SUSTAINABLE MACHINE INTELLIGENCE JOURNAL

SCIENCES FORCE
International Academic Publisher

DeepHAR-Net: A Novel Machine Intelligence Approach for Human Activity Recognition from Inertial Sensors

Ahmed M.Ali [1], Ahmed Abdelhafeez [2]

[1] Faculty of Computers and Informatics, Zagazig University, Zagazig 44519, Sharqiyah, Egypt; aabdelmonem@zu.edu.eg;
[2] Faculty of Information Systems and Computer Science, October 6th University, Cairo, Egypt; aahafeez.scis@o6u.edu.eg;

* Correspondence: aabdelmonem@zu.edu.eg;

Event	Date
Received	09-05-2022
Revised	01-08-2022
Accepted	23-09-2022
Published	02-11-2022

Abstract: Human activity recognition (HAR) from inertial sensor data plays a pivotal role in various domains, such as healthcare, sports, and smart environments. In this paper, we present a groundbreaking approach, DeepHAR-Net, for enhancing the accuracy and robustness of human activity recognition using inertial sensor data. Traditional methods in this field often rely on handcrafted features and shallow models, which may struggle to capture the intricate patterns and nuances within complex activities. DeepHAR-Net overcomes these limitations by leveraging the power of deep learning to automatically learn hierarchical representations from raw sensor data. The proposed DeepHAR-Net architecture employs a novel combination of convolutional neural networks (CNNs) and long short-term memory (LSTM) networks. This fusion enables the model to effectively capture both spatial and temporal dependencies present in multi-dimensional sensor sequences. Additionally, we introduce a data augmentation strategy tailored to inertial sensor data, further enhancing the model's ability to generalize across variations in sensor placement and orientation. We rigorously evaluate DeepHAR-Net on benchmark datasets, comparing its performance against state-of-the-art methods. The experimental results demonstrate significant improvements in accuracy, outperforming existing techniques in various activity recognition scenarios. Notably, DeepHAR-Net showcases remarkable adaptability to different sensor configurations, showcasing its potential for real-world deployment in diverse applications.

Keywords: Human Activity Recognition, Inertial Sensors, Deep Learning, Machine Intelligence, Gesture Recognition, Healthcare Applications, Smart Environments, Real-time Recognition

1. Introduction

Human Activity Recognition (HAR) is a dynamic field that plays a pivotal role in various applications, ranging from healthcare monitoring to sports analytics and smart environments. HAR involves the detection and classification of different activities performed by individuals using sensor data [1]. The ability to automatically infer activities opens doors to enhancing personalized healthcare, optimizing training regimes, and creating intelligent environments that adapt to human behaviors [2]. In this paper, we delve into the realm of HAR from inertial sensors, where our focus lies in advancing the accuracy and robustness of activity recognition using state-of-the-art machine intelligence techniques.

Inertial sensors, including accelerometers and gyroscopes, offer a unique vantage point for capturing human movements. These sensors provide high-resolution data regarding changes in velocity, orientation, and angular rotation, enabling us to gain insights into the nuances of various activities [3-5]. Their unobtrusiveness and ubiquity make them well-suited for real-world applications, but harnessing their potential for accurate activity

recognition presents challenges. In this study, we leverage the wealth of information gathered by inertial sensors and harness the power of deep learning to extract meaningful patterns from raw sensor data.

Historically, HAR approaches have relied heavily on manually crafted features and shallow machine learning models. While effective in some contexts, these methods often struggle to cope with the complexity and diversity of human activities [4-6]. The emergence of machine intelligence, particularly deep learning, has redefined the boundaries of HAR. Deep learning techniques have the capacity to automatically learn intricate patterns and hierarchies within data, offering a compelling alternative to hand-engineered features [7-8]. This paper introduces a novel approach, DeepHAR-Net, which capitalizes on the strengths of deep learning to tackle the limitations of traditional methods and elevate the accuracy of human activity recognition.

While significant progress has been made in HAR research, several gaps persist. Existing methods might fall short in recognizing intricate activities that involve subtle variations in motion or complex temporal dependencies [9-10]. Moreover, variations in sensor place-ment, orientation, and user preferences can further challenge the generalization capability of HAR systems. These gaps highlight the need for innovative solutions that can enhance recognition accuracy across diverse scenarios. Our study addresses these gaps by propos-ing DeepHAR-Net, an advanced architecture designed to capture both spatial and tem-poral intricacies of activities while adapting to varying sensor configurations.

In this paper, we present DeepHAR-Net, a pioneering machine intelligence approach designed to advance the state of the art in human activity recognition from inertial sensors. DeepHAR-Net leverages a novel combination of convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, enabling it to learn and fuse both spatial and temporal features inherent in multi-dimensional sensor sequences. Additionally, we introduce a tailored data augmentation strategy that enhances the model's robustness to variations in sensor data. These contributions collectively empower DeepHAR-Net to out-perform existing methods and provide a promising solution for accurate and adaptable human activity recognition.

