Inteligencia Artificial 28(75), 298-314
doi: 10.4114/intartif.vol28iss75pp298-314

INTELIGENCIA ARTIFICIAL

http://journal.iberamia.org/

IoT-Based System for Human Localization Activity Recognition Using Hybrid Deep Learning Techniques

Mohammed Chachan Younis[1,A], Zaid Jafar Fadil[2], Baydaa Sulaiman Bahnam[3]

[1][2][3] College of Computer Sciences and Mathematics, University of Mosul, Mosul, 41002, Iraq
[1][3] Department of Artificial Intelligence
[2] Department of Networks
[A]mohammed.c.y@uomosul.edu.iq

Abstract Indoor positioning and navigation is an emerging field where accurate location identification and activity recognition with precision are important factors. Due to the emergence of handheld devices with location enabled and their importance in smart homes, industries, health monitoring, and security surveillance, human activity localization is also of great concern and importance. Keeping in mind the importance of accuracy with precision, we have proposed an IoT-based solution for human activity recognition using a hybrid deep learning approach in this article. In our proposed model, we have integrated Convolutional Neural Networks (CNNs) with advanced feature extraction, with an added feature of optimization for enhanced accuracy and precision. Our proposed hybrid model successfully classifies human activities such as "AT SCHOOL," "LOC Home," "Indoor," and "Outdoor" using IoT-based sensor data received from multiple stations. The accuracy of our proposed deep learning hybrid model is 96%, compared to existing techniques for human activity recognition such as Deep Neural Decision Forest (89%), HAR-graph CNN (87%), and Random Forest (87%), with enhanced precision, recall, and F1 score, respectively. For data augmentation and optimization, we have used SMOTE and Yeo-Johnson to address the issues of class imbalance and feature distribution, respectively. Moreover, 5-fold cross-validation is used to ensure the robustness and efficiency of the proposed model for localizing human activities with enhanced accuracy and recognition.

Keywords: Human Localization, Activity Recognition, SMOTE, Yeo-Johnson Transformation, Convolutional Neural Network (CNN).

1 Introduction

Due to the technological advancement in handheld devices, the Internet of Things (IoT) [1-3], smart phones, and other digital gadgets with Global Positioning System (GPS) enabled, and accelerometer provides location identification and activity monitoring in an outdoor environment. In the case of indoor environments, such as smart homes, health care monitoring, and industrial site where activity monitoring [4-5] with accurate location identification is crucial requires continuous GPS signals, dependency on dynamic environment, the accuracy of traditional approaches are not up to the mark due to frequent changes in sensors data transmission and reception. It is evident from the latest research that deep learning models provide higher accuracy and precision for human activity recognition and monitoring, especially in complex indoor environments [6-10]. However, existing deep learning models have issues such as class imbalance, errors in data, and high computational cost to address the large volume of data. To handle these issues in case of a complex and dynamic environment, we have proposed an IoT-based solution for human activity monitoring and navigation using Hybrid Deep Learning Models. Our proposed hybrid solution combines Convolutional Neural Networks (CNNs) [11-13] with advanced signal processing and optimization techniques. This will enhance classification. For noise reduction in data, the

ISSN: 1137-3601 (print), 1988-3064 (on-line)

Butterworth filter is used, and Hamming window-based segmentation and Synthetic Minority Over-sampling Technique (SMOTE) for adjustment to balance the data. For optimization, we integrated the Yeo-Johnson transformation for feature distribution [14] in the dataset and model generalizability. The performance of our proposed hybrid deep learning model is evaluated on the Extrasensory dataset with higher accuracy as compared to the existing models used for human activity monitoring and identification. Our proposed IoT-based deep learning hybrid model is effective in terms of computational complexity, handling in balance data, and robustness. The following are the contributions in this article.

1. Deep learning models require a large volume of datasets. In this article, we have used SMOTE, which is a statistical technique to address the issue of class imbalance, which is required for applying deep learning models so that the accuracy of human activity is enhanced.

2. Optimizes feature sets by way of the Yeo-Johnson transformation, which yields a normalized feature distribution that enhances the efficiency of machine learning models, especially for IoT-focused applications.

2 Related work

Due to recent advancements in multifunctional frameworks, activity prediction now faces extra challenges. The following table summarizes the key characteristics and limitations of various approaches.

Table 1: Comparative Analysis of Related Work.

Study	**Methodology**	**Limitations**	**Comparison with Proposed Hybrid-CNN**
Ordóñez et al. [15]	CNN + LSTM	Lack of data filtration tools leading to reduced performance with multimodal inputs	Hybrid-CNN integrates denoising and feature optimization for better multimodal performance
Research Effort [16]	Multiple IMUs + LSTM	Inadequate feature extraction and data filtering resulting in insufficient accuracy	Our model enhances feature extraction and filtering, leading to improved accuracy
Chavarriaga et al. [17]	Multimodal system using wearable and external sensors	Exclusion of optical sensors	Hybrid-CNN does not rely on specific sensors, making it more adaptable
Chung et al. [18]	Two-level supervised classifier + modified CRF-based classifier	Limited accuracy due to the absence of handcrafted features	Hybrid-CNN autonomously extracts optimal features, eliminating need for handcrafted features
Manokhin et al. [19]	Sensors on wrist and ankle + PCA + nonparametric weighted feature extraction	Limited sensor placement and reliance on wireless communication	Our model is sensor-independent and processes data efficiently without connection issues
A-Bassett et al. [20]	Sensory data to RGB images + multiscale classification features + channel-wise attention mechanism	Trained on small samples; concerns about generalizability and scalability	Hybrid-CNN trained on a larger dataset, ensuring better generalizability

Konak et al. [21]	Feature sets from accelerometer data + various classifiers	Limited dataset size; potential for improved performance with advanced models	Deep learning model improves feature learning and overall accuracy
Chetty et al. [22]	Inertial sensors in smartphones + feature ranking methods + various classifiers	Trained on a single dataset; potential limitations in real-time applications	Hybrid-CNN generalizes better across diverse datasets
Ehatisham-ul-Haq et al. [23]	Activity awareness using accelerometer data + Random Forest	Reliance on accelerometer data; potential misclassification in complex scenarios	Hybrid-CNN incorporates multiple sensor modalities for robust recognition
Cao et al. [24]	Group-based hierarchical structure + context awareness	Performance could be enhanced by incorporating additional sensors	Hybrid-CNN achieves high accuracy without requiring additional sensor integration
Khan et al. [25]	Utilized a Deep Polynomial Neural Network (DPNN) for human activity recognition and localization in IoT environments.	The complexity of the DPNN may lead to increased computational requirements, posing challenges for real-time applications.	Our model employs a hybrid deep learning approach combining CNN and LSTM, which enhances both spatial and temporal feature learning

3 Materials and methods

The proposed system recognizes human localization activity correctly. In the first, we apply a butter worth filters to remove unwanted noise. Secondly, the hamming windows technique is applied to segment large sequence data into small chunks. In the third step, different features were extracted. We observed that some activity has a smaller number of samples, which leads to an imbalanced dataset, so for this purpose, we employed the SMOTE data enhancing method to synthetically improve the diversity of samples. The augmented data is optimized using Yeo-Johnson optimization. Ultimately, the optimized features are sent to a convolutional classifier for the classification of different localization activities. Figure 1 shows the framework architecture.

