

Article

Optimal Sensor Placement and Multimodal Fusion for Human Activity Recognition in Agricultural Tasks

Lefteris Benos ^{1,*}, Dimitrios Tsaopoulos ¹, Aristotelis C. Tagarakis ¹, Dimitrios Kateris ¹
and Dionysis Bochtis ^{1,2}

¹ Institute for Bio-Economy and Agri-Technology (IBO), Centre of Research and Technology-Hellas (CERTH), GR57001 Thessaloniki, Greece; d.tsaopoulos@certh.gr (D.T.); a.tagarakis@certh.gr (A.C.T.); d.kateris@certh.gr (D.K.); d.bochtis@certh.gr (D.B.)

² farmB Digital Agriculture, Doiranis 17, GR54639 Thessaloniki, Greece

* Correspondence: e.benos@certh.gr

Abstract: This study examines the impact of sensor placement and multimodal sensor fusion on the performance of a Long Short-Term Memory (LSTM)-based model for human activity classification taking place in an agricultural harvesting scenario involving human-robot collaboration. Data were collected from twenty participants performing six distinct activities using five wearable inertial measurement units placed at various anatomical locations. The signals collected from the sensors were first processed to eliminate noise and then input into an LSTM neural network for recognizing features in sequential time-dependent data. Results indicated that the chest-mounted sensor provided the highest F1-score of 0.939, representing superior performance over other placements and combinations of them. Moreover, the magnetometer surpassed the accelerometer and gyroscope, highlighting its superior ability to capture crucial orientation and motion data related to the investigated activities. However, multimodal fusion of accelerometer, gyroscope, and magnetometer data showed the benefit of integrating data from different sensor types to improve classification accuracy. The study emphasizes the effectiveness of strategic sensor placement and fusion in optimizing human activity recognition, thus minimizing data requirements and computational expenses, and resulting in a cost-optimal system configuration. Overall, this research contributes to the development of more intelligent, safe, cost-effective adaptive synergistic systems that can be integrated into a variety of applications.

Keywords: Long Short-Term Memory (LSTM) networks; wearable sensors; multi-sensor information fusion; human-robot collaboration; human factors; cost-optimal system configuration



Citation: Benos, L.; Tsaopoulos, D.; Tagarakis, A.C.; Kateris, D.; Bochtis, D. Optimal Sensor Placement and Multimodal Fusion for Human Activity Recognition in Agricultural Tasks. *Appl. Sci.* **2024**, *14*, 8520. <https://doi.org/10.3390/app14188520>

Academic Editors: Petru Alexandru Vlaicu and Basarab Matei

Received: 30 August 2024

Revised: 17 September 2024

Accepted: 17 September 2024

Published: 21 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Agricultural environments, particularly those including open fields, present several unpredictable and varied conditions that hinder effective management and operational efficiency [1,2]. To tackle these challenges, human-robot collaboration has been suggested towards achieving common objectives through effective information sharing and task coordination [3–5]. During human-robot interaction (HRI), robots need to understand human intentions and respond appropriately. This can be achieved through human activity recognition (HAR), which involves the use of wearable sensors, computer vision, and machine learning (ML) to classify human activities [6–8]. Consequently, robots can synchronize their operations with human actions, allowing them to work alongside farmers and assisting with tasks such as weeding, harvesting, and transporting crops to storage [9,10].

Studies on HAR are still scarce in agriculture, mainly due to the complex nature of the environments they involve. Furthermore, the lack of standardized datasets and benchmarks for agricultural HAR research has hindered progress in this field [11]. Motivated by the necessity of creating natural communication systems for facilitating HRI in agriculture,

a vision-based static hand gesture recognition framework was developed in [12]. This framework was successfully tested in an open-field scenario involving an Unmanned Ground Vehicle (UGV) following the participants as they harvested crops and helping transfer crates from the field to a specific area. For the purpose of overcoming the challenges of static gestures, including their limited ability to capture the full range of human intentions in dynamic situations, Moysiadis et al. [9] developed a dynamic movement recognition HRI system. This advancement provided more detailed insights by integrating a wider range of body movements and variations, allowing for more accurate and natural interactions between humans and robots in a similar field scenario with [12]. In both studies, however, as they relied on data coming from an RGB-D camera mounted on the UGV, recognizing human movements accurately in different environments, like with different lighting or backgrounds, proved to be a considerable challenge.

To address the challenges associated with vision-based HAR, wearable sensors can provide a trustworthy alternative option [13–15]. These sensors offer additional layers of information, enabling more robust activity recognition, even in conditions with poor lighting or occlusions. Nevertheless, challenges still remain regarding mainly the fusion and synchronization of multi-sensor data, possible sensor drift, and the increased computational complexity [6,16]. Hand gesture recognition, through data originated from sensors embedded within a specially designed glove, was accomplished and successfully tested in [17] towards controlling wirelessly a robotic arm for removing weeds. In addition, Patil et al. [18] presented a wearable shirt with a smartphone attached to it, allowing the measurement of acceleration signals during several material handling activities and tested the performance of different ML classifiers. In a similar vein, Sharma et al. [19–21] used accelerometer data along with data from a microphone and global positioning system (GPS) from smartphones and evaluated the performance of several ML algorithms. Towards classifying specific agricultural worker's activities, Aiello et al. [22] used two accelerometers fixed to the wrists of operators of vibrating agricultural tools and a k-nearest neighbors (KNN) classifier. In [23], data collection field experiments were conducted to acquire data using wearable sensors throughout a human–robot collaborative harvesting task using two different UGVs for ergonomic purposes. In [24], the obtained signals from wearable sensors were fed into a long short-term memory (LSTM) network for HAR.

Existing studies, although limited in number, highlight the efficacy of wearable sensors in capturing relevant data for HAR in agricultural contexts. However, there is a notable lack of consensus on the most effective anatomical locations for sensor placement, while also the fusion of data from multiple sensors (e.g., accelerometers, gyroscopes, and magnetometers) has not been fully explored. Our research aims to fill these gaps by systematically investigating the optimal anatomical locations for sensor placement, through exploring the effectiveness of three types of sensors (accelerometers, gyroscopes, and magnetometers) and optimizing an ML-based framework for HAR. To acquire the essential sensor data, experimental field tests were carried out involving twenty participants wearing five inertial measurement units (IMUs) positioned on different parts of the body and carrying out six well-defined activities of a collaborative harvesting scenario similar to [9,12]. An LSTM neural network was used for classifying the activity signatures of the participants, while the dataset is shared publicly in [25].

In conclusion, our approach for a strategic selection and placement of sensors ensures that the most informative signals are captured, which is particularly important in agricultural tasks, where activities can be complex and varied. This approach not only improves accuracy, but also reduces computational costs, minimizes the amount of data needed, and lowers overall system costs. Although the primary objective of our work is to improve HAR in agricultural tasks, with a particular focus on human-robot collaboration, its methodology and key findings could also be applied to other domains. Finally, by providing access to these data, we aim to encourage further research and innovation, accelerating progress and overcoming the barriers of data scarcity that have historically limited the scope of agricultural studies [26,27].

The remainder of the paper is organized as follows: Section 2 outlines the methodology for data collection, pre-processing, and the developed LSTM-based HAR. Section 3 presents the main results regarding the optimal sensor placement and multimodal sensor fusion, while in Section 4 the results are discussed from a broader perspective along with future research directions. Finally, Section 5 concludes with the key findings.

