A Residual Deep Learning Method for Accurate and Efficient Recognition of Gym Exercise Activities Using Electromyography and IMU Sensors

Sakorn Mekruksavanich [1] and Anuchit Jitpattanakul [2,3,*]

[1] Department of Computer Engineering, School of Information and Communication Technology, University of Phayao, Phayao 56000, Thailand; sakorn.me@up.ac.th

[2] Department of Mathematics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand

[3] Intelligent and Nonlinear Dynamic Innovations Research Center, Science and Technology Research Institute, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand

* Correspondence: anuchit.j@sci.kmutnb.ac.th

Abstract: The accurate and efficient recognition of gym workout activities using wearable sensors holds significant implications for assessing fitness levels, tailoring personalized training regimens, and overseeing rehabilitation progress. This study introduces CNN-ResBiGRU, a novel deep learning architecture that amalgamates residual and hybrid methodologies, aiming to precisely categorize gym exercises based on multimodal sensor data. The primary goal of this model is to effectively identify various gym workouts by integrating convolutional neural networks, residual connections, and bidirectional gated recurrent units. Raw electromyography and inertial measurement unit data collected from wearable sensors worn by individuals during strength training and gym sessions serve as inputs for the CNN-ResBiGRU model. Initially, convolutional neural network layers are employed to extract unique features in both temporal and spatial dimensions, capturing localized patterns within the sensor outputs. Subsequently, the extracted features are fed into the ResBiGRU component, leveraging residual connections and bidirectional processing to capture the exercise activities' long-term temporal dependencies and contextual information. The performance of the proposed model is evaluated using the Myogym dataset, comprising data from 10 participants engaged in 30 distinct gym activities. The model achieves a classification accuracy of 97.29% and an F1-score of 92.68%. Ablation studies confirm the effectiveness of the convolutional neural network and ResBiGRU components. The proposed hybrid model uses wearable multimodal sensor data to accurately and efficiently recognize gym exercise activity.

Keywords: deep learning; gym exercise activity; human activity recognition; electromyography; IMU sensor

1. Introduction

Wearable sensors to recognize human activities enable continuous monitoring of movement and physical activity patterns. This technology provides insights into self-tracking and cognitive functions, which can be harnessed in digital health and fitness solutions [1–3]. Miniaturized inertial measurement units (IMUs) and surface electromyography (EMG) sensors are capable of capturing time-series data, offering valuable insights into motion and muscle exertion indicators [4]. Recent progress in machine learning has given rise to multimodal data streams, showing promise in identifying complex activities, assessing biomechanical techniques, and contextualizing exercise behaviors to discern their safety or potential harm [5]. Extensive research in human activity recognition (HAR) concentrates on classifying ambulatory activities like walking, running, and daily chores [6]. However, analyzing gym-based free-weight and resistance exercises presents unique challenges due to the intricate biomechanics involved. These challenges stem

from the various types of fitness equipment, movement speeds, and complex muscle engagement associated with these exercises.

Accurate and efficient recognition of gym workout activities is practical in fitness tracking, rehabilitation, sports coaching, and related domains [7]. The growing adoption of wearable devices like IMUs and EMG sensors has opened up opportunities for the employment of sensor-driven gym exercise identification to offer personalized feedback and real-time self-tracking data [8]. However, categorizing gym exercises poses challenges due to the diverse array of encountered equipment, movement styles, and intensity levels.

Recent research has explored various machine learning techniques to detect exercise activities using sensors. These methods encompass traditional methods, as well as modern deep learning models. Past approaches relied on manually chosen features extracted from sensor data and classifiers like support vector machines and random forests to classify exercises [9,10]. However, the success of these methods depends heavily on domain expertise in engineering meaningful features.

Deep learning has become increasingly common in automating and comprehensively categorizing time series data from sensors. This is accomplished by automatically generating robust hierarchical feature representations [11]. Convolutional neural networks (CNNs) excel at extracting consistent patterns from sensor data streams, both in spatial and temporal aspects [12]. As recurrent models, long short-term memory (LSTM) networks can grasp long-range temporal contexts in sequences of exercises [13]. Hybrid networks that combine CNNs and LSTMs have demonstrated remarkable accuracy in various domains, such as recognizing postures and detecting gestures, mainly when applied to multimodal data collected from wearable devices [14]. This investigation focuses on tailored deep learning approaches specifically crafted for identifying multimodal gym workout activities, leveraging insights from prior accomplishments.