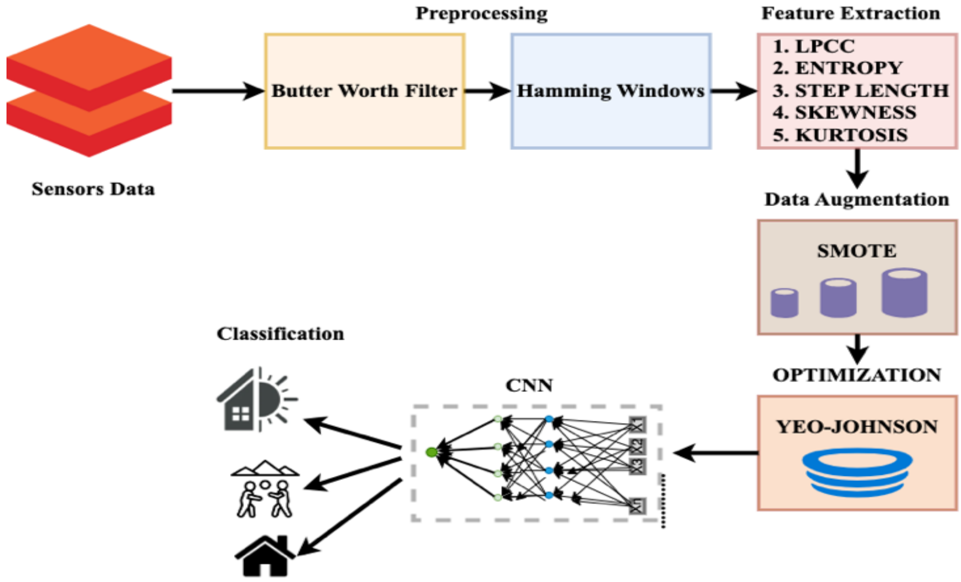

Figure 1. The proposed model’s framework.

3.1 Signal denoising

We applied the Butterworth filter [26] to denoise sensor data from accelerometers, gyroscopes, magnetometers, GPS, and sound sensors. The goal was to enhance the accuracy of recognizing localization activities, such as distinguishing between indoor and outdoor settings. The cutoff frequency is the threshold beyond which frequencies are attenuated.

In our case, f_c= 0.1Hz. The sampling frequency is the rate at which data samples are collected. We set f_c = 1.0Hz. The filter order determines the steepness of the filter's transition band. We used a second-order filter, n = 2. The other half of the sample frequency is the Nyquist frequency:

$$f_N = \frac{f_s}{2} \tag{1}$$

The normalized cutoff frequency is the threshold frequency divided by the Nyquist frequency:

$$\omega_c = \frac{f_c}{f_n} \tag{2}$$

The Butterworth filter is characterized by its transfer function $H\ (s)$, which in the online space is derived using the bilinear transform. For a low-pass Butterworth filter, the design involves determining the filter coefficients b and a such that the filter meets the desired specifications. Given the normalized cutoff frequency ω_c and the filter order n, the filter coefficients are obtained using the butter function in the SciPy signal library. The filter is applied to the data using the filter function after the filter coefficients have been established. This function performs forward and backward filtering to ensure zero phase distortion, yielding the filtered output y. In Figure 2 the noisy vs filtered signal can be seen.

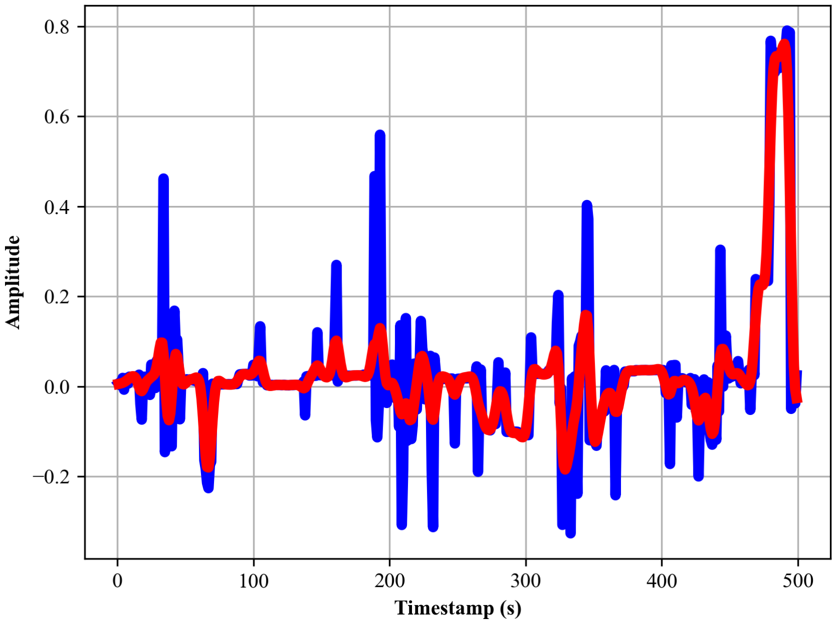

Figure 2. Noisy vs Filtered Accelerometer Signal.

3.2 Hamming Windows

After filtering the data with the Butterworth filter, we apply Hamming windows to smooth the data further and reduce spectral leakage. The Hamming window [27] is defined by the function:

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \quad (3)$$

where n ranges from 0 to $N - 1$ and N is the total number of data points in the signal. Using the filtrated information y, the Hamming window is applied as follows:

$$y_{\text{windowed}} = y \cdot w \quad (4)$$

where y_{windowed} is the final output after windowing, y is the filtered data, and w is the Hamming window.

3.3 Feature Extraction

The process of feature extraction is crucial to machine learning. In our study, we proposed a hybrid deep learning model. We extract different features, including LPCC (Linear Predictive Coefficient), MFCC (Mel Frequency Cepstral Coefficients), skewness, kurtosis, and phase angle.

3.3.1 LPCC (Linear Predictive Coefficient)

In human activity recognition, LPCC helps in analyzing motion-related signals from wearable sensors, effectively distinguishing between dynamic (e.g., walking, running) and static (e.g., sitting, standing) activities. In signal processing and speech, LPCCs [28] are a feature extraction technique that is frequently employed. It entails computing coefficients that symbolize a signal's spectral envelope. It is from the Linear Predictive Coding (LPC) model that LPCC coefficients are obtained. A linear predictive model that can forecast the future sample of a signal based on previous samples is developed using LPC, a method for estimating its parameters. From these LPC coefficients, the LPCC coefficients are then calculated. LPC assumes that the signal $x\ [n]$ may be roughly represented as a linear mixture of its previous P samples:

$$x[n] \approx -\sum_{k=1}^{p} a_k x[n-k] \quad (5)$$

where a_k are the LPC coefficients and P denotes the model's order. Next, the LPC coefficients a_k are used to calculate the cepstral coefficients. The relationship between LPC and cepstral coefficients $c\ n$ c n is given by:

$$c_n = \begin{cases} a_n + \sum_{k=1}^{n-1} \frac{k}{n} c_k a_{n-k}, \text{ for } n \geq 1 \\ \log(\sigma^2), \text{ for } n = 0 \end{cases} \quad (6)$$

where σ^2 is the prediction error variance. The LPCC plotted for different activities can be seen in Figure 3.