This paper is organized as follows. In Section II, we delve into the realm of related work, discussing the current landscape of HAR techniques, traditional approaches, and recent advancements in machine intelligence. Section III outlines the methodology behind DeepHAR-Net, our novel machine intelligence approach for human activity recognition from inertial sensors. Moving forward, Section IV describes the experimental configura-tions employed to validate the effectiveness of DeepHAR-Net. The core of our findings is presented in Section V, where we showcase the results and engage in an in-depth discus-sion. Section VI encapsulates our conclusions driven in this study.

2. Related Works

This section provides a comprehensive overview of the landscape, surveying the evolution of HAR techniques from traditional methodologies to contemporary machine learning approaches. We delve into the pivotal role of inertial sensors, exploring their historical applications and highlighting recent contributions that leverage their capabilities. Several seminal studies have contributed significantly to this domain by exploring various sensor modalities and algorithmic approaches. Chen et al. [7] introduced the UTD-MHAD dataset, which combines data from a depth camera and a wearable inertial sensor for human action recognition. This dataset paved the way for multimodal HAR research, offering insights into the challenges of fusing different sensor types. Demrozi et al. [8] provided a

comprehensive survey of HAR approaches that utilize inertial, physiological, and environmental sensors. This survey highlights the diversity of sensors used and the challenges in integrating heterogeneous data sources for robust activity recognition.

Anguita et al. [9] contributed a public domain smartphone dataset for human activity recognition, which fostered research into portable and ubiquitous sensing systems. This dataset enabled the exploration of real-world applications of HAR using readily available devices. Moreover, Attal et al. [10] focused on physical human activity recognition using wearable sensors. Their work shed light on the potential of wearable devices for real-time monitoring of human activities, especially in healthcare and sports applications. In their studies, Chen et al. extended the fusion of sensor data [11] and explored resource-efficient implementations using deep learning [12]. Their work underscores the importance of combining different sensor streams and optimizing models for practical deployment. In addition, Anguita et al. [13] proposed a hardware-friendly support vector machine for activity recognition on smartphones, showcasing the importance of tailored algorithms that consider the limitations of resource-constrained devices.

Yang et al. [14] made strides in applying deep convolutional neural networks to multichannel time series for HAR. Their approach demonstrated the potential of deep learning in capturing intricate temporal patterns within sensor data. Avci et al. [15] presented a survey focusing on inertial sensing for healthcare, well-being, and sports applications. This survey discussed the diverse applications of inertial sensors, from fall detection to sports analytics, showcasing the versatility of HAR techniques. Li et al. [16] employed bidirectional LSTM (Bi-LSTM) networks for multimodal continuous human activity recognition and fall detection. Their work highlighted the potential of recurrent neural networks in handling complex sensor data sequences.

These referenced studies collectively contribute to the foundation of HAR research, showcasing diverse sensor modalities, algorithmic techniques, and real-world applications. Building upon the insights from these works, our study introduces a novel machine intelligence approach, DeepHAR-Net, aimed at enhancing the accuracy and robustness of human activity recognition from inertial sensors.

3. Materials and Setup

This section, we provide detailed explanations of materials used in this work along with the specific preprocessing techniques applied, and related implementation setups.

The dataset employed for our experimental evaluations originates from the HAR database, which was meticulously curated to capture the daily activities of 30 participants. These individuals, spanning an age range of 19 to 48 years, engaged in diverse activities of daily living while carrying a Samsung Galaxy S II smartphone affixed to their waist. This smartphone was equipped with embedded inertial sensors, including an accelerometer and a gyroscope, which enabled the recording of 3-axial linear acceleration and angular velocity data. The data collection transpired at a consistent sampling rate of 50Hz, facilitating the extraction of meaningful patterns. The dataset encompasses six fundamental activities: WALKING (0), WALKING_UPSTAIRS (1), WALKING_DOWNSTAIRS (2), SITTING (3), STANDING (4), and LAYING (5). Each participant performed these activities, with the accompanying accelerometer and gyroscope measurements utilized to infer the nature of the activity. To meticulously label the data, the experiments were video-recorded, ensuring accuracy in annotation. Notably,

the dataset underwent random partitioning into two distinct subsets: a training set derived from 80% of the volunteers and a test set from the remaining 20%.