The proposed deep learning architecture for multimodal gym exercise activity identification offers the following significant advancements:

- This research introduces a novel hybrid deep learning design called CNN-ResBiGRU. This architecture effectively combines a CNN, residual connections, and bidirectional gated recurrent units (BiGRUs) to accurately identify gym workout patterns using multimodal sensor data. The proposed model effectively captures essential spatial and temporal features from EMG and IMU data, focusing on speed and computational efficiency.
- An extensive evaluation of the CNN-ResBiGRU model was conducted using the Myogym dataset, which serves as a thorough benchmark comprising data from 10 individuals performing 30 different gym exercises. The model showcased outstanding performance, achieving a remarkable accuracy of 97.29% and an F1-score of 92.68% when combining IMU and EMG data. This outperforms existing deep-learning approaches across various sensor configurations.
- To explore the impacts of different elements within the CNN-ResBiGRU structure, thorough ablation experiments are underway. These studies play a crucial role in unraveling the contributions of convolutional blocks in capturing spatial patterns and the ResBiGRU block in modeling temporal relationships, especially concerning EMG signals. The findings highlight the importance of each component in enhancing the model's effectiveness and robustness.
- The primary objective of this research is to investigate the cooperative interaction between IMU and EMG sensor types when identifying gym activities. Through extensive testing, this study demonstrates the enhanced classification accuracy attained by integrating IMU and EMG data. It emphasizes the importance of utilizing multiple data sources for accurate activity detection.

The subsequent parts of this research are organized as follows: Section 2 outlines HAR; delves into relevant sensor types, particularly IMU and EMG sensors; and assesses prior deep learning methods, along with their limitations. Section 3 details the proposed

approach. It illustrates how the CNN-ResBiGRU architecture is tailored for simulating gym workouts and highlights significant enhancements in the model components tailored for this purpose. Section 4 elaborates on the experimental setup, dataset attributes, implementation specifics, and comprehensive quantitative results, comparing them to state-of-the-art techniques. Thorough ablation studies offer valuable insights into the specific contributions of each model component, complemented by qualitative visualizations showcasing distinctive learned features. Section 5 analyzes the experimental findings, while Section 6 provides an overall summary, identifies potential constraints, and suggests avenues for further research in this domain.

2. Related Works

Recent progress in wearable sensors for identifying human movements has been notable. However, prior studies have focused on simple actions like walking and daily tasks, paying little attention to complex exercise routines. This section reviews pertinent literature on diverse sensor types, both basic and advanced machine learning methods, and techniques for merging data to analyze activities detected by sensors.

2.1. Sensor Modalities

IMUs comprise a gyroscope; an accelerometer; and, in some cases, a magnetometer. The accelerometer measures acceleration along the x, y, and z axes, facilitating the detection of both gravitational forces and linear movement. Conversely, the gyroscope [15] quantifies the rotation rate, while the magnetometer detects and measures the Earth's magnetic fields [16]. IMU sensors are commonly integrated into smart devices, wearables, and smartphones. The widespread use of mobile phones and similar smart gadgets has made IMU data readily accessible and widely available for motion sensing and HAR purposes [17].

EMG sensors have the ability to quantify the electrical waveforms generated by muscular movements [18]. This technique is employed in medical assessments to diagnose and investigate the functionality of nerves and muscles. Consequently, it has considerable importance in the field of rehabilitation and the identification of neuromuscular diseases. Furthermore, these biosignals are used in other domains, such as mechanical actuators employed in human–machine interfaces and for the regulation of robot devices and gaming, among others applications. There are two distinct kinds of EMG sensors, namely intramuscular EMG (iEMG) and surface EMG (sEMG) sensors. The use of sEMG in wearable non-invasive HAR systems has been reported [19]. The authors of [20] devised a method using sEMG to accurately identify ten distinct hand motions.

2.2. Machine Learning Approaches

Various machine learning techniques have demonstrated their effectiveness in analyzing sEMG data for diverse purposes [21], such as recognizing gestures [22], detecting muscle fatigue [23], and investigating human–machine interaction [24]. Different models, including the multi-layer perceptron, have been utilized to classify neuromuscular issues [25]. Support vector machines have been predominantly employed in categorizing sEMG signals [26] and identifying physiological patterns and parameters [10]. Extensive efforts have been dedicated to characterizing walking activities, mainly focusing on classifying gait phases and assessing gait quality [27,28].

2.3. Deep Learning Approaches

The field of HAR has displayed significant interest in EMG information due to its ability to record muscle activation patterns during various movements. Systems for recognizing human movements based on EMG data can offer thorough insights into muscle involvement, fatigue, and coordination. Therefore, they play a crucial role in various applications, such as monitoring fitness progress, rehabilitation progress, and recognizing gestures [4].

Recent advancements in deep learning methodologies have revolutionized the methodology of recognizing human movements using EMG data. Deep learning models like CNNs and recurrent neural networks (RNNs) have demonstrated outstanding proficiency in extracting distinctive features from unprocessed EMG data and precisely classifying human actions [11].

CNNs have shown exceptional efficacy in collecting both spatial and temporal patterns in EMG data. CNNs have the ability to acquire hierarchical features that remain unaffected by translation and scaling by using convolutional filters on the unprocessed EMG signals [12]. The capacity to extract resilient features has resulted in significant improvements in the accuracy of EMG-based HAR compared to conventional machine learning techniques that depend on manually designed features [9].

On the other hand, RNNs have demonstrated notable promise in effectively capturing the temporal connections observed in EMG data. LSTM networks and gated recurrent units (GRUs) are popular designs of RNNs that have proven effective in grasping long-term dependencies and adeptly managing EMG sequences of varying lengths [13]. By leveraging temporal context, RNNs can improve their comprehension of muscle activation patterns and refine the recognition of intricate human movements [10].