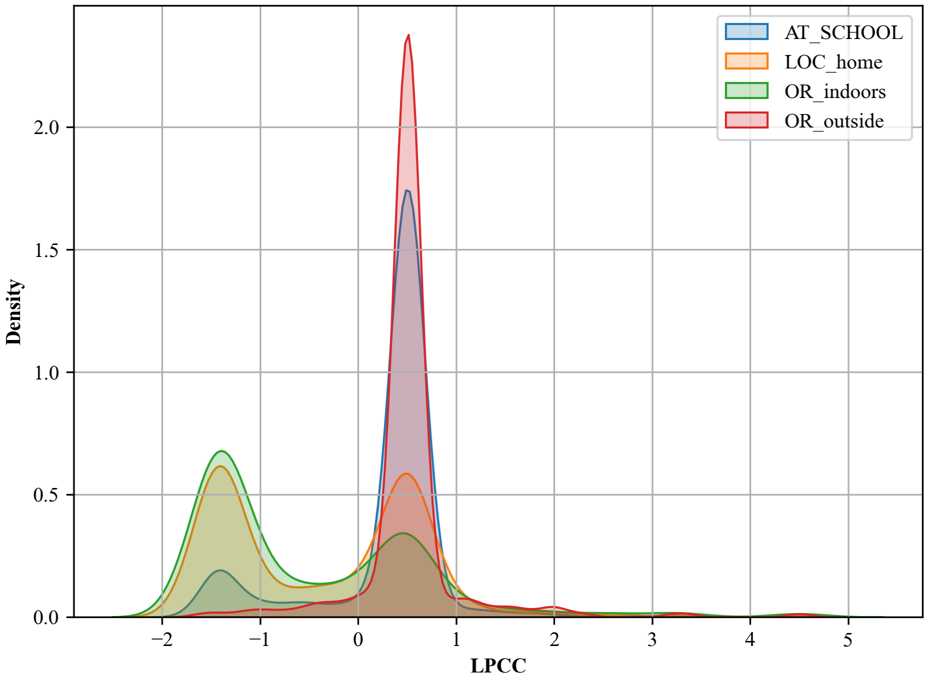

Figure 3. LPCC calculated for different localization activities.

3.3.2 MFCC (Mel Frequency Cepstral Coefficients)

This feature is crucial in HAR as it improves the differentiation of subtle movement variations, such as differentiating between walking at a normal pace vs. brisk walking. Implementing a transformation linear cosine of the exponential power spectrum on a complex mel frequency scale, MFCCs [29] depict the immediate power spectrum of a sound signal. Compared to the vertically spaced frequency ranges utilized in the typical cepstrum, MFCCs are intended to more precisely mimic the response of the human auditory system. They are widely employed in many different applications, including speaker identification, music genre classification, and voice recognition. To amplify high frequencies in the signal, a pre-emphasis filter is initially used. By doing this, the signal's spectral flatness and signal-to-noise ratio are both increased. Next, the signal is divided into smaller frames that coincide. This is due to the assumption that the signal will remain steady for brief intervals. To reduce discontinuities at the start and finish of each frame, a window function is multiplied by each frame. Every frame undergoes an FFT to modify the signal from the temporal domain into the frequency domain. The frequency spectrum of each frame is provided in this step. A filter bank is used to transfer each frame's power spectrum onto the mel scale. It is a sensory scale of sounds that are perceived by listeners as being equally spaced from one another. To change multiplicative components into additive components, one takes the mel spectrum logarithm. The log-mel spectrum is processed using DCT to get the mel-frequency cepstral coefficients. The most important information is packed into the first few coefficients in this stage, which decorrelates the coefficients. To provide a reliable feature vector for classification tasks, the mean of the MFCCs across all frames is finally calculated. The MFCC graph is shown in Figure 4.

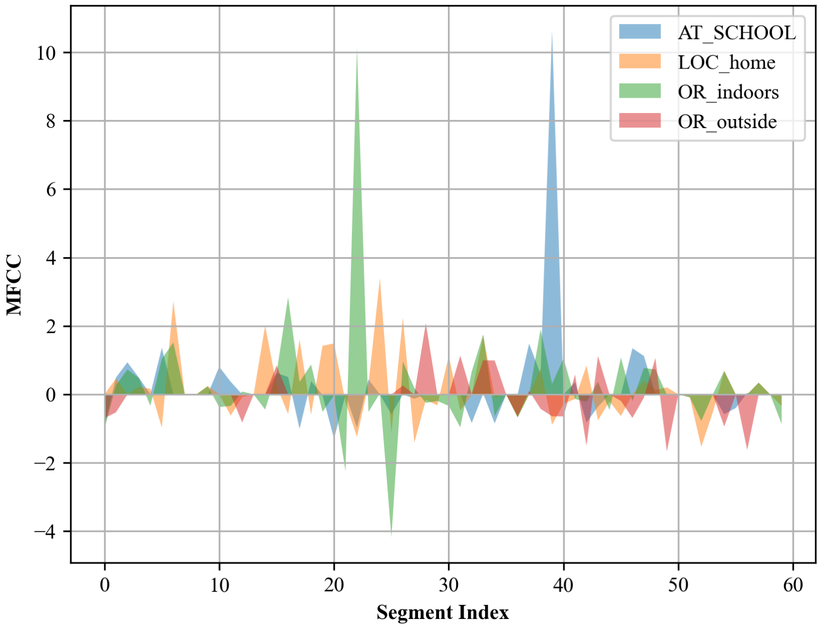

Figure 4. MFCC calculated for different localization activities

3.3.3 Skewness

A statistical metric known as skewness [30] quantifies how asymmetrically a data distribution is around its mean. It gives the degree and direction of the data distribution's asymmetry. In signal processing, skewness is employed to analyze the distribution of signal values within a given time series. The basic concept is that positively skew distributions have long right tails and negatively skew distributions have long left tails. This suggests that the variation with zero skewness is balanced and that most data points are centered on the right side, with fewer data points stretching to the left, meaning the data points are evenly distributed around the mean. Skewness mathematical form is given below:

$$\text{Skewness}=\frac{x}{(x-1)(x-2)}\sum_{i=1}^{x}\left(\frac{z_i-\bar{z}}{s}\right)^3 \qquad (7)$$

where x is the data points, z_i in each data point, $\bar{z}$ is the mean data point and s is the standard deviation. The formula standardizes the data points by subtracting the mean and dividing by the standard deviation, then cubes the result and averages it across all data points. The skewness value indicates the direction and degree of skew. We calculated skewness for different activities presented in Figure 5.