2. Materials and Methods

The overall workflow of the LSTM-based HAR framework used in this analysis is illustrated in Figure 1, and an explanation of the key steps is provided in the following subsections. The workflow begins with data acquisition, where sensor data are collected from various body locations. In the next stage, various data pre-processing techniques are implemented to ensure that the raw data are cleaned and structured, while more informative features are also extracted to enhance model learning. The resulting dataset is then split into training and test sets. Training involves the model learning patterns and relationships within the data that can be used to make predictions. Cross-validation is conducted to find the best model parameters. After training and tuning the model, its performance is evaluated on the test data with metrics like the F1-score and cross-entropy loss function.

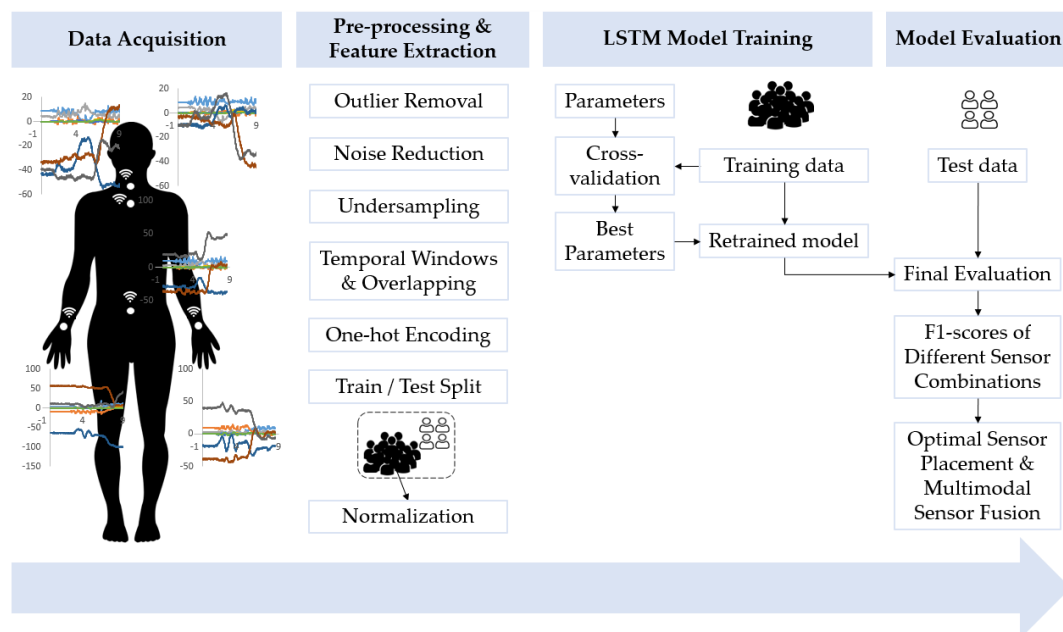


Figure 1. Overview of the proposed machine learning-based human activity recognition framework workflow using wearable sensor data.

2.1. Data Acquisition

The experimental sessions were conducted on a farm located in the region of Volos, central Greece. They involved twenty participants, evenly split between males and females. The average age of the participants was 30.13 years (with a standard deviation (SD) of approximately 4.13 years), their average height was 1.71 m (SD \approx 0.10 m), and their average weight was 70.20 kg (SD \approx 16.10 kg). To participate in this study, all subjects were required to have no history of surgeries or musculoskeletal injuries within the last year that could potentially affect their performance. Each participant provided informed consent, which was approved by the Institutional Ethical Committee, prior to the commencement of any experimental procedures.

Each participant was required to: (a) Remain still until the start signal was given; (b) Walk straight 3.5 m without a crate; (c) Bend down to pick up the crate; (d) Lift it from the ground to an upright stance; (e) Walk back 3.5 m while carrying the crate; and

(f) Position the crate onto a UGV. For the experimental setup, a UGV, known as Thorvald (SAGA Robotics SA, Oslo, Norway) was utilized, having a crate deposit height equal to 80 cm. This robot is commonly used in agricultural environments, due to its versatility and ability to navigate various types of terrain, making it a reliable choice for research and practical agricultural applications [4,28]. The participants were required to handle either an empty crate (with a tare weight of 1.5 kg) or a crate loaded with weight plates to achieve a total mass equivalent to 20% of the participant's body weight [24,29]. The weight plates, available in 1 kg and 2.5 kg increments, allowed for easy adjustment to the required weight. The open plastic crates used in this study had handles positioned 28 cm above the base and dimensions of 31 cm (height) × 53 cm (width) × 35 cm (depth).

Each participant performed both sub-cases (empty crate and loaded crate) three times in a randomized order, moving at their own pace. To minimize the risk of injury, all participants were instructed to perform a five-minute warm-up before beginning the task. The inclusion of a diverse group of participants, varying in gender, age, weight, and height, was intentional to ensure that the collected data captured a wide range of variability. This variability is essential for training ML models that can accurately identify activities under different conditions.

As far as the sensors used to acquire the essential data are concerned, five IMUs (Blue Trident, Vicon, Nexus, Oxford, UK) were implemented, which have been extensively used in relative literature [16,23,30]. Considering the number of available IMUs, we prioritized placing sensors on the upper body and core for effectively recognizing the specific activities performed in this study. These sensors were attached to specific body locations: the chest (over the breastbone), cervix (near the T1 vertebra), lumbar region (near the L4 vertebra), and both wrists. The sensors on the wrists were secured using special Velcro straps, while the other sensors were attached using double-sided tape. Each IMU was equipped with a tri-axial accelerometer, a tri-axial gyroscope, and a tri-axial magnetometer, providing detailed motion data across multiple axes. During the experiments, a sampling frequency of 50 Hz was utilized, which is considered sufficient for capturing the dynamics of the tasks being performed [23,24].

In brief, the chest provides a central view of overall body dynamics and posture, while the cervix and the lumbar region offer insights into upper and lower body motion and alignment, fundamental for activities like bending and lifting, where the spinal region plays a significant role in maintaining posture and balance [31,32]. The wrist sensors were selected for detecting hand and arm movements, such as when participants were grasping, lifting, and carrying the crate [33]. When standing, sensors on the chest, cervix, and lumbar region can record minimal acceleration and angular velocity, indicating stability with low variability [34]. During walking, the sensors can capture rhythmic patterns of acceleration and changes in angular velocity at the chest, reflecting the gait cycle [35]. The cervix and lumbar region sensors recorded dynamic posture shifts and alignment changes, while the wrist sensors recorded the movement of the arms.

The end-to-end workflow for capturing human activity signatures using IMUs is illustrated in Figure 2. The left section of the graph shows a view of the IMU's internal components, while the central section demonstrates the placement of the sensors on the human body. The data capture process is then depicted, showing also how the IMUs are synchronized through the Capture.U 1.4.1 software [36], installed on an Apple iPad mini (64 GB). The required data (CSV files) are saved directly to the five IMUs, each of which contains information about the specific location of the body, and connected to a computer for supporting further analysis. Capture.U software, when paired with the iPad, enabled simultaneous video recording of the ongoing experiments. This capability was especially valuable for manually differentiating between activities and identifying critical moments of transition between them. Following the methodology of [23,24], each activity begins with the subject standing still (designated as activity "0"), which serves a dual purpose: to establish a clear "idle" activity baseline and to enable accurate synchronization of the sensors before commencing the sequence. Subsequently, activity "1" begins as one

foot leaves the ground, marking the start of the stance phase of gait. Activity “2” starts when the participant begins bending their trunk, kneeling, or doing both (known as stoop, squat, and semi-squat techniques [37]) to approach the crate. Activity “3” begins when the participant starts lifting the crate from the ground. Activity “4” starts when the participant enters the stance phase of gait while carrying the crate. Finally, activity “5” begins with bending, kneeling, or both actions and ends when the crate is fully placed onto the UGV. The continuous nature of these tasks means that the start of one activity marks the end of the previous one. Adhering to these criteria was vital for reliable results.