Researchers have also investigated hybrid deep learning architectures integrating CNNs and RNNs for EMG-based HAR. This study seeks to use the advantages of both architectures by first extracting spatial characteristics via CNNs and then modeling the temporal relationships using RNNs [14]. Hybrid techniques have shown enhanced efficacy in collecting the spatio-temporal attributes of EMG data and improving the resilience of HAR systems [5].

3. Methodology

In this section, a framework for sensor-based HAR (S-HAR) is presented, with a focus on identifying gym exercises. The S-HAR framework, developed for this study, employs deep learning algorithms to analyze the activities performed by users wearing the Myo armband device, utilizing inputs from IMU and EMG sensors.

3.1. Overview of the S-HAR Framework for Gym Exercise Activity Recognition

This section presents a detailed overview of the operational framework of the proposed S-HAR system. The initial step involves acquiring data from IMU and EMG sensors, then pre-processing procedures like noise elimination, handling of missing data, and standardization of the data. Data segmentation is then performed to convert multi-dimensional sensor data into suitable samples for model training. The concept of temporal windows is outlined, including their definition, the idea of overlap within these windows, and the process of assigning classes and labeling. The sample data are divided into training and testing sets using the 5-fold cross-validation method in the data generation process. Then, various versions of deep learning models are trained in the subsequent stage. Five primary deep learning architectures (CNN, LSTM, bidirectional LSTM (BiLSTM), GRU, and BiGRU) are employed alongside our custom residual deep learning model, CNN-ResBiGRU. The performance of these models is then evaluated using metrics such as accuracy, precision, recall, and F1-score. The operational flow of the proposed S-HAR system is illustrated in Figure 1.
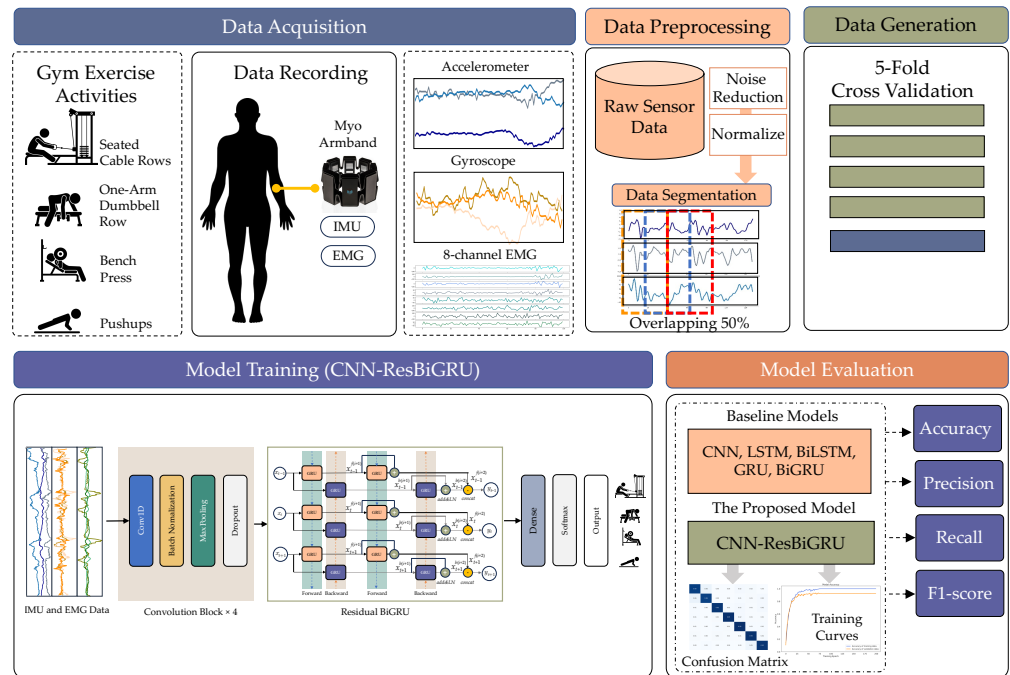
Figure 1. The proposed framework of S-HAR for gym exercise activity recognition.

3.2. Data Acquisition

The dataset utilized in this research, referred to as the Myogym dataset [29], was derived from a sample group of 10 individuals participating in 30 unique gym exercises, each performed ten times, as detailed in Table 1. An additional NULL class denotes periods when no activity was undertaken. The Myogym collection represents an exhaustive assemblage of raw sensor readings captured by the Myo armband during exercise sessions conducted in a gymnasium setting. This dataset encompasses measurements from IMU and EMG. The dimensions of the data are as follows:

- The information obtained from the IMUs encompasses measurements from a three-dimensional accelerometer and a three-dimensional gyroscope. Together, these form a feature domain spanning six dimensions in total. Every data point in this collection represents the instantaneous values for acceleration and angular velocity. These values were recorded at a sampling frequency of 50 measurements per second. Figure 2 illustrates a subset of IMU time-series recordings obtained from the multi-sensor wearable dataset. Panel (a) showcases readings from the triaxial accelerometer, while panel (b) presents corresponding gyroscope rotation signals.