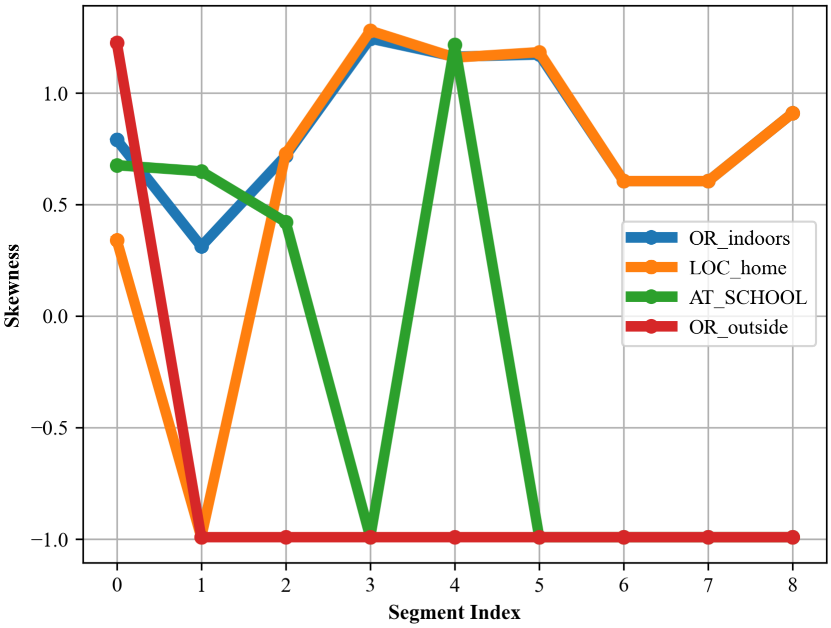

Figure 5. Skewness calculated for different localization activities

3.3.4 Kurtosis

It sheds light on the distribution's form, with special attention to the peak and tails. A distribution with a high kurtosis [31] has fat tails and a sharp peak. In comparison to a standard distribution, this shows that data points have a greater concentration around the mean and have more extreme values, or outliers. A distribution with thin tails and a flatter apex is characterized by low kurtosis. In comparison to a normal distribution, this proposes that data points are less centered on average and have fewer high values. A normal distribution is similar to a distribution with zero kurtosis. It has tails and a moderate summit. Kurtosis can be computed with the following formula:

$$\text{Kurtosis}=\frac{k(k+1)}{(k-1)(k-2)(k-3)}\sum_{i=1}^{k}\left(\frac{x_i-\bar{x}}{s}\right)^4-\frac{3(k-1)^2}{(k-2)(k-3)} \qquad (8)$$

The formula takes the mean and divides it by the standard deviation to normalize the data values, then raises the result to the fourth power and averages it across all data points. The excess kurtosis (subtracting 3) is used to compare the kurtosis to that of a normal distribution. The kurtosis calculated can be seen in Figure 6.

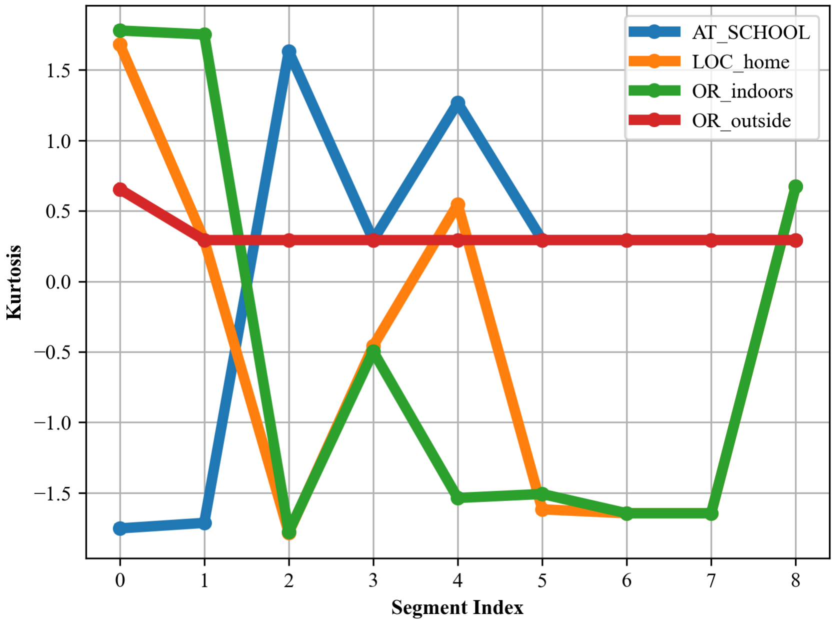

Figure 6. Kurtosis calculated for different activities.

3.3.5 Phase angle

Phase angle [32] describes the angle of the complex representation of a signal, providing information about the signal's phase relative to a reference point. Initially, a mathematical indication is computed, analytic signal is a complex signal derived from a real-valued signal using the Hilbert transform. It consists of the original signal (real part) and its Hilbert transform (imaginary part). Then Hilbert transform is used to generate the imaginary part of the analytic signal. It shifts the phase of the original signal by 90 degrees, creating a complex signal that represents the original signal in the complex plane. Lastly, the phase angle is computed. Given the real valued signal $x\ (t)$ and an analytic signal $z(t)$, which is an intricate signal derived from it:

$$z(t) = x(t) + j \cdot \hat{x}(t) \tag{9}$$

where j is the hypothetical unit, $\hat{x}(t)$ is the Hilbert transform of $x(t)$, and $x(t)$ is the original real-valued signals. The phase angle $\theta(t)$ can be computed as follows:

$$\theta(t) = \arg\big(z(t)\big) = \arctan\left(\frac{\hat{x}(t)}{x(t)}\right) \tag{10}$$

The phase angles are presented in Figure 7.

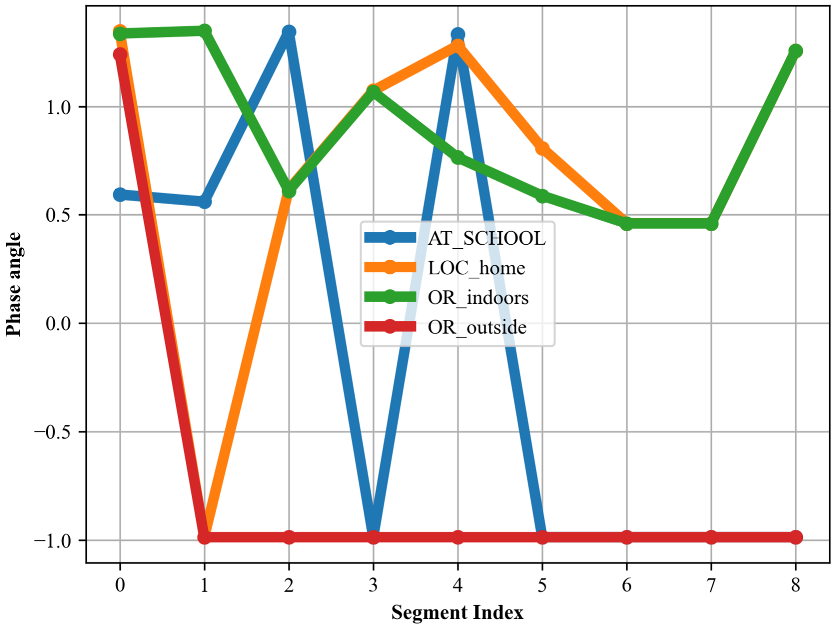

Figure 7. Phase angles calculated for different activities.

3.3.6 Data Augmentation

The over-sampling method known as SMOTE [33][34] is employed to generate artificial samples for the minority class. The method involves picking samples that are near the feature space, dividing the samples in the minority class along this line, and then creating new samples along this line. By not only replicating the current minority class samples, this contributes to the creation of a more balanced dataset. We found that the activity labels in our dataset were not evenly distributed, especially for the "outside" activity, which had fewer samples (about 12k) than the other activities. This inequality may result in skewed models that underperform in the minority class. We employed SMOTE (Synthetic Minority Over-sampling Technique) to solve this problem to balance the dataset. Mathematically, for each sample x_i in the minority class, identify its k-nearest neighbors. Typically, *k* is set to 5, then randomly select one of the k-nearest neighbors, x_{nn}, and produces an entirely new synthetic sample and x_{new} using formula (11). This disparity may result in skewed models that underperform in the minority class. We balanced the dataset using SMOTE (Synthetic Minority Over-sampling Technique) to solve this problem. Find the k-nearest neighbors of each sample x_i in the minority class using mathematics. Typically, $x_{nn}$is the k-nearest neighbor 5 is chosen at random, and x_{new} is the new synthetic sample is created using the following formula:

$$x_{\text{new}} = x_i + \lambda \cdot (x_{nn} - x_i) \tag{11}$$

Here, λ is arbitrary, ranging from 0 to 1. By utilizing this process, SMOTE generates synthetic samples that are variations of existing samples in the minority class, helping to balance the dataset.