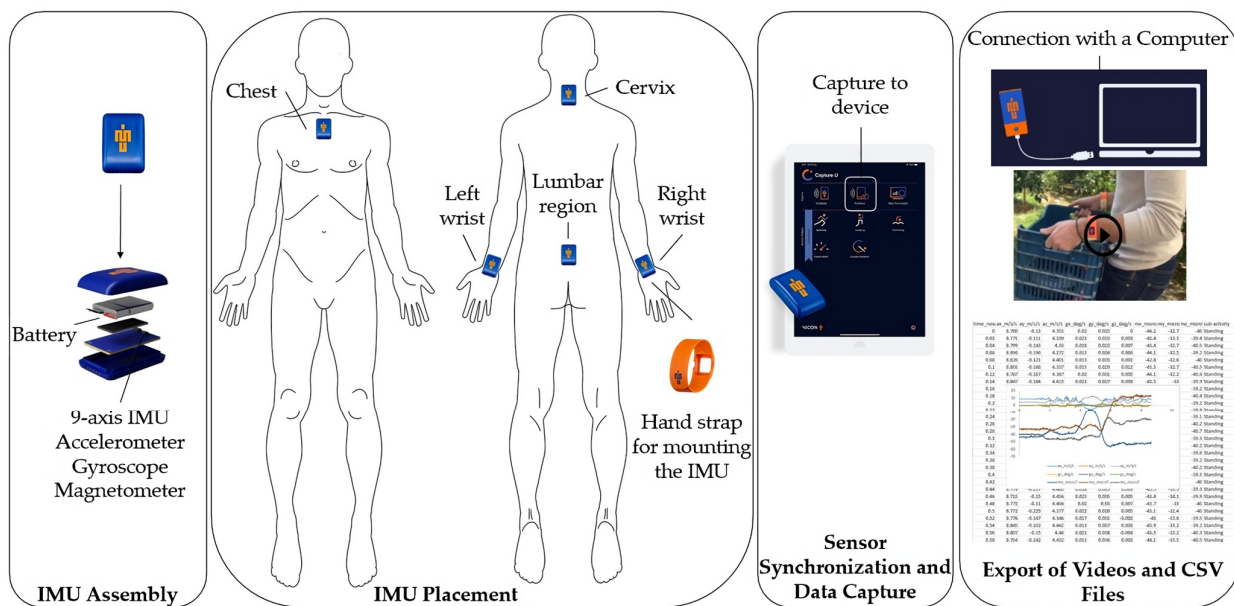


Figure 2. End-to-end workflow of setting up IMU sensors [38] on a human subject, synchronizing the sensors, capturing the data, and exporting them for detailed analysis.

2.2. Data Pre-Processing and Feature Extraction

Data pre-processing and feature extraction are closely related sequential steps in the data preparation pipeline for time series analysis. Data pre-processing focuses on cleaning, preparing, and transforming the raw data into a usable format, while feature extraction comes next to create additional features that are fed into the ML model to capture underlying patterns more effectively.

2.2.1. Handling of Outliers and Unsynchronized Sensor Data

During experiments, sensors may intermittently fail to record data, causing interruptions in the dataset. Hardware malfunctions may also lead to irregular values in the dataset. To address the former issue, any unsynchronized sensor data during processing (pointed out by Capture.U) led to the exclusion of measurements from the remaining sensors to ensure reliable and consistent results throughout the analysis. Additionally, any outliers in the dataset were identified and removed in the early stages through a statistical z-score technique (involving how many standard deviations (SDs) the data are from the mean, with values beyond 3 SDs considered outliers [39]). In the present analysis, the most common challenge was the non-synchronization of the sensors, while outliers were very limited in number and most of the time could be removed manually.

2.2.2. Noise Reduction

Signals data usually contain unwanted components resulting from sensor drift over time caused by the wearer’s movements. To ensure accurate analysis, calibration adjustments and filtering techniques are commonly used. As a means of removing the noise of the captured data, a median filter with eleven taps was used [24]. The median filter effectively

smooths data by replacing each data point with the median value of its surrounding points. This technique helps to reduce noise while preserving significant details and features within the dataset [40].

2.2.3. Activity Count and Class Imbalance in Sensor Data

The duration of the investigated activities varied significantly. Specifically, “walking with the crate” and “walking without the crate” took considerably more time compared to other activities, including “standing still”, “bending” to approach the crate, “lifting crate”, and “placing crate” onto the UGV. The duration of each activity directly impacts the classification imbalance (Figure 3a), due to the nature of the data collection process. Consequently, longer activities generate more sensor data points than shorter activities. Furthermore, our approach of excluding all sensor data when any dataset issue was detected led to a balanced dataset for all body parts, as illustrated in Figure 3b. To handle this imbalance, an under-sampling technique was used, similarly to [24].

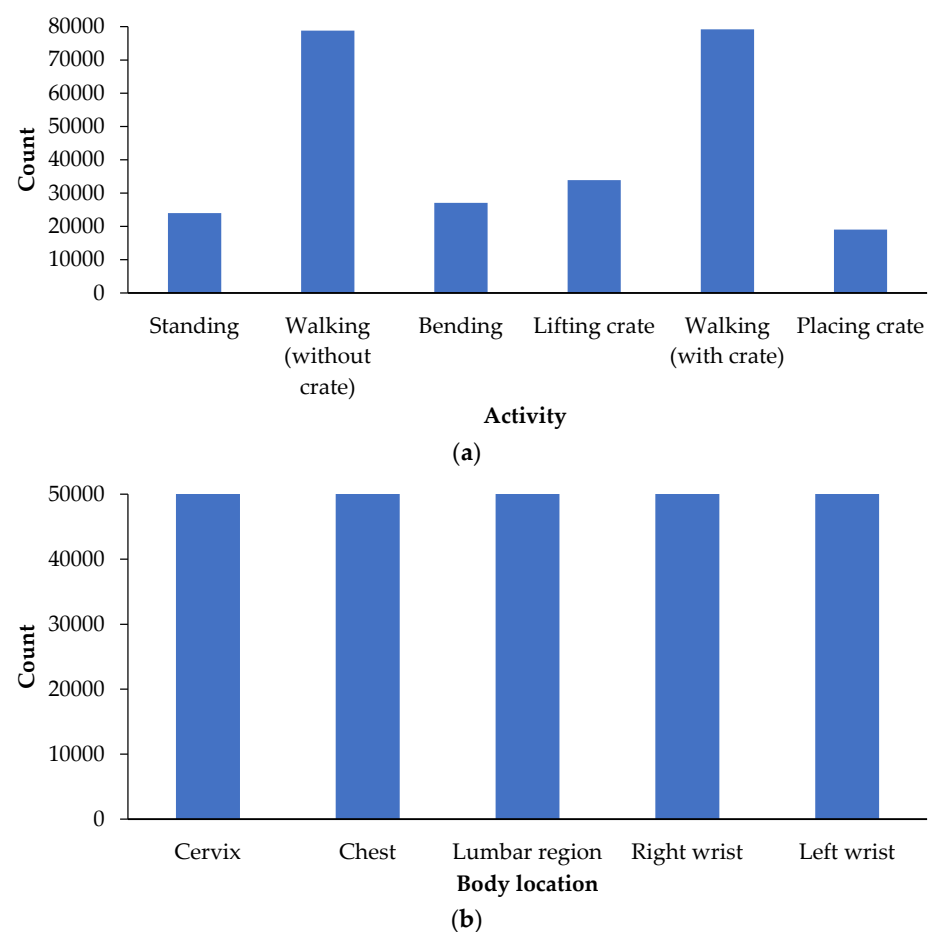


Figure 3. Histograms showing counts of each: (a) activity from all five locations and (b) body location on the present dataset.