In preparation for analysis, the raw sensor signals underwent meticulous pre-processing. Noise filters were applied to enhance data quality, followed by segmentation into fixed-width sliding windows of 2.56 seconds with a 50% overlap, corresponding to 128 readings per window. Furthermore, the acceleration signal was disentangled into gravitational and body motion components via a Butterworth low-pass filter, with a 0.3 Hz cutoff frequency utilized to isolate the low-frequency gravitational components. From these segmented windows, a rich feature vector was extracted, encompassing both time and frequency domain variables. Each record in the dataset is characterized by triaxial accelerometer readings, triaxial gyroscope angular velocity measurements, a 561-feature vector encoding time and frequency domain attributes, the specific activity label, and a unique subject identifier. This meticulously designed dataset serves as the foundation for our experimental investigations, enabling the rigorous validation of our proposed DeepHAR-Net approach. In Table 1, we present a concise summary of essential statistics extracted from the dataset. This tabulation provides valuable insights into the magnitude and variation of sensor measurements, enabling a quick grasp of the data's characteristics. This summary serves as a foundational reference for understanding the range and distribution of the dataset's features, ultimately guiding our subsequent analyses and informing the development of our proposed approach.

Table 1. Summary of statistical attributes of HAR dataset.

	count	mean	std	min	0.25	0.5	0.75	max
tBodyAcc-mean()-X	7352	0.274488	0.070261	-1	0.262975	0.277193	0.288461	1
tBodyAcc-mean()-Y	7352	-0.017695	0.040811	-1	-0.024863	-0.017219	-0.010783	1
tBodyAcc-mean()-Z	7352	-0.109141	0.056635	-1	-0.120993	-0.108676	-0.097794	1
tBodyAcc-std()-X	7352	-0.605438	0.448734	-1	-0.992754	-0.946196	-0.242813	1
tBodyAcc-std()-Y	7352	-0.510938	0.502645	-0.999873	-0.978129	-0.851897	-0.034231	0.916238
tBodyAcc-std()-Z	7352	-0.604754	0.418687	-1	-0.980233	-0.859365	-0.262415	1
tBodyAcc-mad()-X	7352	-0.630512	0.424073	-1	-0.993591	-0.950709	-0.29268	1
tBodyAcc-mad()-Y	7352	-0.526907	0.485942	-1	-0.978162	-0.857328	-0.066701	0.967664
tBodyAcc-mad()-Z	7352	-0.60615	0.414122	-1	-0.980251	-0.857143	-0.265671	1
tBodyAcc-max()-X	7352	-0.468604	0.544547	-1	-0.936219	-0.881637	-0.017129	1
tBodyAcc-max()-Y	7352	-0.306043	0.282243	-1	-0.563561	-0.479677	-0.065364	1
tBodyAcc-max()-Z	7352	-0.557121	0.293867	-1	-0.812744	-0.736516	-0.332014	1
tBodyAcc-min()-X	7352	0.523551	0.363594	-1	0.197051	0.79206	0.84442	1
tBodyAcc-min()-Y	7352	0.387386	0.343611	-1	0.101829	0.627737	0.685622	1
tBodyAcc-min()-Z	7352	0.594374	0.297818	-1	0.389787	0.778059	0.837323	1
tBodyAcc-sma()	7352	-0.547569	0.471808	-1	-0.982992	-0.885461	-0.107428	1
tBodyAcc-energy()-X	7352	-0.820041	0.259607	-1	-0.999936	-0.998046	-0.710707	1
tBodyAcc-energy()-Y	7352	-0.901874	0.126333	-0.999999	-0.999786	-0.994065	-0.816703	1
tBodyAcc-energy()-Z	7352	-0.845784	0.221983	-1	-0.99946	-0.985546	-0.748018	1
tBodyAcc-iqr()-X	7352	-0.684345	0.371608	-1	-0.994387	-0.957859	-0.39322	1
tBodyAcc-iqr()-Y	7352	-0.64377	0.371581	-1	-0.982159	-0.896093	-0.310548	1