3.3.7 Yeo-Johnson Optimization

We applied the Yeo-Johnson modification to optimize features. Optimization is important to adjust the distribution of data for machine learning models. This Statistical model removes nonnormal distribution of data so that the performance may be improved. We forwarded the original feature vector to the Yeo-Johnson transformation [35] process and observed that the optimized features exhibited a more normal distribution for several features. By reducing variance, this normalization approach improves the data's suitability for a wide range of analytical techniques and models. The Yeo-Johnson modification is beneficial since it can handle both positive and negative values, unlike the Box-Cox conversion, which can only work with positive numbers. We were able to reduce skewness and obtain a more Gaussian-like distribution by utilizing Yeo-Johnson to convert our features. This is crucial to enhance the performance of machine learning models for accurate prediction when the optimization models are used. Figure 8 elaborates the original and altered data feature vectors. Imbalanced data, in contrast to optimized data patterns, results in poor accuracy and predictions.

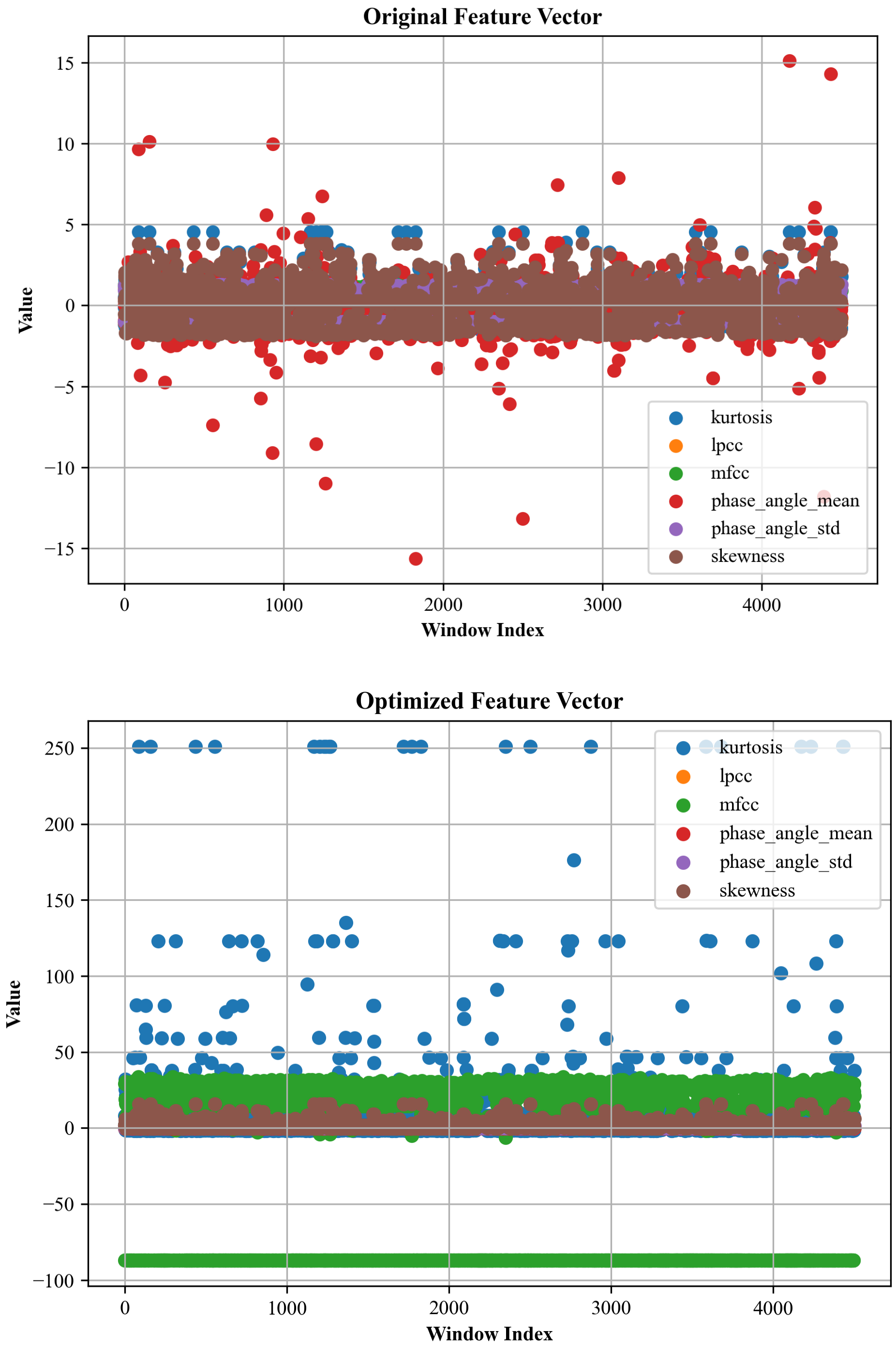

Figure 8. Feature vectors: original (top) and optimized (bottom).

3.3.8 Dataset Description

The ExtraSensory Dataset [36] is a publicly available dataset designed for human activity and context recognition. It was collected by researchers at the University of California, San Diego, using data from 60 participants who carried smartphones and wore smartwatches during their daily activities. The dataset includes sensor readings from multiple sources, such as accelerometers, gyroscopes, magnetometers, GPS, and microphones. These sensors capture a variety of signals related to movement, orientation, location, and ambient sound, providing a detailed view of user activities. A key feature of this dataset is the diversity of activities recorded, covering over 50 activity classes, including common actions like walking, sitting, standing, running, and more specific contexts such as listening to music, talking, or being in a vehicle. Data collection was done in natural settings, meaning participants followed their usual routines rather than performing scripted activities in a lab. This real-world approach makes the dataset useful for developing models that can handle the variability found in everyday human behavior. One challenge with the ExtraSensory dataset is missing data, as participants could choose to turn off sensors at any time. This makes it a good test case for building robust models that can handle incomplete information. The dataset is widely used in human activity recognition research and is freely available for academic and scientific studies.

.

3.3.9 Convolutional Neural Network Classifier

Convolutional neural networks (CNN) have been used to classify location-based tasks [37-39]. A 1-dimensional convolutional layer with 32 filters using ReLU as the activation function makes up the model's structure. Every filter has a kernel size of 3 by 3. This layer successfully extracts distinct patterns and attributes from the incoming data. Next, to reduce the dimensionality and hence the computational load and minimize overfitting, we employed a max pooling layer with a pool size of two. To prepare it for the fully connected layers, the 2-dimensional output from the levels before it was flattened to form a 1-dimensional vector. We added a thick layer with 50 units and ReLU activation to analyze the data further gathered and develop complex predictions by integrating characteristics. In a dropout layer with a 0.5 dropout rate, half of the input units were randomly assigned to 0 during training to avoid overfitting. The model terminates with a dense output layer, in which a softmax activation function the unit's number that corresponds to the number of classes in the dataset is used to submit each input sample to a probability distribution over the classes. utilizing categorical cross-entropy, the model was developed for multi-class classification problems utilizing the Adam optimizer, which is well known for its customizable learning rate capabilities. A proposed CNN layout is displayed in Figure 9.