2.2.4. Temporal Window Selection and Overlapping Approach

In accordance with the methodology detailed in [24], a temporal window length of 2 s was set after thoroughly evaluating several potential durations. Each temporal window was labeled with the activity being performed during that interval. To ensure sufficient coverage, the temporal windows were overlapped; specifically, each window began in the middle of the previous one, resulting in a 50% overlap between consecutive windows.

2.2.5. Encoding Categorical Data

One-hot vectors were used to represent categorical variables in a way that avoids introducing unintended relationships or order among categories. When categorical data are encoded as integers, it can falsely imply an ordinal relationship. One-hot encoding addresses this issue by assigning each category a unique binary vector where all categories are treated equally. This approach is compatible with many ML algorithms that require numerical input, converting categorical data into a format that algorithms can process without bias [41]. Specifically:

- Standing activity with a “0” assigned value got a [1,0,0,0,0] one-hot vector;
- Walking (without crate) with a “1” assigned value got a [0,1,0,0,0] one-hot vector;
- Bending with a “2” assigned value got a [0,0,1,0,0] one-hot vector;
- Lifting crate with a “3” assigned value got a [0,0,0,1,0] one-hot vector;
- Walking (with crate) with a “4” assigned value got a [0,0,0,0,1] one-hot vector;
- Placing crate (with crate) with a “4” assigned value got a [0,0,0,0,1] one-hot vector.

2.2.6. Train/Test Split

The resulting dataset was divided into two segments: (a) a training portion, which includes the examples used to train the model, and (b) a testing portion, which contains the examples used to assess the model’s performance. Given the extensive data collected during the experimentation phase, an 80/20 split was chosen for the training and testing datasets, similarly to [24]. This split was carried out at the subject level to ensure that the model’s performance could be evaluated on the unique movement characteristics of unseen subjects. For testing, data from four randomly selected subjects were used to have the trained model predict their recorded activities. The remaining data from the sixteen subjects were reserved for training the ML algorithm.

2.2.7. Feature Scaling

Feature scaling is a crucial step in data pre-processing aimed at normalizing data to a specific range. This process is essential for speeding up calculations within ML algorithms. When working with frameworks like Scikit-learn, feature scaling is necessary, especially when the dataset contains variables with differing scales. In this analysis, feature scaling was applied to adjust the dataset, making it more compatible with ML methods. Among various scaling techniques available, the StandardScaler() was utilized [42] exclusively to the training dataset to avoid any information leakage into the test dataset. This scaler standardizes the dataset by transforming it so that the resulting distribution has a mean of zero and an SD of one. The transformation is achieved by subtracting the mean from the original value and then dividing by the SD:

$$z = \frac{x - \mu}{SD}. \quad (1)$$

In the above equation, z represents the transformed feature value, x denotes the original value, μ is the mean, and SD stands for the standard deviation of the training samples.

2.3. LSTM Model Training and Evaluation

LSTM is a type of recurrent neural network (RNN) commonly used for feature recognition in time-dependent data, as it can capture long-term dependencies [43,44]. Its gating mechanisms allow it to selectively remember or forget information, leading to the enhancement of memory process [45] and making it effective in tasks like HAR, where understanding patterns over time is of central importance [46].

The LSTM model was constructed using TensorFlow’s Keras API, featuring two LSTM layers with 50 units each, a dense output layer with six units and a softmax activation function. The model was trained to capture temporal patterns in the data using the Adam optimizer, after first testing the impact on model performance of the several optimizers

(Adam, RMSprop, SGD, Adagrad, and Adadelata). In addition, the cross-entropy loss function was utilized to quantify the error between the predicted values and the actual values, while a low learning rate of 0.001 was employed to improve the model's fitting capability. A range of batch sizes was also tested to ensure both high F1-scores and low generalization gaps. This approach aimed to balance the model's ability to accurately classify the data (as indicated by high F1-scores) while minimizing overfitting, as reflected in low generalization gaps between the training and testing datasets.

For the purpose of avoiding the dependence of the results on a specific random choice, the trained model was validated using a 10-fold cross-validation. Thus, the training set was split into 10 smaller sets. The LSTM model was trained using nine of the folds as training data, while the resulting model was validated upon the rest of the data (i.e., they were utilized as test data to calculate performance measures like F1-score). The performance metrics provided by the 10-fold cross-validation were the average of the calculated values during the loop [47]. A flowchart outlining the cross-validation workflow in model training is provided in Figure 1. It begins with the dataset in the form made after data pre-processing and feature extraction, which is split into training and test data. The parameters for the model are initialized, and cross-validation is performed only on the training data to tune these parameters. Based on the cross-validation results, the best parameters are identified, and the model is retrained using the optimal parameters. Finally, the test data are evaluated on the retrained model to produce a final evaluation of its performance using appropriate metrics. Early stopping was applied to prevent overfitting by halting training when the validation loss did not improve for 10 consecutive epochs.

Finally, in order to identify the most effective combinations of anatomical locations for the IMUs, a systematic analysis was performed. This involved generating all possible combinations of the five anatomical locations—cervix, chest, lumbar region, right wrist, and left wrist. Additionally, all possible combinations of sensor types were evaluated. This included single sensor types (e.g., accelerometers only), as well as combinations of multiple sensors (e.g., accelerometers and gyroscopes, or accelerometers, gyroscopes, and magnetometers). F1-score, preferred in datasets with imbalanced classes [48], was calculated for each set to determine which combinations provided the best classification accuracy.

3. Results

3.1. Effect of Hyper-Parameters on Model Performance

Hyper-parameters play a crucial role in shaping the model's learning process of LSTM models. Their tuning can significantly impact the model's accuracy, generalization ability, and computational efficiency. The set of key hyper-parameters includes the number of LSTM units, learning rate, dropout rate, activation function, optimizer, and batch size. In brief, increasing the number of units initially enhanced model performance, but beyond a certain point overfitting risks increased, with 50 units balancing the model's capacity and generalization. Moreover, a learning rate of 0.001, a dropout rate of 0.4, the Tanh activation function, the Adam optimizer, and a batch size of 200 provided stable and efficient training.

Indicatively, Figure 4 illustrates the relationship between the batch size and two key metrics; F1-score and generalization gap when all types of sensors (i.e., accelerometers, gyroscopes, and magnetometers) were placed in all available body positions. F1-score is a measure of model performance with higher values indicating better overall accuracy, whereas the generalization gap is the difference between a model's performance on training data and its performance on unseen test data. A smaller generalization gap implies better model generalization, meaning that the model is less likely to overfit to the training data. Based on Figure 4, choosing a batch size of 200 is justified as it achieves a strong balance between high F1-score and effective generalization. F1-score seems to reach a plateau or slightly decrease after a batch size of around 200. This suggests that increasing the batch size beyond 200 might not yield significant improvements in model performance. In turn, the generalization gap is significantly lower than smaller batch sizes like 32, 64, and 128. Additionally, while larger batch sizes can improve training speed, they also require more

memory and computational efficiency [48]. Hence, a batch size of 200 offers an optimal trade-off between performance, generalization, and computational efficiency making it a well-rounded choice.