- EMG data: The EMG data were acquired through an array of eight distinct electromyography sensors, generating an eight-dimensional feature space. Each channel recorded the electrical impulses the underlying muscular tissues produced throughout the gymnasium-based exercise routines. The electromyographic signals were acquired at a sampling rate of 50 measurements per second, synchronized with the data collected from the inertial measurement units. As stated by Jung et al. [30], the Myo armband employed a WEMG-8 commercial electromyography sensor manufactured by Laxtha Co., Ltd. (Daejeon, Republic of Korea). This sensor was equipped with a wireless transmitter operating at a frequency of 2.4 GHz and incorporated an analog bandpass filter within the electromyography electrode unit itself. The frequency range of this filter spanned from 13 Hz to 430 Hz. The Myogym dataset's sampling frequency for collecting electromyographic data was set at 50 Hz. Figure 3 depicts a set of EMG time-series streams recorded from arm muscles during various resistance training sessions. The Myogym dataset is a comprehensive compilation of multi-modal time-series data incorporating inertial and EMG sensors to capture information throughout gym sessions. The rationale behind employing a multi-sensor, multi-subject dataset is

to facilitate the advancement and evaluation of algorithms for activity recognition and sensor fusion capable of autonomously identifying and classifying typical strength training and cardiovascular exercises.
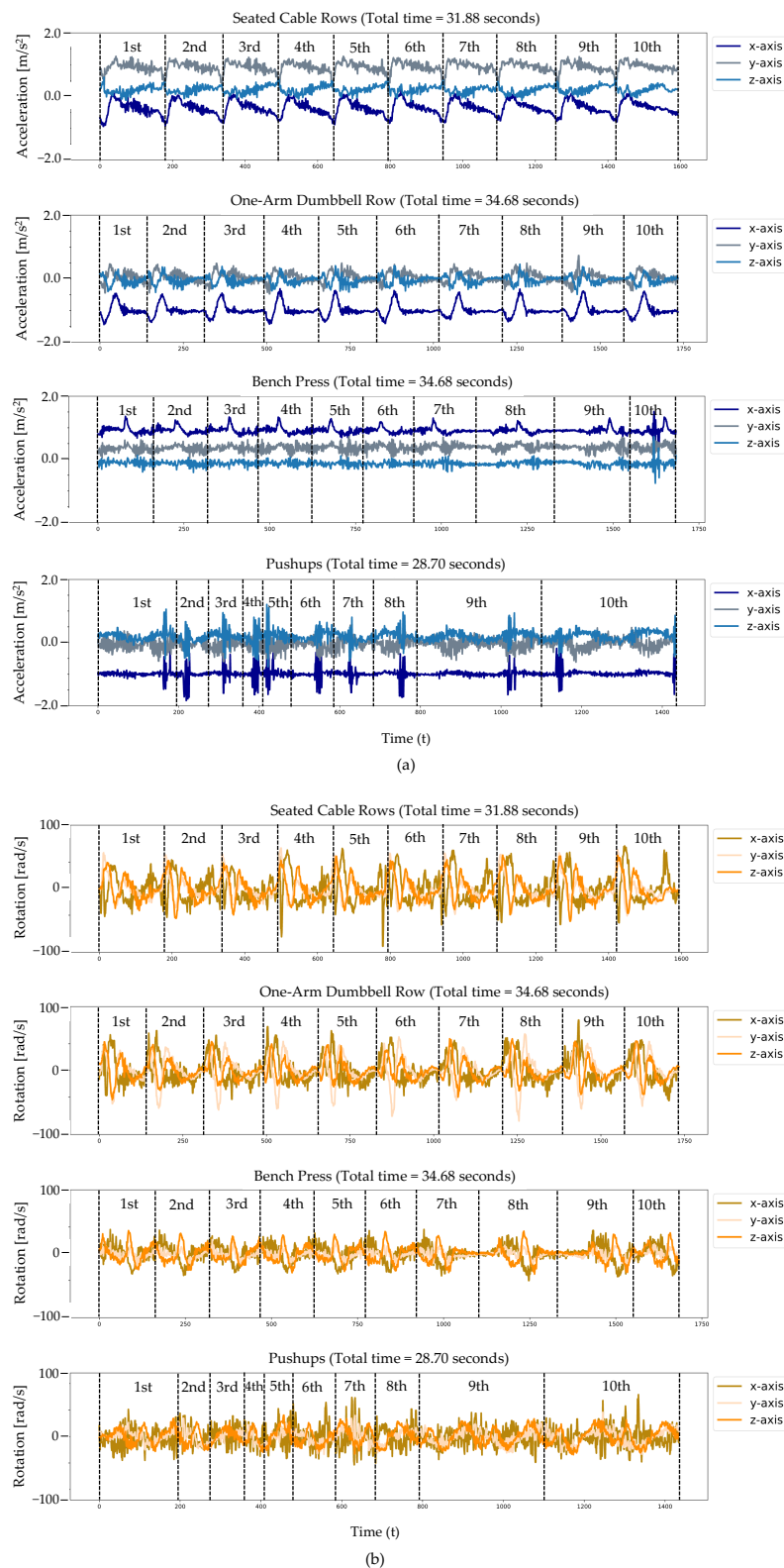
Figure 2. Some samples of IMU data from the Myogym dataset: (**a**) accelerometer data; (**b**) gyroscope data.
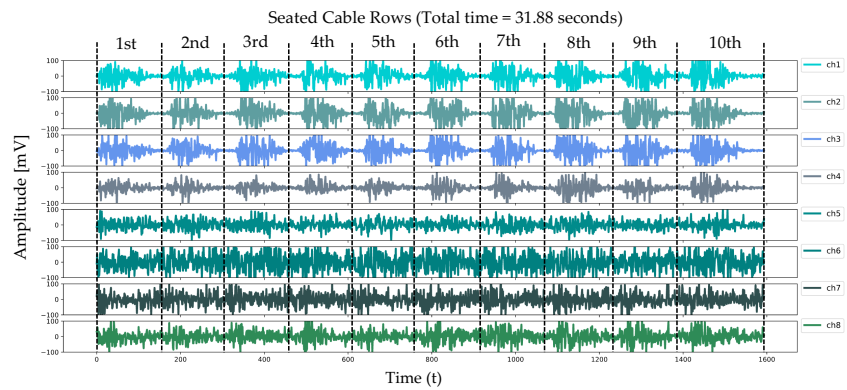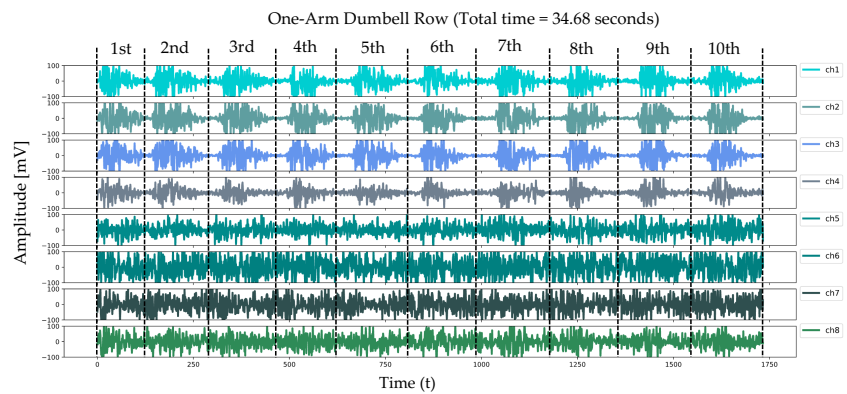
Figure 3. Some samples of EMG data from the Myogym dataset: (**a**) seated cable rows; (**b**) one-arm dumbbell row; (**c**) bench press; (**d**) pushups.