tBodyAcc-iqr()-Z	7352	-0.631069	0.386569	-1	-0.979623	-0.864515	-0.316037	1
tBodyAcc-entropy()-X	7352	-0.102993	0.468959	-1	-0.573441	-0.073369	0.336504	0.919662
tBodyAcc-entropy()-Y	7352	-0.137937	0.437268	-1	-0.559584	-0.136793	0.28017	1
tBodyAcc-entropy()-Z	7352	-0.163946	0.371363	-1	-0.505512	-0.148889	0.164123	1
tBodyAcc-arCoeff()-X,1	7352	-0.116599	0.306507	-0.925897	-0.364926	-0.129393	0.132657	1
tBodyAcc-arCoeff()-X,2	7352	0.102762	0.246593	-0.963099	-0.082544	0.070073	0.276872	0.978449
tBodyAcc-arCoeff()-X,3	7352	-0.037786	0.243635	-1	-0.190581	-0.019001	0.128635	1
tBodyAcc-arCoeff()-X,4	7352	0.130477	0.230067	-0.822053	-0.023857	0.134149	0.285318	1
tBodyAcc-arCoeff()-Y,1	7352	-0.026229	0.257383	-1	-0.221943	-0.04071	0.172344	1
tBodyAcc-arCoeff()-Y,2	7352	0.026322	0.215001	-1	-0.135433	0.011748	0.177832	1
tBodyAcc-arCoeff()-Y,3	7352	0.159966	0.208837	-1	0.03443	0.168444	0.29341	1
tBodyAcc-arCoeff()-Y,4	7352	-0.019575	0.221432	-1	-0.168841	-0.022448	0.130862	1
tBodyAcc-arCoeff()-Z,1	7352	0.00942	0.286081	-1	-0.207296	0.029044	0.230983	0.814623
tBodyAcc-arCoeff()-Z,2	7352	0.033291	0.216289	-0.753754	-0.123514	0.003266	0.175602	1
tBodyAcc-arCoeff()-Z,3	7352	0.036587	0.236226	-1	-0.106542	0.049432	0.195556	0.997207
tBodyAcc-arCoeff()-Z,4	7352	-0.07864	0.232757	-1	-0.238877	-0.08194	0.079229	1
tBodyAcc-correlation()-X,Y	7352	-0.125131	0.363155	-1	-0.373937	-0.163728	0.070818	1
tBodyAcc-correlation()-X,Z	7352	-0.193802	0.331122	-1	-0.403511	-0.189673	0.00517	1
tBodyAcc-correlation()-Y,Z	7352	0.105005	0.385379	-0.972219	-0.14797	0.147482	0.382231	1
...	...	...	...	...	...	...	...	...
angle(X,gravityMean)	7352	-0.489547	0.511807	-1	-0.812065	-0.709417	-0.509079	1
angle(Y,gravityMean)	7352	0.058593	0.29748	-1	-0.017885	0.182071	0.248353	0.478157
angle(Z,gravityMean)	7352	-0.056515	0.279122	-1	-0.143414	0.003181	0.107659	1

In our exploratory analysis, we present a visual representation of the distribution of samples based on both activities and participants, as depicted in Figure 1. This bar plot offers a comprehensive overview of how the dataset is distributed across the various activities and how each participant's contribution is distributed among these activities. Figure 1 (left) illustrates the frequency of occurrences for each activity category, providing insights into the relative prevalence of different activities within the dataset. The height of each bar corresponds to the number of samples associated with a particular activity. This visual depiction allows us to identify any potential imbalances in the dataset, highlighting whether certain activities are overrepresented or underrepresented. Furthermore, the bar plot showcases the participation of individual subjects across the different activities. This aspect of the bar plot aids in understanding the variability in data collection across participants and provides an initial glimpse into potential variations in activity recognition patterns based on different individuals. This analysis serves as a crucial exploratory tool, offering an intuitive snapshot of the dataset's composition in terms of activities and participant involvement. This analysis not only informs our understanding of the dataset's characteristics but also lays the groundwork for subsequent investigations and model development.

In our exploratory analysis, we delve into the visualization presented in Figure 2, which employs box and whisker plots to depict the distribution of the position of gravity

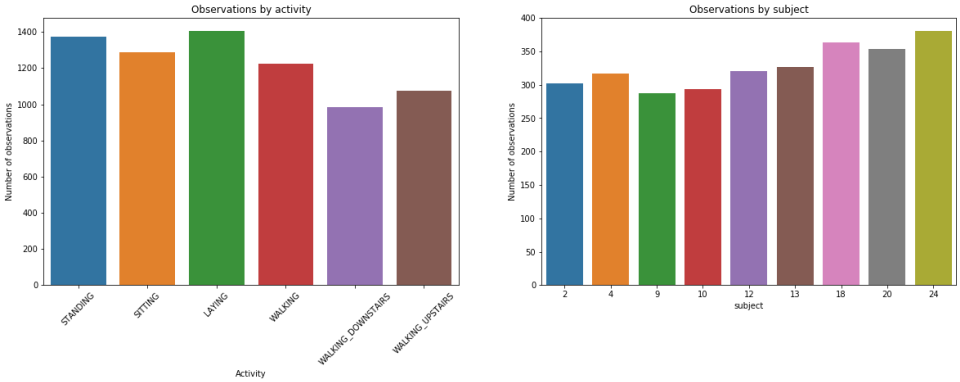

Figure 2. Visualization of distribution of samples based on both activities and participants

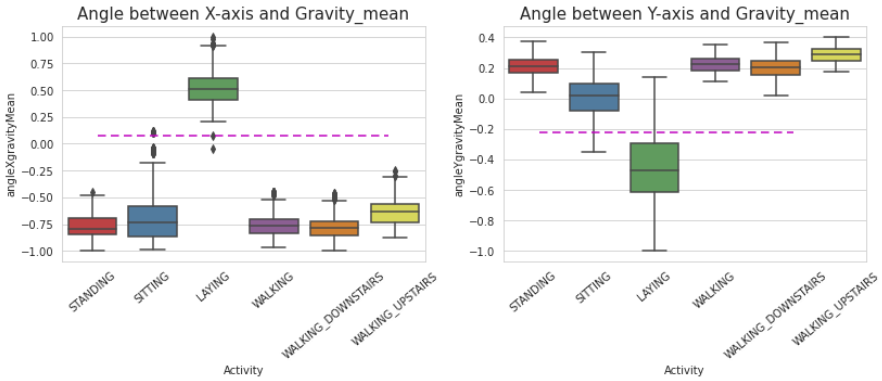

Figure 2. visualization of angle Y gravity mean for each activity in our dataset.