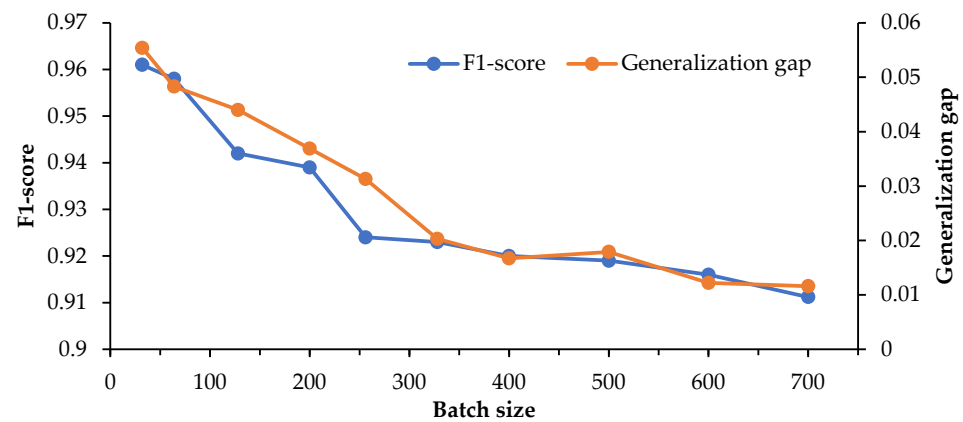


Figure 4. Effect of batch size on model performance.

3.2. Sensor Placement Combinations

Given the present set of five anatomical locations (cervix, chest, lumbar region, right wrist, and left wrist), the total number of combinations is:

$$C(5, 1) + C(5, 2) + C(5, 3) + C(5, 4) + C(5, 5) = 31, \quad (2)$$

where $C(n, k) = \frac{n!}{k!(n-k)!}$, with n being the total number of anatomical locations (5 in the present analysis), k is the number of anatomical locations to choose, and $!$ denotes factorial, namely the product of all positive integers up to that number. Table 1 presents the F1-scores corresponding to different combinations of sensor placements on various anatomical locations, by considering the data fusion of all sensors that IMUs include; accelerometer, gyroscope, and magnetometer.

Interestingly, the highest F1-score of 0.939 is achieved with a single IMU placed on the chest, indicating this location's superior effectiveness in capturing the necessary data for the present activities. However, the cervix and lumbar region are nearly as effective as the chest for sensor placement, potentially offering practical alternatives. In contrast, the wrist-mounted sensors, especially the left wrist, underperform compared to the chest-mounted sensor, suggesting that they are not as effective in capturing the relevant activities. The wrists' greater variability in movement patterns contributes to the difficulty of accurate activity recognition. Unlike the torso, which is directly involved in most activities, the wrists have a more indirect role, leading to inconsistent data and noise. This makes it challenging to identify clear patterns, as wrist movements can sometimes resemble those of other activities or involve irrelevant motions, such as arm swinging during standing or walking without a crate. Besides, activities like carrying a crate introduce further complexity through varying wrist orientations and different carrying techniques, adding task-specific variability that complicates classification.

For the two anatomical location combinations, the F1-scores tend to be slightly lower than the highest score achieved with the single chest IMU. The combination of the chest with lumbar region and cervix yields the highest F1-scores of 0.932 and 0.931, respectively, among the dual placements, indicating that these sites are effective in capturing reliable data. Conversely, combinations involving the left wrist generally lead to lower F1-scores, such as right wrist and left wrist at 0.898.

Concerning the three anatomical location combinations, the F1-scores demonstrate varied effectiveness, with some combinations achieving results close to those seen with two sensors. The combination of cervix, chest, and lumbar region yields the highest F1-

score of 0.931. However, not all three-IMU setups enhance data accuracy, and some may even introduce redundancy or noise, reducing overall performance.

Table 1. Sensor placement combinations and corresponding F1-scores by considering data fusion of accelerometer, gyroscope, and magnetometer signals.

Combinations	F1-Score
1 anatomical location	
Cervix	0.933
Chest	0.939
Lumbar region	0.932
Right wrist	0.919
Left wrist	0.888
2 anatomical locations	
Cervix, Chest	0.931
Cervix, Lumbar region	0.927
Cervix, Right wrist	0.915
Cervix, Left wrist	0.906
Chest, Lumbar region	0.932
Chest, Right wrist	0.919
Chest, Left wrist	0.908
Lumbar region, Right wrist	0.922
Lumbar region, Left wrist	0.904
Right wrist, Left wrist	0.898
3 anatomical locations	
Cervix, Chest, Lumbar region	0.931
Cervix, Chest, Right wrist	0.918
Cervix, Chest, Left wrist	0.913
Cervix, Lumbar region, Right wrist	0.920
Cervix, Lumbar region, Left wrist	0.909
Cervix, Right wrist, Left wrist	0.901
Chest, Lumbar region, Right wrist	0.924
Chest, Lumbar region, Left wrist	0.909
Chest, Right wrist, Left wrist	0.894
Lumbar region, Right wrist, Left wrist	0.904
4 anatomical locations	
Cervix, Chest, Lumbar region, Right wrist	0.922
Cervix, Chest, Lumbar region, Left wrist	0.914
Cervix, Chest, Right wrist, Left wrist	0.906
Cervix, Lumbar region, Right wrist, Left wrist	0.905
Chest, Lumbar region, Right wrist, Left wrist	0.909
5 anatomical locations	
Cervix, Chest, Lumbar region, Right wrist, Left wrist	0.908

For the four anatomical location combinations, the F1-scores reflect a slight decline compared to some of the three-location setups. The combination of chest, cervix, lumbar region, and right wrist achieves the highest F1-score of 0.922 among the four-location configurations, suggesting that this setup captures the examined activities sufficiently. These results also indicate diminishing returns with the inclusion of the left wrist, as expected.

Finally, the combination of all five anatomical locations provides an F1-score equal to 0.908, which is 3.4% lower than the corresponding metric of the single sensor positioned on the chest. Consequently, the inclusion of all five sensors seems to introduce unnecessary complexity, which can lead to a decrease in the effectiveness of the LSTM model. This reinforces the idea that strategic selection of fewer, more impactful sensor locations can be more beneficial than using many locations.

In summary, the single chest IMU consistently outperforms other placements, demonstrating its superiority in capturing the necessary data for the investigated activities. Indicatively, Tables 2 and 3 provide the confusion matrices of the proposed LSTM model by considering all the anatomical locations and only the IMU at the chest, respectively. This comparison highlights the practicality of achieving high accuracy with fewer sensors, thereby reducing system complexity and cost. These confusion matrices offer a thorough breakdown of the model's predictions towards identifying activities where the model is performing well or poorly. The LSTM network with both all anatomical locations and using only the chest IMU demonstrate considerable overall performance, as shown by the diagonal dominance in the confusion matrices. Nevertheless, some classes, such as "Bending" and "Lifting crate", appear to exhibit slightly higher rates of misclassification. This could be attributed to the inherent similarity between these activities [24]. The more zeros appearing in Table 3 indicate that the LSTM model based on the data provided by the IMU on the chest is performing better in accurately classifying the data and minimizing errors.

Table 2. Confusion matrix of the proposed LSTM network by considering all the anatomical locations.

Confusion Matrix		Predicted Label					
		Standing	Walking (without Crate)	Bending	Lifting Crate	Walking (with Crate)	Placing Crate
True label	Standing	4456	497	3	3	0	0
	Walking (without crate)	548	14,685	303	59	42	2
	Bending	5	524	4385	350	31	0
	Lifting crate	3	111	366	5950	425	0
	Walking (with crate)	0	37	20	260	15,280	394
	Placing crate	2	18	1	2	767	2979

Table 3. Confusion matrix of the proposed LSTM network by considering only the IMU at the chest.