Table 1. Activity list of the Myogym dataset.

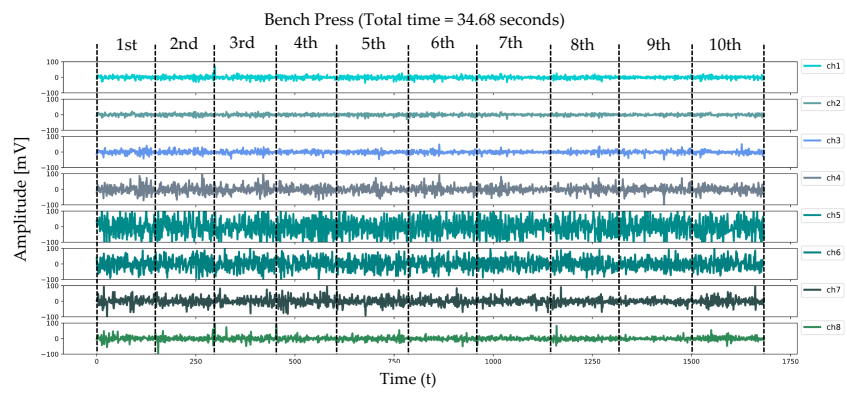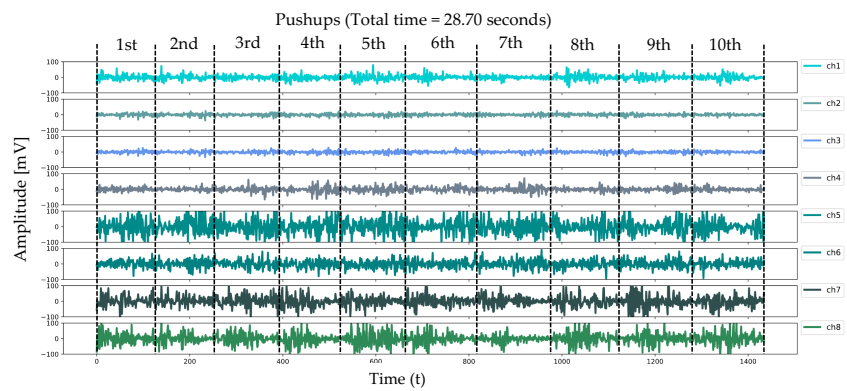| Muscle Group	Gym Exercise Activities	Posture	One-Arm, Both or Alternate	Equipment Used
Middle Back	Seated cable rows	Seated	Both	Cable machine
	One-Arm dumbbell row	Bent over	One-	dumbbell
	Wide-Grip pulldown behind the neck	Seated	Both	Cable machine
	Bent-over barbell row	Bent over	Both	Barbell
	Reverse-grip bent-over row	Bent over	Both	Barbell
	Wide-grip front pulldown	Seated	Both	Cable machine
Chest	Bench press	On back	Both	Barbell
	Inclined dumbbell flyers	Seated, inclined	Both	Dumbbell
	Inclined dumbbell press	Seated, inclined	Both	Dumbbell
	Dumbbell flyers	On back	Both	Dumbbell
	Pushups	Prone with hands and toes grounded	Both	Body weight
	Leveraged chest press	Seated	Both	Barbell
	Closed-grip barbell bench press	On back	Both	Barbell
Triceps	Bar skullcrusher	On back	Both	Barbell
	Tricep pushdown	Standing	Both	Cable machine
	Bench dip/dip	Seated, inclined	Both	Body weight
	Overhead tricep extension	Standing	Both	Barbell
	Tricep dumbbell kickback	Bent over	One arm	Dumbbell
Biceps	Spider curl	On stomach	Both	Barbell
	Dumbbell alternate bicep curl	Standing	Alternate	Dumbbell
	Inclined hammer curl	Seated, inclined	Both	Dumbbell
	Concentration curl	Seated	One arm	Dumbbell
	Cable curl	Standing	Both	Cable machine
	Hammer curl	Standing	Alternate	Dumbbell
Shoulders	Upright barbell row	Standing	Both	Barbell
	Side lateral raise	Standing	Both	Dumbbell
	Front dumbbell raise	Standing	Alternate	Dumbbell
	Seated dumbbell shoulder press	Seated	Both	Dumbbell
	Car drivers	Standing	Both	Barbell plate
	Lying rear-delt raise	On stomach	Both	Dumbbell