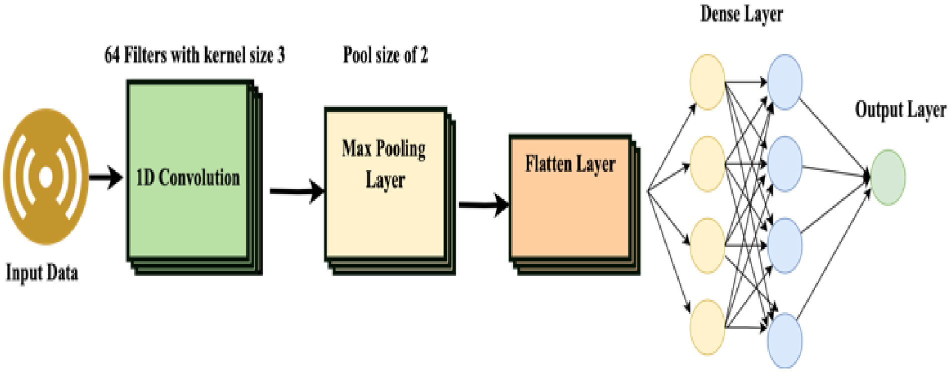

Figure 9. Proposed Architecture for CNN.

4 Experimental Results

An experiment was conducted on a macOS computer equipped with a 3.1 GHz Intel Core i5 processor and 16 GB of RAM. Python and pertinent libraries for deep learning, feature extraction, and data processing were used to implement the suggested model. The Convolutional Neural Network training process, feature extraction, and denoising required complicated operations that could be handled by the computational resources available.

4.1 Experiment 1: (Confusion Matrix)

The suggested model's performance in several localization tasks is depicted in the confusion matrix, Details of confusion matrix structure can be found in [40-42]. According to the results, the model was exceptionally accurate in predicting the activities, correctly classifying the majority of true labels into the relevant groups. The model had good predictive outcomes for the "label: AT_SCHOOL" class, correctly identifying 1,800 out of 1,876 samples with little misclassification. With all 1,876 samples properly identified, the "location home" class shows good accuracy, it also illustrates how well the models perform in detecting home-based activities.

With 42 samples in the "indoors" class mistakenly labeled as "AT SCHOOL" and 46 as "location home," there were only slight misclassifications in this class; nonetheless, the model correctly identified 1,732 out of 1,876 samples. Ultimately, the "outside" class exhibits the model's value in differentiating between outside activities by being perfectly classified with no errors. Figure 10 displays the confusion matrix.

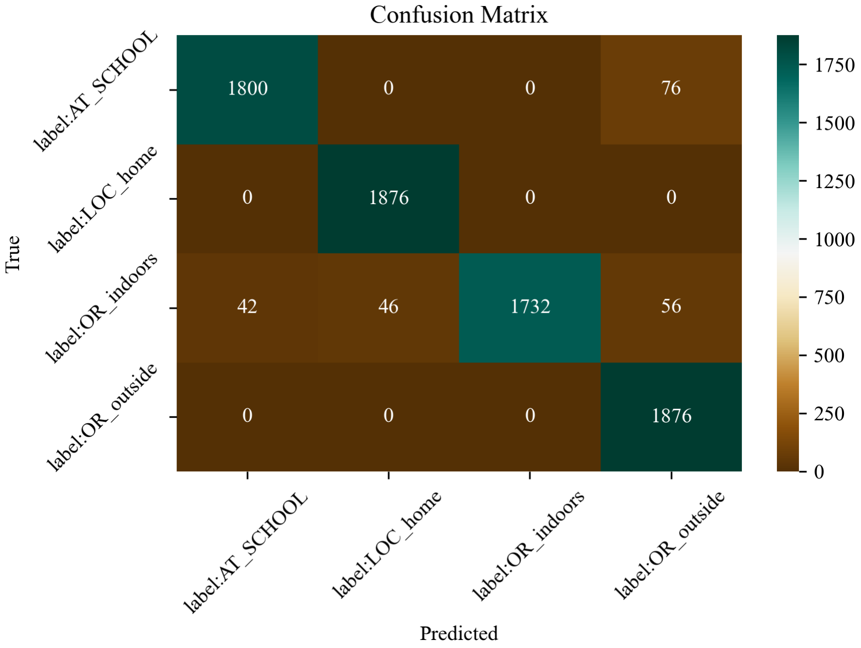

Figure 10. Confusion Matrix over Extrasensory Localization activities.

4.2 Experiment 2: (Precision, Recall, and F1-Score)

The research investigation comprises precision, recall, and F1-score evaluations to estimate how well the suggested model accomplishes in categorizing human localization tasks. The model assembles an F1-score of 0.97, a recall of 0.96, and a precision of 0.98 for the "label: AT_SCHOOL" class. This demonstrates that the model has a small percentage of false positives and false negatives and is very successful at recognizing education-related activities. With an F1-score of 0.99, a perfect recall of 1.00, and an accuracy of 0.98, the "location home" class performed significantly well. This implicit that there was almost no misclassification and that the model identified the activities that take place at home almost ideally. The recall for the " indoors" class decreased to 0.92 from a perfect precision of 1.00, yielding an F1-score of 0.96. Although the algorithm recognizes samples as being indoors when expected, it did not pick up on all cases where the activity was indoors. With an F1-score of 0.97, a perfect recall of 1.00, and an accuracy of 0.93, the "outside" class likewise fared well. This illustrates how effectively the algorithm detects outdoor activities with a low rate of false negatives. Every activity's detail is depicted in Table 2.

Table 2: Precision, Recall, and F1-score.

Classes	Precision	Recall	F1 Score
label:AT_SCHOOL	0.98	0.96	0.97
label: LOC_home	0.98	1.00	0.99
label: OR_indoors	1.00	0.92	0.96
label: OR_outside	0.93	1.00	0.97

4.3 Experiment 3: (K-fold Cross-validation)

In this experiment, a 5-fold cross-validation method was used to examine the robustness and generalizability of the proposed model across different subsets of the data. Cross-validation is essential in assessing how the model performs when trained and tested on different splits of the data, validating that the model's accomplishment is not overly reliant on a specific partition. The model was trained and evaluated over five folds, with each fold representing a distinct split of the dataset. The accuracy and loss metrics were recorded for each fold, and the results demonstrated consistently high accuracy across all folds, with an average accuracy of approximately 97.17% and a low average loss of 0.0305. The standard deviation of the accuracy across the folds was minimal, indicating the model's stability and reliability in classifying localization activities. In Table 3 the detail for each fold is presented.

Table 3: 5-fold cross-validation.

Fold No.	Loss	Accuracy
1	0.0230	99.6003%
2	0.0270	98.1339%
3	0.0312	97.2672%
4	0.0402	98.8008%
5	0.0311	99.0667%

4.4 Experiment 4: (Comparisons with existing systems)

In a comparative analysis with other models and, the most advanced techniques in this field, our proposed Hybrid-CNN model demonstrated superior performance in recognizing human localization activities. As shown in Table 4, the accuracy of the Hybrid-CNN model reached 96%, significantly outperforming other approaches. For instance, the Deep Neural Decision Forest achieved an accuracy of 89%, while both the HAR-Graph CNN and Random Forest methods reached 87%. Linear Regression lagged with an accuracy of 83%.