Confusion Matrix		Predicted Label					
		Standing	Walking (without Crate)	Bending	Lifting Crate	Walking (with Crate)	Placing Crate
True label	Standing	934	60	0	0	0	0
	Walking (without crate)	127	2947	39	4	0	0
	Bending	0	33	1043	43	0	0
	Lifting crate	0	4	46	1238	48	0
	Walking (with crate)	0	0	0	23	3064	83
	Placing crate	0	0	0	0	126	640

3.3. Multimodal Sensor Fusion

As proved in the previous section, the IMU placed on the chest was the most effective sensor location for the LSTM model. Its superior performance is attributed to its strategic location on the torso, which provides stable, consistent data relevant to many upper-body activities. Hence, its central placement helps in accurately capturing core movements and reducing interference from other body parts, resulting in better performance in activity classification. Focusing on this case consideration, a characteristic plot of the training and validation loss decreasing over the epochs is depicted in Figure 5. The very low generalization gap, evidenced by the close alignment of training and validation loss curves, indicates that the model is performing well both on the training data and on unseen validation data. Furthermore, the smooth and decreasing curves demonstrate that the model is learning effectively and making consistent progress in minimizing loss over time.

As detailed in Section 2.1, this study considers signals acquired from accelerometer, gyroscope, and magnetometer sensors. These sensors are all integral components of the IMUs utilized in this research. Sensor fusion was applied towards combining results for multimodal data similarly to studies like [49]. For the evaluation, F1-score was used in

order to guarantee the fairness and consistency of the subsequent comparisons, as it is a useful performance indicator for imbalanced classes like the present one [24,48]. In addition, only the IMU at the chest was considered in this section, as it proved to yield the best model performance compared with the other body positions.

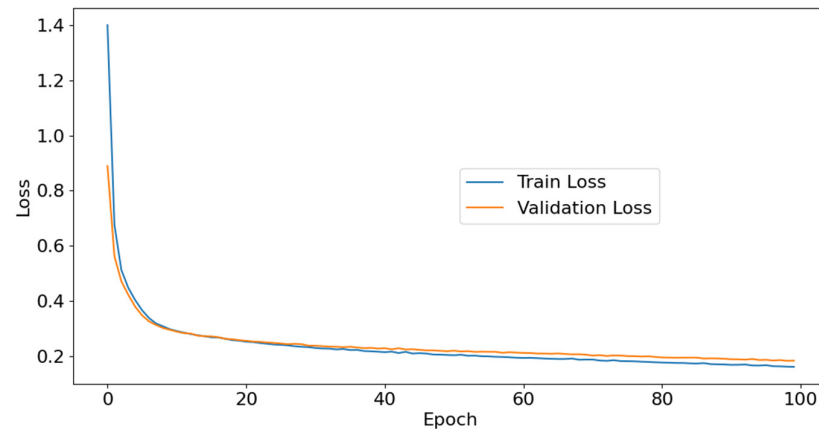


Figure 5. Training and validation plot loss, considering only the IMU at the chest.

First, the overall performance of single sensor modality was assessed for predicting the investigated activities. As depicted in Table 4, the magnetometer consistently outperforms the accelerometer and gyroscope in terms of F1-score, by showing an F1-score equal to 0.783. The lowest performance was achieved by taking into account only the accelerometers. Combining data from multiple sensors proved to lead to improved performance compared to using a single sensor, stressing the value of multimodal sensor fusion in enhancing classification accuracy. Indicatively, the records of the magnetometer when fused with those of either accelerometers or gyroscopes provided F1-scores equal to 0.887 or 0.915, respectively, whereas the combination of accelerometers and gyroscopes resulted in the lowest model performance (F1-score equal to 0.774). The highest F1-score of 0.939 was achieved by combining all three sensors, indicating that the complementary information from the accelerometer, gyroscope, and magnetometer is very important for accurate activity classification.

Table 4. Multimodal sensor fusion, considering only the IMU at the chest, and corresponding F1-scores.

Accelerometer	Gyroscope	Magnetometer	F1-Score
✓			0.554
	✓		0.637
		✓	0.783
✓	✓		0.774
✓		✓	0.887
	✓	✓	0.915
✓	✓	✓	0.939

4. Discussion

This study deals with the impact of sensor placement and multimodal sensor fusion on the performance of an LSTM-based model for activity classification. The investigated activities are related to a collaborative human-robot scenario in which a UGV follows workers during the harvesting process to transport full crates out of the field, as presented in [9,12]. Overall, by optimizing key hyper-parameters, the model achieved high accuracy in classifying activities.

In terms of sensor placement, the results suggest the chest as the most effective anatomical location for capturing the necessary data for activity classification, achieving the highest F1-score of 0.939. Its stable position on the torso ensures consistent and accurate

data gathering, principally for movements involving the upper body. Moreover, the chest is less prone to interference compared to the wrists, for example, making it ideal for reliable HAR associated with the present experimental setup. Combinations of the chest with the lumbar region or cervix also performed well, although they did not surpass the single chest sensor's F1-score. Conversely, the inclusion of wrist-mounted sensors, particularly on the left wrist, generally resulted in lower F1-scores, suggesting that these placements are less effective for the tasks considered in this study. This can be attributed to dominance and usage patterns, since 90% of the participants had the right hand as dominant, leading to more intense and frequent movements on the right side [50,51]. In contrast, the left wrist, being less dominant, may experience less consistent movement patterns, resulting in less reliable data.

Multimodal sensor fusion demonstrated the value of combining data from different sensor types to enhance classification accuracy. Specifically, the magnetometer outperformed the accelerometer and gyroscope, emphasizing its ability to capture critical orientation and motion information that the other sensors might miss. However, the highest performance was achieved by fusing data from all three sensors. This all-inclusive approach exploited the unique strengths of each sensor; the accelerometer's measurement of linear acceleration, the gyroscope's detection of rotational movements, and the magnetometer's ability to sense changes in orientation or location. Hence, by integrating these diverse data sources, the LSTM model benefited from a more complete representation of the activities, leading to improved HAR.

Given the promising results achieved in this study, future research could incorporate a more diverse dataset with a wider range of agricultural material handling activities and participant demographics that have the potential to improve generalizability and robustness of the present LSTM model. Also, investigating other sensor types, including electromyography (EMG) and electrocardiogram (ECG) for measuring muscle response and heart rate variability, respectively, could provide complementary information and potentially improve classification accuracy for specific tasks. Experimenting with different combinations of sensor placements, like legs, shoulders and elbows, might identify other most effective combinations for various use cases, leveraging both the upper and lower body data points. It is also essential to conduct real-world agricultural collaborative robotics experiments and collect feedback from workers to ensure that these systems meet practical needs and address any usability issues. Finally, long-term studies to evaluate the adaptability of these collaborative systems over extended periods, in conjunction with workers' perspectives, are also important for guaranteeing their viability in real-world agricultural settings.

From a broader perspective, this study advances the field of HRI by integrating insights from sensor technology and data fusion to enhance HAR in agricultural settings. As agricultural practices increasingly embrace robotization [52], these contributions are instrumental for improving the efficiency and safety of agricultural operations [53,54]. They align with the broader goal of developing adaptive and intelligent robotic systems that can seamlessly interact with human workers. Additionally, the insights gained from this study offer a framework for future human-centric, cost-effective innovations across various domains, making technological progress more accessible and beneficial in everyday work environments.