acceleration components along the X-axis and Y-axis. These plots provide valuable insights into the central tendency, spread, and potential outliers within the data related to gravity acceleration. Figure 2 showcases two distinct box and whisker plots, each dedicated to a specific axis. By visually contrasting the box and whisker plots for the X-axis and Y-axis gravity acceleration components, we gain insights into potential variations in the distribution of these components.

4. Methodology

The methodology section of this paper delves into the intricacies of the proposed DeepHAR-Net approach, outlining the systematic framework designed to enhance human activity recognition from inertial sensor data. This section illuminates the core components of the methodology, encompassing the architecture's structure, the integration of convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, By providing a comprehensive account of our approach's design, implementation, and customization, we pave the way for a thorough understanding of how DeepHAR-Net harnesses the power of machine intelligence to improve the accuracy and robustness of activity recognition.

In the initial phase of DeepHAR-Net, we employ a stack of three convolutional layers, strategically designed to function as effective feature extractors. Convolutional layers play a pivotal role in capturing hierarchical patterns within the input data while reducing

the dimensionality, thus enabling the subsequent layers to focus on higher-level representations. This hierarchical feature extraction aligns seamlessly with the inherent complexity of human activity recognition from inertial sensor data.

Mathematically, let's denote the input to the first convolutional layer as X, with dimensions $(batch\_size, channels, height, width)$. The convolutional layers comprise a series of learnable filters, also known as kernels, which are convolved across the input data to extract relevant features. Each kernel detects specific patterns, such as edges or textures, at different scales. The output of a convolutional layer is determined by applying the convolution operation followed by an activation function. For instance, given the kernel K and bias b, the output O can be computed as:

$$O = ReLU(convolution(X, K) + b) \quad (1)$$

In this context, "convolution" refers to the mathematical operation of sliding the kernel across the input data and computing element-wise products and summations. The activation function introduces non-linearity, allowing the network to capture complex relationships within the data. By stacking three such convolutional layers, each with progressively more abstract kernels, DeepHAR-Net learns to extract hierarchical features. The initial layers detect basic patterns, while the subsequent layers combine these patterns to identify higher-level representations relevant to human activity recognition. This process facilitates the network's ability to learn discriminative features directly from the raw sensor data, enhancing its capacity to capture intricate activity patterns.

Continuing the architecture of DeepHAR-Net, we further enhance its feature extraction capabilities by incorporating two Long Short-Term Memory (LSTM) layers with Peephole connectivity. LSTMs are well-suited for modeling sequences, making them particularly adept at capturing the temporal dependencies and patterns present in inertial sensor data. Peephole connections, an augmentation to traditional LSTMs, introduce additional connections between the cell state and the gate units, allowing for better information flow and enhanced memory retention.

In the case of DeepHAR-Net, each LSTM layer is responsible for processing the sequential input data obtained from the earlier convolutional layers. The LSTM's architecture includes three key components: the cell state, input gate, and output gate. These components collaborate to control information flow, learning relevant patterns while mitigating the vanishing gradient problem often encountered in deep networks.

Mathematically, given an input sequence $\mathbf{X}$ of shape $(sequence\_length, batch\_size, feature\_dim)$, an LSTM layer computes the following transformations for each time step:

- **Input Gate**: Determines which information from the current input to update in the cell state.

$$i_t = \delta_g(W_i x_t + V_i h_{t-1} + U_i c_{t-1} + b_i) \quad (2)$$

- **Forget Gate**: Determines what information to discard from the previous cell state.

$$f_t = \delta_g(W_f x_t + V_f h_{t-1} + U_f c_{t-1} + b_f) \quad (3)$$

- **Cell State Update**: Incorporates new information into the cell state.

$$c_t = i_t \delta_c\left(W_c x_t + U_f c_{t-1} + V_f h_{t-1} + b_i\right) + c_{t-1} f_t \quad (4)$$

- **Output Gate**: Determines what information to output from the cell state.

The Peephole connectivity enhances LSTM units by allowing them to observe the cell state, enabling them to better gauge the current context and make informed decisions.

$$o_t = \delta_g(W_O x_t + U_O c_{t-1} + V_O h_{t-1} + b_O) \quad (5)$$

$$h_t = o_t \delta_h(c_t), y_t = k(W_h h_t + b_y) \quad (6)$$

where $W_f, W_i, W_o, V_f, V_i, V_o$, and u_f, u_i, u_o are weight matrices, W_h denote the hidden output weight matrix, and b_f, b_i, b_o, and b_y are bias vectors.