The unprocessed sensor data are complemented by appropriate activity labels, which indicate the precise physical exercise done at each time interval in the gym. The Myogym dataset provides a comprehensive and complex description of sensor information, allowing for the creation and assessment of machine learning models for the recognition of gym exercises.

The choice of this dataset was driven by its importance and widespread adoption in the realm of HAR using wearable sensors. It has been utilized in numerous studies as an ideal benchmark for comparing different methodologies and assessing the achieved accuracy levels across various implementations. The designers of the dataset study meticulously supervised the data collection process to ensure data quality and movement consistency. Consequently, this dataset is well-suited for conducting thorough analyses of its various activities. It also proves effective in identifying symmetric movements. As the dataset used in this study is publicly available, the researchers obtained necessary approvals from pertinent authorities, given the involvement of human participants in the study.

Table 1 displays detailed information on 25 variations of resistance training exercises commonly performed during strength training sessions, targeting major muscle groups in the upper and lower body. These exercises encompass a diverse range of equipment, such as free weights like barbells and dumbbells for 13, cable machines for 7, and bodyweight exercises for 5. Among these exercises, 19 involve bilateral arm movements, 3 are unilateral, and 3 involve alternate arm movements. The documentation covers various positions, including sitting, standing, lying down, and prone positions. Additional recorded contextual data include the specific targeted muscle groups, the symmetry of the movement, and details about the utilized resistance methods. This information facilitates the analysis of sensor data collected during the real-world execution of complex multi-joint gym routines

commonly encountered in exercise programs, aiding in the development and assessment of recognition algorithms.

3.3. Data Pre-Processing

The raw Myogym dataset comprises intricate time-series data from various sensors embedded within the Myo armband wearable gadgets the study participants wore. These sensors encompass a 3-axis accelerometer and gyroscope adept at capturing motion dynamics. Moreover, eight channels of sEMG recordings are utilized to gauge muscle activations. However, the collected data streams require appropriate pre-processing and refinement steps to convert them into suitable input formats for developing and applying deep learning models in subsequent analyses.

Initially, denoising algorithms are employed to identify and eliminate irregularities that could result from sensor recording errors or wireless transmission issues. If addressed, these anomalies could positively affect the accuracy of classification. To ensure consistent significance ranking among various sensor inputs, normalization algorithms are utilized to adjust the cleaned signals to a uniform range. Subsequently, the continuous recordings are segmented into overlapping windows of a predetermined size to extract contextual segments suitable for processing by deep neural networks.

The main goal of this systematic data preparation procedure is to enhance the overall quality, compatibility, and utility of the Myogym dataset, enabling efficient training of sensor-based models for human activity recognition. Through a combination of denoising, normalization, and segmentation, dependable time-series tensor representations are generated. These representations are adept at capturing unique features that distinguish different types of gym exercises.