Table 4: Comparison with other methods.

Methods	Accuracy (%)
Deep Neural Decision Forest. [43]	0.89
HAR-Graph CNN. [44]	0.87
Random Forest [45]	0.87
Linear Regression [46]	0.83
Proposed Hybrid-CNN	0.96

5 Discussion and Limitations

The outcome of our proposed IoT-based Hybrid Deep Learning Model effectively recognizes human activity with enhanced accuracy compared to existing schemes, especially in complex and dynamic environments. The reason for the enhanced performance in terms of precision, recall, and F1 score across all activities is the integration of a Butterworth filter and data segmentation with Hamming windows for removing noise from IoT signals and the Yeo-Johnson transformation with the integration of SMOTE for data augmentation. This allows the deep learning models to achieve efficient and robust predictions and recognition. Our proposed model is most feasible for intelligent and context-aware dynamic environments where traditional approaches do not perform well.

Despite the high accuracy and precision, there are a few limitations to the proposed model. Its validation is recommended for deployment on other datasets or in real-world scenarios. The results may differ with larger datasets containing additional environment-specific features. Further testing and training are required to validate its computational cost and robustness. Secondly, the training phase may require significant memory and computing power, which might not be ideal for devices with limited computational resources. Thirdly, the performance of our proposed system may vary across different hardware platforms.

6 Conclusions

This research study proposed an IoT-based deep learning hybrid model that recognizes human activity with high accuracy and precision. Our proposed model consists of a CNN and advanced feature extraction and optimization techniques, which are ideal and recommended for complex and dynamic environments where frequent changes are expected. Its comparison with existing models further validates its performance in terms of high accuracy, precision, recall, and F1 score for object and human activity monitoring, tracking, and navigation in dynamic environments. Our proposed model is robust, accurate, and computationally cost-effective, feasible for smart homes, healthcare facilities, and child and disabled persons' activity monitoring. Looking ahead, we plan to test the model on more diverse datasets to ensure it performs well in different real-world environments. Another area of improvement is making the model lighter and more efficient for deployment on resource-limited devices like smartphones, smartwatches and embedded systems. We also aim to integrate Explainable AI (XAI) techniques to provide better interpretation of model decisions. Finally, our plan is to develop mobile application for Android and ios to make this technology more accessible, allowing real-time adaptability.

Acknowledgments

The authors would like to thank the ExtraSensory Dataset [36], which is publicly available for research purposes, as well as IMHESR and the University of Mosul.

7 References

[1] T. Silvio, Animation: The New Performance, Journal of Linguistic Anthropology, 20(2), pp. 422–438, 2010. https://doi.org/10.1111/j.1548-1395.2010.01078.

[2] L. Cheng, A. Zhao, K. Wang, H. Li, Y. Wang, and R. Chang, Activity recognition and localization based on UWB indoor positioning system and machine learning, in Proc. 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), (Vancouver, BC, Canada, 2020), pp. 0528–0533, https://doi.org/10.1109/IEMCON51383.2020.9284937.

[3] D. Khan, M. Alonazi, M. Abdelhaq, N. Al Mudawi, A. Algarni, A. Jalal, and H. Liu, Robust human locomotion and localization activity recognition over multisensory, Front. Physiol. 15 (2024) 1344887, https://doi.org/10.3389/fphys.2024.1344887.

[4] D. Moreira, M. Barandas, T. Rocha, P. Alves, R. Santos, R. Leonardo, P. Vieira, and H. Gamboa, Human activity recognition for indoor localization using smartphone inertial sensors, Sensors 21(18) (2021) 6316, https://doi.org/10.3390/s21186316.

[5] K. Ouchi and M. Doi, Indoor-outdoor activity recognition by a smartphone, in Proc. 2012 ACM Conf. Ubiquitous Computing (UbiComp '12), (New York, NY, USA, 2012), p. 537, https://doi.org/10.1145/2370216.2370297.

[6] A. Bilbao-Jayo, X. Cantero, A. Almeida, L. Fasano, T. Montanaro, I. Sergi, and L. Patrono, Location-based indoor and outdoor lightweight activity recognition system, Electronics 11(3) (2022) 360, https://doi.org/10.3390/electronics11030360.

[7] E. Bulbul, A. Cetin and I. A. Dogru, Human activity recognition using smartphones, in Proc. 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), (Ankara, Turkey, 2018), pp. 1–6, https://doi.org/10.1109/ISMSIT.2018.8567275.

[8] M. Alema Khatun and M. Abu Yousuf, Human activity recognition using smartphone sensor based on selective classifiers, in Proc. 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), (Dhaka, Bangladesh, 2020), pp. 1–6, https://doi.org/10.1109/STI50764.2020.9350486.

[9] A. V. Vesa, A. Carlan, D. Olaru, R. Pascariu, R. Pop, and G. Cosma, Human activity recognition using smartphone sensors and beacon-based indoor localization for ambient assisted living systems, in Proc. 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP), (Cluj-Napoca, Romania, 2020), pp. 205–212, https://doi.org/10.1109/ICCP51029.2020.9266158.

[10] A. Saha, T. Sharma, H. Batra, A. Jain and V. Pal, Human action recognition using smartphone sensors, in Proc. 2020 International Conference on Computational Performance Evaluation (ComPE), (Shillong, India, 2020), pp. 238–243, https://doi.org/10.1109/ComPE49325.2020.9200169.

[11] A. K. Muhammad Masum, S. Jannat, E. H. Bahadur, M. Golam Rabiul Alam, S. I. Khan and M. Robiul Alam, Human activity recognition using smartphone sensors: A dense neural network approach, in Proc. 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), (Dhaka, Bangladesh, 2019), pp. 1–6, https://doi.org/10.1109/ICASERT.2019.8934657.

[12] P. Podder, T. Z. Khan, M. H. Khan, and M. M. Rahman, Comparative performance analysis of Hamming, Hanning, and Blackman window, Int. J. Comput. Appl. 96(18) (2014) 1–5, https://doi.org/10.5120/16891-6927.

[13] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas, Window size impact in human activity recognition, Sensors 14(4) (2014) 6474–6499, https://doi.org/10.3390/s140406474.

[14] T.-H. Tan, J.-H. Tian, A. K. Sharma, S.-H. Liu, and Y.-F. Huang, Human activity recognition based on deep learning and micro-Doppler radar data, Sensors 24(8) (2024) 2530, https://doi.org/10.3390/s24082530.

[15] G. Dogan, S. S. Ertas, and İ. Cay, Human activity recognition using convolutional neural networks, in Proc. 2021 IEEE Conf. Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), (Melbourne, Australia, 2021), pp. 1–5, https://doi.org/10.1109/CIBCB49929.2021.9562906.

[16] F. Ordóñez and D. Roggen, Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition, Sensors 16(2016) 115, https://doi.org/10.3390/s16010115.

[17] D. De, P. Bharti, S. K. Das, and S. Chellappan, Multimodal wearable sensing for fine-grained activity recognition in healthcare, IEEE Internet Comput. 19 (2015) 26–35, https://doi.org/10.1109/MIC.2015.72.

[18] R. Chavarriaga, H. Sagha, A. Calatroni, S. T. Digumarti, G. Tröster, J. D. R. Millán, and D. Roggen, The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition, Pattern Recognit. Lett. 34 (2013) 2033–2042, https://doi.org/10.1016/j.patrec.2012.12.014.