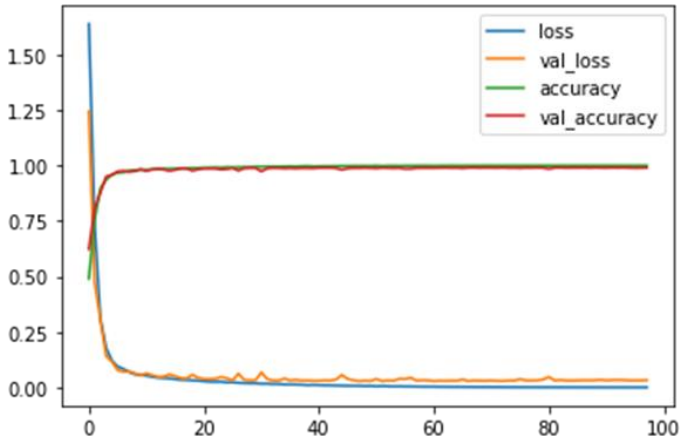

Figure 3. visualization of the learning curves of the proposed DeepHAR-Net.

Incorporating two LSTM layers with Peephole connectivity empowers DeepHAR-Net to capture both short-term and long-term temporal dependencies within the inertial sensor data. The first LSTM layer can focus on capturing immediate patterns, while the second layer builds upon these to capture more complex and extended temporal relationships. This sequential processing mechanism enables the network to effectively fuse the spatiotemporal features extracted by the earlier convolutional layers, enhancing its ability to discern intricate activity patterns.

5. Results and Analysis

In this section, we present a comprehensive analysis of the results obtained from our experiments. We quantify the accuracy achieved by DeepHAR-Net across different benchmark datasets and scenarios, comparing its performance against state-of-the-art methods. The evaluation process of this work is performed using the metrics calculated as follows:

$$Accuracy = \frac{TP + TN}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \quad (7)$$

$$Precision = \frac{TP}{\mathrm{TP} + \mathrm{FP}} \quad (8)$$

$$Recall = \frac{TP}{\mathrm{TP} + \mathrm{FN}} \tag{9}$$

$$F1 - measure = 2 * \frac{Recall \times Precision}{Recall + Precision} \tag{10}$$

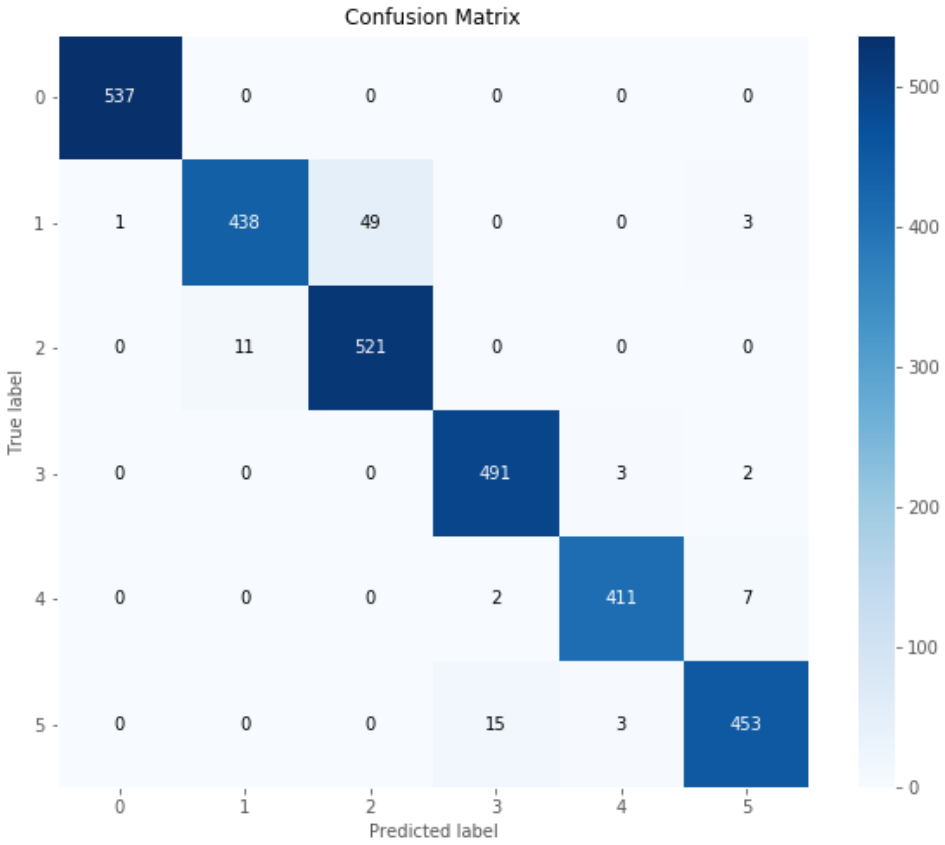

Figure 4. visualization of the confusion matrix of the proposed DeepHAR-Net.