3.3.1. Data Denoising

To enhance the signal quality and minimize the impact of noise sources, a denoising filter is applied to the raw multi-channel time-series data. This study employs a sixth-order zero-phase Butterworth infinite impulse response (IIR) bandpass filter. For the accelerometer and gyroscope measurements, cutoff frequencies of 20 Hz and 1 Hz are set for the low-pass and high-pass filters, respectively. To accommodate the wider frequency range of EMG signals, a band ranging from 10 to 400 Hz is utilized.

3.3.2. Data Normalization

Before inputting the Myogym dataset into the CNN-ResBiGRU model, we used min–max normalization to adjust the sensor input to a consistent range. Min–max normalization is a frequently employed approach that linearly converts the data to a predetermined range, usually from 0 to 1. The normalization step is conducted individually for each sensor channel using the following equation:

$$x_i^{norm} = \frac{x_i - x_i^{min}}{x_i^{max} - x_i^{min}}, i = 1, 2, 3, \dots \tag{1}$$

where x_i^{norm} represents the normalized data, n represents the number of channels, and x_i^{max} and x_i^{min} are the maximum and minimum values of the i-th channel, respectively.

Many reasons determined the selection of min–max normalization. First, it guarantees uniformity in scaling throughout several sensor modalities (IMU and EMG) by adjusting all channels to an identical range. It is crucial to include this factor when combining data from several sensors, as it avoids any one kind of sensor overpowering the others, owing to variations in magnitude. Furthermore, min–max normalization ensures the initial data distribution is maintained throughout the scaling process, preserving the corresponding connections across data points. This quality is essential for preserving the intrinsic structures and properties of the sensor signals.

Furthermore, applying data normalization to a predetermined range is suitable for deep learning models and stabilizes the learning procedure. By normalizing the data to

a range of 0 to 1, we ensure that the input to the CNN-ResBiGRU model is suitable for practical training and optimization. Min–max normalization is more resistant to outliers than other normalizing procedures, such as z-score normalization. Since the Myogym dataset consists of practical sensor data that can sometimes contain outliers, the use of min–max normalization helps to reduce the influence of these outliers on the general distribution of the data.

Min–max normalization was applied to each sensor channel separately, such as the x-axis of the accelerometer, the y-axis of the gyroscope, and the first EMG channel. The normalizing approach was only applied to some channels collectively. This method guarantees that the scaling procedure is tailored to the distinct attributes of each specific channel, thus maintaining the proportional differences between various sensor modalities. By independently normalizing each channel, we preserve the unique characteristics and variations specific to each sensor signal while adjusting them to a standardized scale.

3.3.3. Data Segmentation

The detailed documentation of an exercise needs to be more extensive and directly presented as input to the classifier. Instead, the data are divided into smaller segments known as windows. These windows are created by segmenting the body movement data from the sensor device, while the exercise-related information obtained from the mobile application serves as class labels. Each section is labeled based on the predominant class mode. Previous studies [31,32] explored various window lengths with overlapping. In our study, we used a window size of 2 s for activity recognition (as shown in Figure 4), resulting in 100 samples. The results are presented both with and without overlapping windows. Our investigation considered the window dimensions and the extent of overlap. The continuous nature of our exercise activity also contributes to the need for overlapping. This overlap ensures that each successive window retains some information from the preceding one.
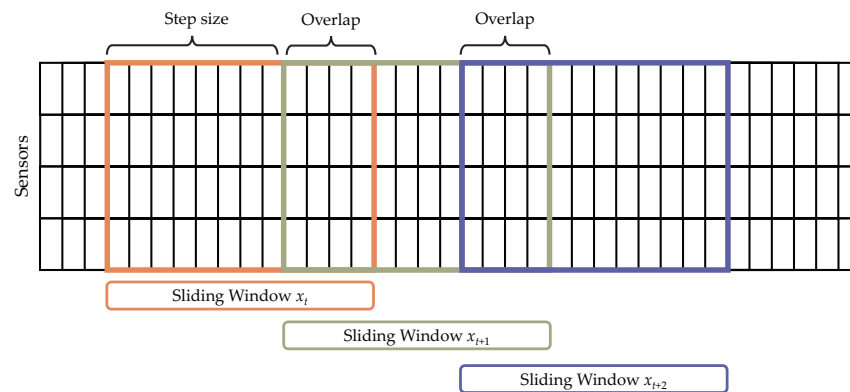
Figure 4. A visualization of the sliding window segmentation with a fixed-width window size of 13 and an overlapping proportion.

3.3.4. Data Generation

To ensure the robustness and versatility of our CNN-ResBiGRU model, we employed a rigorous 5-fold cross-validation approach for training and assessment. This widely accepted method for evaluating a model's performance involves dividing the dataset into five equal-sized subsets, or folds, and iteratively using each fold as a validation set while training the model on the other four-folds.

The Myogym dataset, which consists of sensor data from 10 subjects doing 30 different gymnasium activities, was divided into five folds using random partitioning in our analysis. Each fold included around 20% of the sample set, guaranteeing an equitable data distribution across all folds. The CNN-ResBiGRU model was then trained and evaluated five times, with a different fold being used as the validation set in each iteration.