[19] 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 19(2019) 1716, https://doi.org/10.3390/s19071716.

[20] M. Manokhin, P. Chollet, and P. Desgreys, Towards flexible and low-power wireless smart sensors: Reconfigurable analog-to-feature conversion for healthcare applications, Sensors (Basel) 24(3) (2024) 999, https://doi.org/10.3390/s24030999.

[21] M. Abdel-Basset, H. Hawash, V. Chang, R. K. Chakrabortty, and M. Ryan, Deep learning for heterogeneous human activity recognition in complex IoT applications, IEEE Internet Things J. 9 (2022) 5653–5665, https://doi.org/10.1109/jiot.2020.3038416.

[22] S. Konak, F. Turan, M. Shoaib, and Ö. D. Incel, Feature engineering for activity recognition from wrist-worn motion sensors, in Proc. International Conference on Pervasive and Embedded Computing and Communication Systems (2016), https://doi.org/10.5220/0006007100760084.

[23] G. Chetty, M. White, and F. Akther, Smart phone based data mining for human activity recognition, Procedia Comput. Sci. 46 (2015) 1181–1187, https://doi.org/10.1016/j.procs.2015.01.031.

[24] M. Ehatisham-ul-Haq and M. A. Azam, Opportunistic sensing for inferring in-the-wild human contexts based on activity pattern recognition using smart computing, Future Gener. Comput. Syst. 106 (2020) 374–392, https://doi.org/10.1016/j.future.2020.01.003.

[25] D. Khan et al., "Advanced IoT-Based Human Activity Recognition and Localization Using Deep Polynomial Neural Network," in IEEE Access, vol. 12, pp. 94337-94353, 2024, doi: 10.1109/ACCESS.2024.3420752.

[26] L. Cao, Y. Wang, B. Zhang, Q. Jin, and A. V. Vasilakos, GCHAR: An efficient group-based context-aware human activity recognition on smartphone, J. Parallel Distributed Comput. 118 (2017) 67–80, https://doi.org/10.1016/j.jpdc.2017.05.007.

[27] X. Zhang and S. Jiang, Application of Fourier transform and Butterworth filter in signal denoising, in Proc. 2021 6th Int. Conf. Intelligent Computing and Signal Processing (ICSP), (Xi'an, China, 2021), pp. 1277–1281, https://doi.org/10.1109/ICSP51882.2021.9408933.

[28] Z. Chen, C. Cai, T. Zheng, J. Luo, J. Xiong, and X. Wang, RF-based human activity recognition using signal adapted convolutional neural network, IEEE Trans. Mobile Comput. 22(1) (2023) 487–499, https://doi.org/10.1109/TMC.2021.3073969.

[29] H. Gupta and D. Gupta, LPC and LPCC method of feature extraction in speech recognition system, in Proc. 2016 6th Int. Conf. Cloud System and Big Data Engineering (Confluence), (Noida, India, 2016), pp. 498–502, https://doi.org/10.1109/CONFLUENCE.2016.7508171.

[30] M. A. Hossan, S. Memon, and M. A. Gregory, A novel approach for MFCC feature extraction, in Proc. 2010 4th Int. Conf. Signal Processing and Communication Systems, (Gold Coast, QLD, Australia, 2010), pp. 1–5, https://doi.org/10.1109/ICSPCS.2010.5709752.

[31] T. Cho, U. Sunarya, M. Yeo, B. Hwang, Y. S. Koo, and C. Park, Deep-ACTINet: End-to-end deep learning architecture for automatic sleep-wake detection using wrist actigraphy, Electronics 8(12) (2019) 1461, https://doi.org/10.3390/electronics8121461.

[32] B. K. Pappachan, W. Caesarendra, T. Tjahjowidodo, and T. Wijaya, Frequency domain analysis of sensor data for event classification in real-time robot-assisted deburring, Sensors 17(6) (2017) 1247, https://doi.org/10.3390/s17061247.

[33] H.-L. Le, D.-N. Nguyen, T.-H. Nguyen, and H.-N. Nguyen, A novel feature set extraction based on accelerometer sensor data for improving the fall detection system, Electronics 11 (2022) 1030, https://doi.org/10.3390/electronics11071030.

[34] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, SMOTE: Synthetic minority over-sampling technique, J. Artif. Intell. Res. 16(1) (2002) 321–357, https://doi.org/10.1613/jair.953.

[35] W. Satriaji and R. Kusumaningrum, Effect of synthetic minority oversampling technique (SMOTE), feature representation, and classification algorithm on imbalanced sentiment analysis, in Proc. 2018 2nd Int. Conf. Informatics and Computational Sciences (ICICoS), (Semarang, Indonesia, 2018), pp. 1–5, https://doi.org/10.1109/ICICOS.2018.8621648.

[36] I.-K. Yeo and R. A. Johnson, A new family of power transformations to improve normality or symmetry, Biometrika 87(4) (2000) 954–959, https://doi.org/10.1093/biomet/87.4.954.

[37] Y. Vaizman, K. Ellis, and G. Lanckriet, Recognizing detailed human context in the wild from smartphones and smartwatches, IEEE Pervasive Comput. 16(4) (2017) 62–74, https://doi.org/10.1109/MPRV.2017.3971131.

[38] R. K. Murthy, S. Dhanraj, T. N. Manjunath, A. N. Prasad, P. K. Pareek, and H. N. Kumar, A human activity recognition using CNN and long term short term memory, Int. J. Health Sci. 6(S6) (2022) 10797–10809, https://doi.org/10.53730/ijhs.v6nS6.12919.

[39] M. Younis, R. Ramo, and B. Bahnam, Explainable deep learning for hyperparameters optimization-based prediction of solar radiation intensity classification. Vietnam Journal of Computer Science (2025). https://doi.org/10.1142/S2196888825500058.

[40] R. I. Talal, and F. M. Ramo, Classification of personality traits by using pretrained deep learning models. In 2021 7th International Conference on Contemporary Information Technology and Mathematics (ICCITM), pp. 42-47. IEEE, 2021. https://doi.org/10.1109/ICCITM53167.2021.9677668

[41] M. C. Younis, E. Keedwell, and D. Savic, An investigation of pixel-based and object-based image classification in remote sensing. In 2018 International Conference on Advanced Science and Engineering (ICOASE) (pp. 449-454), (2018, October). IEEE. https://doi.org/10.1109/ICOASE.2018.8548845

[42] L. Alharbawee, and N. Pugeault. Generative adversarial networks for facial expression recognition in the wild. *International Journal of Computing and Digital Systems* 16, no. 1 (2024): 1259-1282. http://dx.doi.org/10.12785/ijcds/160193

[43] N. Gupta, S. K. Gupta, and V. Jain, Human activity recognition using CNN and LSTM for inertial sensors activity data, AIP Conf. Proc. 3072(1) (2024) 020019, https://doi.org/10.1063/5.0198752.

[44] 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 23 (2023) 7363, https://doi.org/10.3390/s23177363.

[45] M. Abduallah, 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. 2022 IEEE Int. Conf. Pervasive Computing and Communications Workshops and Other Affiliated Events, (New York, NY, USA, 2022), pp. 335–340, https://doi.org/10.1109/PerComWorkshops53856.2022.9767259.

[46] 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 Journal, 20 (2020) 4361–4371, https://doi.org/10.1109/JSEN.2020.2964278.