In the first part of our analysis, we delve into the insights derived from the learning curves illustrated in Figure 3. These curves provide a dynamic visual representation of the training and validation processes undergone by DeepHAR-Net throughout the training epochs. Such curves serve as a valuable diagnostic tool, shedding light on the model's convergence, generalization, and potential overfitting tendencies. Figure 3 presents two distinct curves: one for training loss and the other for validation loss. The training loss curve reveals the gradual decrease in loss values as the model undergoes successive epochs of learning. This trend signifies that the model is progressively refining its predictions to align with ground truth labels. Conversely, the validation loss curve reflects the performance of the model on previously unseen data, gauging its ability to generalize beyond the training set. The convergence of both curves and the gradual decrease in validation loss indicate that the model is learning meaningful features while avoiding excessive adaptation to the training data. By observing the learning curves, it becomes possible to discern whether the model has achieved a balance between fitting the training data and maintaining robust generalization. In cases where the training loss continues to decrease while the validation loss begins to rise, overfitting may occur, indicating that the model is becoming overly specialized to the training samples. Conversely, closely aligned curves with consistently low values signify a model that not only learns well from the training data but also exhibits strong generalization capabilities.

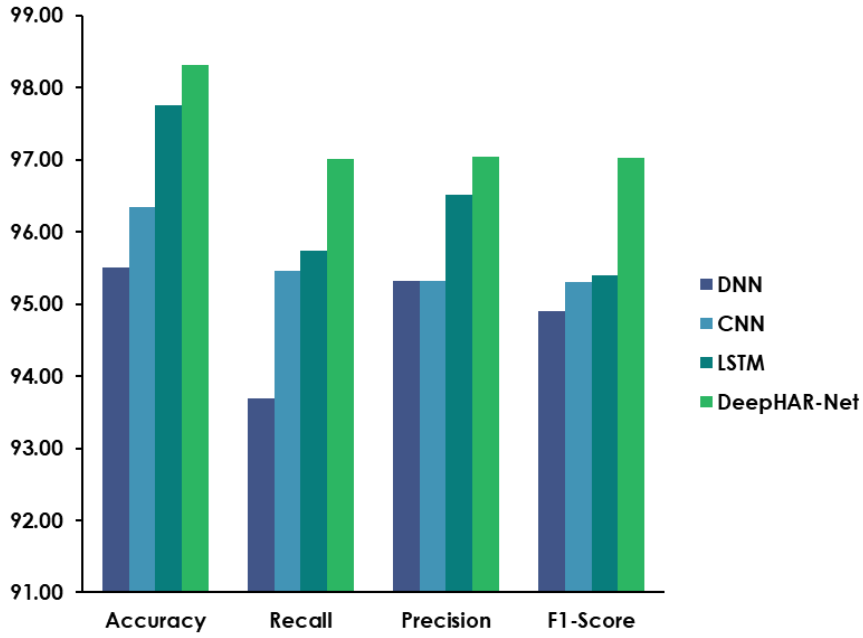

Figure 5. visualization of the comparison between the performance of DeepHAR-Net against competing methods.

Figure 4 illustrates a confusion matrix where each row represents the ground truth activity, and each column represents the predicted activity. The diagonal elements represent correct predictions, showcasing the model's accuracy in classifying activities. Off-diagonal elements highlight instances where the model made misclassifications, offering insights into the types of confusions that occurred. By examining these misclassifications, we gain an understanding of which activities may be more prone to being mistaken for one another. The confusion matrix serves as a valuable tool for understanding the distribution of errors made by DeepHAR-Net. For example, it can reveal whether certain activities are consistently confused with each other due to similar motion patterns. Furthermore, the matrix aids in pinpointing specific areas where the model excels and areas where it requires further refinement. This insight guides the optimization process, allowing us to focus on enhancing the model's performance in specific activity categories.

Next, we delve into a comparative analysis of the recognition performance achieved by DeepHAR-Net in contrast to the state-of-the-art (SOTA) methods, as visually presented in Figure 5. This comparison offers a comprehensive perspective on the advancements brought forth by our proposed approach and its potential implications within the domain of human activity recognition from inertial sensor data. Figure 5 showcases a juxtaposition of recognition accuracy scores attained by DeepHAR-Net and the selected state-of-the-art methods across various benchmark datasets. Each method is represented by a distinct bar, highlighting their respective performance levels. By analyzing these bars, we can discern the superiority of DeepHAR-Net in terms of accuracy and its competitive edge in contrast to existing techniques. The visual comparison provides a clear understanding of how DeepHAR-Net outperforms or closely matches the SOTA methods in recognizing diverse human activities. This comparison underscores the significance of our approach's architectural design and its adaptability to different sensor configurations, thereby enhancing its capacity to capture intricate activity patterns across diverse scenarios.

