.

DANYAL KHAN received the M.S. degree in computer science from Air University, Islamabad. He is currently a Research Assistant with Air University. His research interests include machine learning, deep learning, camera and sensor-based gesture recognition, and virtual reality.

ABDULLAH ALSHAHRANI received the degree in computer science from King Khalid University, in 2008, the M.Sc. degree in computer science from La Trobe University, Melbourne, Australia, in 2010, and the Ph.D. degree from The Catholic University of America, USA, in 2018. He is currently an Assistant Professor with the Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Saudi Arabia. His research interests include wireless sensor networks, network security, parallel computing, smart home systems, the IoTs, and data science.

ABRAR ALMJALLY received the bachelor's degree from King Saud University, in 2007, the master's degree in information management security from Syracuse University, USA, in 2014, and the Ph.D. degree in informatics from the University of Sussex, Brighton, U.K. She is currently a Faculty Member with the Information Technology Department, Imam Mohammad Ibn Saud Islamic University (IMSIU). Her research interests include computing education, the Internet of Things, education technology, machine learning, and artificial intelligence.

NAIF AL MUDAWI received the master's degree in computer science from Australian La Trobe University, in 2011, and the Ph.D. degree from the College of Engineering and Informatics, University of Sussex, Brighton, U.K., in 2018. He is currently an Assistant Professor with the Department of Computer Science and Information Systems, Najran University. He has published research and scientific articles in many prestigious journals in various disciplines of computer science. During his academic journey to obtain the master's degree, he was a member of the Australian Computer Science Committee.

ASAAD ALGARNI received the Ph.D. degree in software engineering from North Dakota State University, USA. He is currently an Assistant Professor with the Department of Computer Sciences, College of Computing and Information Technology, Northern Border University, Kingdom of Saudi Arabia. His research interests include revolve around software engineering, computer vision applications, and machine learning.

KHALED ALNOWAISER received the Ph.D. degree in computer science from Glasgow University, Scotland. He is currently an Assistant Professor with the Computer Engineering Department, Prince Sattam Bin Abdulaziz University, Saudi Arabia. His research interests include computer vision, optimization techniques, and performance enhancement.

AHMAD JALAL received the Ph.D. degree from the Department of Biomedical Engineering, Kyung Hee University, Republic of Korea. He is currently an Associate Professor with the Department of Computer Science and Engineering, Air University, Pakistan. He is a Postdoctoral Research Fellowship with POSTECH. His research interests include multimedia contents and artificial intelligence.