5. Conclusions

In conclusion, this research demonstrated that strategic sensor placement and effective integration of diverse sensor data can significantly enhance the model's accuracy. The key findings are summarized as follows:

- Model performance: overall, the present LSTM-based model demonstrated high accuracy in classifying the investigated activities;
- Optimal sensor placement: by prioritizing sensor placement on the upper body to effectively capture the examined activities, we found that the chest was the most

effective anatomical location for activity classification, achieving the highest F1-score of 0.939.

- Multimodal sensor fusion: fusing data from all sensors, namely accelerometers, gyroscopes, and magnetometers, substantially enhanced classification accuracy.

Author Contributions: Conceptualization, L.B., D.T. and D.B.; methodology, L.B. and D.T.; software, L.B.; validation, D.T. and A.C.T.; writing—original draft preparation, L.B.; writing—review and editing, L.B., D.T., A.C.T., D.K. and D.B.; visualization, L.B. and D.K.; supervision, D.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethical Committee under the identification code 1660 on 3 June 2020.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The dataset used in this work is publicly available in [25].

Conflicts of Interest: Author Dionysis Bochtis is employed by the company farmB Digital Agriculture S.A. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Bechar, A.; Vigneault, C. Agricultural robots for field operations: Concepts and components. *Biosyst. Eng.* **2016**, *149*, 94–111. [[CrossRef](#)]
2. Huo, D.; Malik, A.W.; Ravana, S.D.; Rahman, A.U.; Ahmedy, I. Mapping smart farming: Addressing agricultural challenges in data-driven era. *Renew. Sustain. Energy Rev.* **2024**, *189*, 113858. [[CrossRef](#)]
3. Benos, L.; Moysiadis, V.; Kateris, D.; Tagarakis, A.C.; Busato, P.; Pearson, S.; Bochtis, D. Human-Robot Interaction in Agriculture: A Systematic Review. *Sensors* **2023**, *23*, 6776. [[CrossRef](#)] [[PubMed](#)]
4. Lytridis, C.; Kaburlasos, V.G.; Pachidis, T.; Manios, M.; Vrochidou, E.; Kalampokas, T.; Chatzistamatis, S. An Overview of Cooperative Robotics in Agriculture. *Agronomy* **2021**, *11*, 1818. [[CrossRef](#)]
5. Vasconez, J.P.; Kantor, G.A.; Auat Cheein, F.A. Human-robot interaction in agriculture: A survey and current challenges. *Biosyst. Eng.* **2019**, *179*, 35–48. [[CrossRef](#)]
6. Liu, H.; Gamboa, H.; Schultz, T. Human Activity Recognition, Monitoring, and Analysis Facilitated by Novel and Widespread Applications of Sensors. *Sensors* **2024**, *24*, 5250. [[CrossRef](#)]
7. Bhola, G.; Vishwakarma, D.K. A review of vision-based indoor HAR: State-of-the-art, challenges, and future prospects. *Multimed. Tools Appl.* **2024**, *83*, 1965–2005. [[CrossRef](#)] [[PubMed](#)]
8. Donisi, L.; Cesarelli, G.; Pisani, N.; Ponsiglione, A.M.; Ricciardi, C.; Capodaglio, E. Wearable Sensors and Artificial Intelligence for Physical Ergonomics: A Systematic Review of Literature. *Diagnostics* **2022**, *12*, 3048. [[CrossRef](#)]
9. Moysiadis, V.; Benos, L.; Karras, G.; Kateris, D.; Peruzzi, A.; Berruto, R.; Papageorgiou, E.; Bochtis, D. Human-Robot Interaction through Dynamic Movement Recognition for Agricultural Environments. *AgriEngineering* **2024**, *6*, 2494–2512. [[CrossRef](#)]
10. Upadhyay, A.; Zhang, Y.; Koparan, C.; Rai, N.; Howatt, K.; Bajwa, S.; Sun, X. Advances in ground robotic technologies for site-specific weed management in precision agriculture: A review. *Comput. Electron. Agric.* **2024**, *225*, 109363. [[CrossRef](#)]
11. Tagarakis, A.C.; Benos, L.; Kyriakarakos, G.; Pearson, S.; Sørensen, C.G.; Bochtis, D. Digital Twins in Agriculture and Forestry: A Review. *Sensors* **2024**, *24*, 3117. [[CrossRef](#)] [[PubMed](#)]
12. Moysiadis, V.; Katikaridis, D.; Benos, L.; Busato, P.; Anagnostis, A.; Kateris, D.; Pearson, S.; Bochtis, D. An Integrated Real-Time Hand Gesture Recognition Framework for Human-Robot Interaction in Agriculture. *Appl. Sci.* **2022**, *12*, 8160. [[CrossRef](#)]
13. Han, C.; Zhang, L.; Tang, Y.; Huang, W.; Min, F.; He, J. Human activity recognition using wearable sensors by heterogeneous convolutional neural networks. *Expert Syst. Appl.* **2022**, *198*, 116764. [[CrossRef](#)]
14. Minh Dang, L.; Min, K.; Wang, H.; Jalil Piran, M.; Hee Lee, C.; Moon, H. Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognit.* **2020**, *108*, 107561. [[CrossRef](#)]
15. Bian, S.; Liu, M.; Zhou, B.; Lukowicz, P. The State-of-the-Art Sensing Techniques in Human Activity Recognition: A Survey. *Sensors* **2022**, *22*, 4596. [[CrossRef](#)]
16. Rana, M.; Mittal, V. Wearable Sensors for Real-Time Kinematics Analysis in Sports: A Review. *IEEE Sens. J.* **2021**, *21*, 1187–1207. [[CrossRef](#)]
17. Gokul, S.; Dhikshith, R.; Sundaresh, S.A.; Gopinath, M. Gesture Controlled Wireless Agricultural Weeding Robot. In Proceedings of the 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 15–16 March 2019; pp. 926–929.