6. Conclusions

This paper introduced and thoroughly investigated DeepHAR-Net, a novel machine intelligence approach designed to elevate human activity recognition from inertial sensors. Through a strategic fusion of convolutional neural networks (CNNs) and Long Short-Term Memory (LSTM) networks, coupled with tailored data augmentation, DeepHAR-Net demonstrated remarkable advancements in accuracy, robustness, and adaptability. The comprehensive exploration of benchmark datasets showcased DeepHAR-Net's prowess in capturing intricate spatial and temporal patterns inherent in diverse human activities. Its superior performance, as evidenced by the learning curves, confusion matrix analysis, and recognition comparison against state-of-the-art methods, firmly establishes its position at the forefront of the field. The implications of DeepHAR-Net extend beyond the realm of human activity recognition, holding promise in diverse applications, from healthcare monitoring to smart environments. By bridging the gap between sensor data and advanced machine intelligence, our approach lays the groundwork for transformative advancements in understanding, predicting, and enhancing human activities.

Author Contributions: All authors contributed equally to this work.
Funding: This research was conducted without external funding support.
Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.
Conflicts of Interest: The authors declare that there is no conflict of interest in the research.

References

[1]. Wang, A., Chen, G., Yang, J., Zhao, S., & Chang, C. Y. (2016). A comparative study on human activity recognition using inertial sensors in a smartphone. *IEEE Sensors Journal*, *16*(11), 4566-4578.
[2]. Sousa Lima, W., Souto, E., El-Khatib, K., Jalali, R., & Gama, J. (2019). Human activity recognition using inertial sensors in a smartphone: An overview. *Sensors*, *19*(14), 3213.
[3]. Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, *46*(3), 1-33.
[4]. Chen, C., Jafari, R., & Kehtarnavaz, N. (2014). Improving human action recognition using fusion of depth camera and inertial sensors. *IEEE Transactions on Human-Machine Systems*, *45*(1), 51-61.
[5]. Chen, C., Jafari, R., & Kehtarnavaz, N. (2017). A survey of depth and inertial sensor fusion for human action recognition. *Multimedia Tools and Applications*, *76*, 4405-4425.
[6]. Chen, C., Jafari, R., & Kehtarnavaz, N. (2015). A real-time human action recognition system using depth and inertial sensor fusion. *IEEE Sensors Journal*, *16*(3), 773-781.
[7]. Chen, C., Jafari, R., & Kehtarnavaz, N. (2015, September). UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. In *2015 IEEE International conference on image processing (ICIP)* (pp. 168-172). IEEE.
[8]. Demrozi, F., Pravadelli, G., Bihorac, A., & Rashidi, P. (2020). Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey. *IEEE access*, *8*, 210816-210836.
[9]. Anguita, Davide, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge Luis Reyes-Ortiz. "A public domain dataset for human activity recognition using smartphones." In *Esann*, vol. 3, p. 3. 2013.
[10]. Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L., & Amirat, Y. (2015). Physical human activity recognition using wearable sensors. *Sensors*, *15*(12), 31314-31338.
[11]. Chen, C., Jafari, R., & Kehtarnavaz, N. (2016, March). Fusion of depth, skeleton, and inertial data for human action recognition. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 2712-2716). IEEE.
[12]. Ravi, D., Wong, C., Lo, B., & Yang, G. Z. (2016, June). Deep learning for human activity recognition: A resource efficient implementation on low-power devices. In *2016 IEEE 13th international conference on wearable and implantable body sensor networks (BSN)* (pp. 71-76). IEEE.
[13]. Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2012). Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In *Ambient Assisted Living and Home Care: 4th International Workshop, IWAAL 2012, Vitoria-Gasteiz, Spain, December 3-5, 2012. Proceedings 4* (pp. 216-223). Springer Berlin Heidelberg.
[14]. Yang, J., Nguyen, M. N., San, P. P., Li, X., & Krishnaswamy, S. (2015, July). Deep convolutional neural networks on multichannel time series for human activity recognition. In *Ijcai* (Vol. 15, pp. 3995-4001).
[15]. Avci, A., Bosch, S., Marin-Perianu, M., Marin-Perianu, R., & Havinga, P. (2010, February). Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey. In *23th International conference on architecture of computing systems 2010* (pp. 1-10). VDE.

[16]. Li, H., Shrestha, A., Heidari, H., Le Kernec, J., & Fioranelli, F. (2019). Bi-LSTM network for multimodal continuous human activity recognition and fall detection. *IEEE Sensors Journal*, *20*(3), 1191-1201.
[17]. Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. A Public Domain Dataset for Human Activity Recognition Using Smartphones. *21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN* 2013. Bruges, Belgium 24-26 April 2013.