18. Patil, P.A.; Jagyasi, B.G.; Raval, J.; Warke, N.; Vaidya, P.P. Design and development of wearable sensor textile for precision agriculture. In Proceedings of the 7th International Conference on Communication Systems and Networks (COMSNETS), Bangalore, India, 6–10 January 2015.
19. Sharma, S.; Raval, J.; Jagyasi, B. Mobile sensing for agriculture activities detection. In Proceedings of the IEEE Global Humanitarian Technology Conference (GHTC), GHTC 2013, San Jose, CA, USA, 20–23 October 2013; pp. 337–342.
20. Sharma, S.; Raval, J.; Jagyasi, B. Neural network based agriculture activity detection using mobile accelerometer sensors. In Proceedings of the Annual IEEE India Conference (INDICON), Pune, India, 11–13 December 2014.
21. Sharma, S.; Jagyasi, B.; Raval, J.; Patil, P. AgriAct: Agricultural Activity Training using multimedia and wearable sensing. In Proceedings of the IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops), St. Louis, MO, USA, 23–27 March 2015; pp. 439–444.
22. Aiello, G.; Catania, P.; Vallone, M.; Venticinque, M. Worker safety in agriculture 4.0: A new approach for mapping operator’s vibration risk through Machine Learning activity recognition. *Comput. Electron. Agric.* **2022**, *193*, 106637. [[CrossRef](#)]
23. Tagarakis, A.C.; Benos, L.; Aivazidou, E.; Anagnostis, A.; Kateris, D.; Bochtis, D. Wearable Sensors for Identifying Activity Signatures in Human-Robot Collaborative Agricultural Environments. *Eng. Proc.* **2021**, *9*, 5. [[CrossRef](#)]
24. Anagnostis, A.; Benos, L.; Tsaopoulos, D.; Tagarakis, A.; Tsolakis, N.; Bochtis, D. Human activity recognition through recurrent neural networks for human-robot interaction in agriculture. *Appl. Sci.* **2021**, *11*, 2188. [[CrossRef](#)]
25. Open Datasets—iBO. Available online: <https://ibo.certh.gr/open-datasets/> (accessed on 16 September 2024).
26. Rozenstein, O.; Cohen, Y.; Alchanatis, V.; Behrendt, K.; Bonfil, D.J.; Eshel, G.; Harari, A.; Harris, W.E.; Klapp, I.; Laor, Y.; et al. Data-driven agriculture and sustainable farming: Friends or foes? *Precis. Agric.* **2024**, *25*, 520–531. [[CrossRef](#)]
27. Atik, C. Towards Comprehensive European Agricultural Data Governance: Moving Beyond the “Data Ownership” Debate. *IIC-Int. Rev. Intellect. Prop. Compet. Law* **2022**, *53*, 701–742. [[CrossRef](#)]
28. Botta, A.; Cavallone, P.; Baglieri, L.; Colucci, G.; Tagliavini, L.; Quaglia, G. A Review of Robots, Perception, and Tasks in Precision Agriculture. *Appl. Mech.* **2022**, *3*, 830–854. [[CrossRef](#)]
29. Lavender, S.A.; Li, Y.C.; Andersson, G.B.; Natarajan, R.N. The effects of lifting speed on the peak external forward bending, lateral bending, and twisting spine moments. *Ergonomics* **1999**, *42*, 111–125. [[CrossRef](#)]
30. Winter, L.; Bellenger, C.; Grimshaw, P.; Crowther, R.G. Analysis of Movement Variability in Cycling: An Exploratory Study. *Sensors* **2023**, *23*, 4972. [[CrossRef](#)]
31. Jackie, D.; Zehr Danielle, R.; Carnegie, T.N.W.; Beach, T.A.C. A comparative analysis of lumbar spine mechanics during barbell- and crate-lifting: Implications for occupational lifting task assessments. *Int. J. Occup. Saf. Ergon.* **2020**, *26*, 1439872. [[CrossRef](#)]
32. Huysamen, K.; Power, V.; O’Sullivan, L. Elongation of the surface of the spine during lifting and lowering, and implications for design of an upper body industrial exoskeleton. *Appl. Ergon.* **2018**, *72*, 10–16. [[CrossRef](#)] [[PubMed](#)]
33. Hlucny, S.D.; Novak, D. Characterizing Human Box-Lifting Behavior Using Wearable Inertial Motion Sensors. *Sensors* **2020**, *20*, 2323. [[CrossRef](#)]
34. Ghislieri, M.; Gastaldi, L.; Pastorelli, S.; Tadano, S.; Agostini, V. Wearable Inertial Sensors to Assess Standing Balance: A Systematic Review. *Sensors* **2019**, *19*, 4075. [[CrossRef](#)]
35. Nazarahari, M.; Rouhani, H. Detection of daily postures and walking modalities using a single chest-mounted tri-axial accelerometer. *Med. Eng. Phys.* **2018**, *57*, 75–81. [[CrossRef](#)]
36. Capture U-ImeasureU. Available online: <https://imeasureu.com/capture-u/> (accessed on 19 August 2024).
37. Vecchio, L. Del Choosing a Lifting Posture: Squat, Semi-Squat or Stoop. *MOJ Yoga Phys. Ther.* **2017**, *2*, 56–62. [[CrossRef](#)]
38. VICON Blue Trident Inertial Measurement Unit. Available online: <https://www.vicon.com/hardware/blue-trident/> (accessed on 26 August 2024).
39. Sullivan, J.H.; Warkentin, M.; Wallace, L. So many ways for assessing outliers: What really works and does it matter? *J. Bus. Res.* **2021**, *132*, 530–543. [[CrossRef](#)]
40. Afsar, M.M.; Saqib, S.; Aladfaj, M.; Alatiyyah, M.H.; Alnowaiser, K.; Aljuaid, H.; Jalal, A.; Park, J. Body-Worn Sensors for Recognizing Physical Sports Activities in Exergaming via Deep Learning Model. *IEEE Access* **2023**, *11*, 12460–12473. [[CrossRef](#)]
41. Li, J.; Xu, Y.; Shi, H. Bidirectional LSTM with Hierarchical Attention for Text Classification. In Proceedings of the IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 20–22 December 2019; pp. 456–459.
42. Sklearn.Preprocessing. StandardScaler—Scikit-Learn 0.24.1 Documentation. Available online: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html> (accessed on 20 January 2021).
43. Hewamalage, H.; Bergmeir, C.; Bandara, K. Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. *Int. J. Forecast.* **2021**, *37*, 388–427. [[CrossRef](#)]
44. Rithani, M.; Kumar, R.P.; Doss, S. A review on big data based on deep neural network approaches. *Artif. Intell. Rev.* **2023**, *56*, 14765–14801. [[CrossRef](#)]
45. Landi, F.; Baraldi, L.; Cornia, M.; Cucchiara, R. Working Memory Connections for LSTM. *Neural Netw.* **2021**, *144*, 334–341. [[CrossRef](#)]
46. Hu, Y.; Zhang, X.-Q.; Xu, L.; Xian He, F.; Tian, Z.; She, W.; Liu, W. Harmonic Loss Function for Sensor-Based Human Activity Recognition Based on LSTM Recurrent Neural Networks. *IEEE Access* **2020**, *8*, 135617–135627. [[CrossRef](#)]

47. Scikit-Learn User Guide. Available online: https://scikit-learn.org/stable/modules/cross_validation.html (accessed on 12 September 2024).
48. Xia, K.; Huang, J.; Wang, H. LSTM-CNN Architecture for Human Activity Recognition. *IEEE Access* **2020**, *8*, 56855–56866. [[CrossRef](#)]
49. Chung, S.; Lim, J.; Noh, K.J.; Kim, G.; Jeong, H. Sensor Data Acquisition and Multimodal Sensor Fusion for Human Activity Recognition Using Deep Learning. *Sensors* **2019**, *19*, 1716. [[CrossRef](#)]
50. Nataraj, R.; Sanford, S.; Liu, M.; Harel, N.Y. Hand dominance in the performance and perceptions of virtual reach control. *Acta Psychol. (Amst.)* **2022**, *223*, 103494. [[CrossRef](#)]
51. Nazari, F.; Mohajer, N.; Nahavandi, D.; Khosravi, A.; Nahavandi, S. Comparison Study of Inertial Sensor Signal Combination for Human Activity Recognition based on Convolutional Neural Networks. In Proceedings of the 15th International Conference on Human System Interaction (HSI), Melbourne, Australia, 28–31 July 2022; pp. 1–6.
52. Marinoudi, V.; Benos, L.; Villa, C.C.; Lampridi, M.; Kateris, D.; Berruto, R.; Pearson, S.; Sørensen, C.G.; Bochtis, D. Adapting to the Agricultural Labor Market Shaped by Robotization. *Sustainability* **2024**, *16*, 7061. [[CrossRef](#)]
53. Benos, L.; Bechar, A.; Bochtis, D. Safety and ergonomics in human-robot interactive agricultural operations. *Biosyst. Eng.* **2020**, *200*, 55–72. [[CrossRef](#)]
54. Giallanza, A.; La Scalia, G.; Micale, R.; La Fata, C.M. Occupational health and safety issues in human-robot collaboration: State of the art and open challenges. *Saf. Sci.* **2024**, *169*, 106313. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.