During each iteration of the 5-fold cross-validation procedure, the model was trained using 80% of the data (four folds) and validated with the remaining 20% (one fold). This pro-

cess was repeated five times, with each fold serving as the validation set exactly once. The final performance measures, such as accuracy and F1-score, were computed by averaging the results from the five iterations, providing a reliable estimate of the model's performance.

The 5-fold cross-validation approach is essential because it guarantees that all Myogym dataset samples are used for training and validation. This helps to reduce the danger of overfitting and provides a more accurate evaluation of the model's capacity to generalize.

3.4. The Proposed CNN-ResBiGRU Model

The proposed approach entails creating a hybrid deep learning model comprising convolution, residual BiGRU blocks, and a hybrid architecture. This end-to-end model's overall structure is depicted in Figure 5.
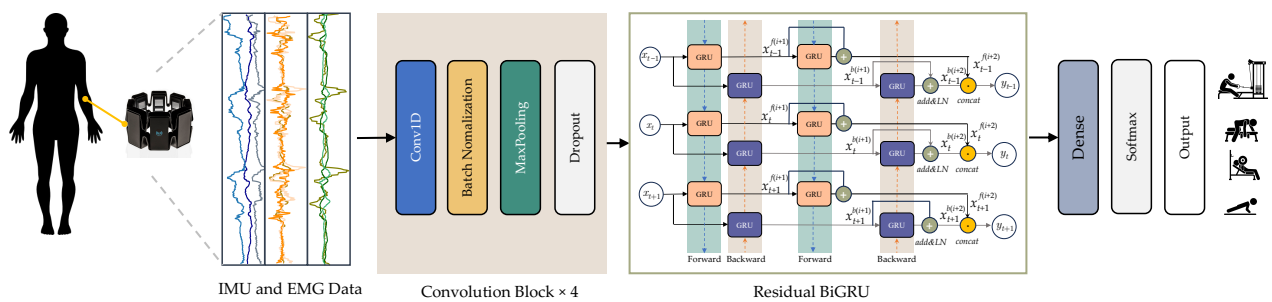
Figure 5. Detailed architecture of the proposed CNN-ResBiGRU model.

In our proposed framework, the first component, the convolution block, is responsible for extracting spatial features from the pre-processed input data. To reduce the time-series length, we adjust the step size of the convolution kernel, thus speeding up the recognition process. Following this, we employ the BiGRU network to extract temporal characteristics from the data processed by the convolution block. This enhances the model's ability to capture long-term dependencies in the time-series data by leveraging the benefits of a BiGRU. Integrating these components makes the model better equipped to understand complex temporal patterns, thereby improving recognition accuracy. The behavior data are then classified using a fully connected layer and the SoftMax algorithm. The output of this classification process serves as the recognition result, predicting the specific activity being performed. We delve deeper into each component in subsequent sections, outlining their roles and contributions within our proposed framework.

3.4.1. Convolution Block

In the realm of CNNs, employing a predefined set of elements is typical. These neural networks are frequently utilized in supervised learning scenarios. Generally, they establish connections between each neuron and all others in subsequent layers. The activation function within these networks converts input neuron values into output values. The effectiveness of this function is influenced by two key factors, namely sparsity and the capacity of lower layers to handle diminished gradient flow. In CNNs, dimensionality reduction is often achieved through pooling techniques. The widely adopted method is maximum pooling, while average pooling is also commonly utilized.

Our study utilizes convolutional blocks (ConvBs) to capture basic features from raw sensor data. A ConvB comprises the following four layers: a 1D convolutional layer (Conv1D), batch normalization layer (BN), max-pooling layer (MP), and dropout layer. The Conv1D model contains multiple trainable convolutional kernels, each producing a distinct feature map. We incorporated the BN layer into the architecture to expedite and improve training.

3.4.2. Residual BiGRU Block

Rather than relying solely on the convolution block for spatial feature extraction, activity identification is required due to the temporal aspect of human activities. It is crucial to take into account the entire activity sequence over time. RNNs offer beneficial traits in processing time-series data. However, as the time series lengthens, RNN models may encounter problems like gradient vanishing and loss of information.

Hochreiter et al. [33] suggested LSTM, which, in contrast to basic RNNs, is a recurrent neural network that uses gates to efficiently preserve temporal information over extended periods of time. Furthermore, it surpasses basic RNNs in managing extended time series. However, it should be noted that behavioral data are subject to the effects of both previous and succeeding events.

LSTM has demonstrated its ability to address the gradient vanishing challenge observed in RNNs, yet its utilization of memory cells increases memory usage. The GRU network introduced by Cho et al. in 2014 [34] offers a novel approach within the realm of RNNs. Unlike LSTM, GRU does not include dedicated memory cells [35]. Instead, it features update and reset gates, which regulate the extent of modification for each hidden state, determining which information is retained and discarded. BiGRU, a variant of GRU, integrates both forward and backward information in its computation, enhancing time-series feature extraction by capturing bidirectional dependencies. Consequently, employing BiGRU networks to extract time-series features from behavioral data is a suitable strategy.

The BiLSTM network excels in extracting time-series features but has constraints in capturing spatial details. Additionally, escalating the number of stacked layers exacerbates the challenge of vanishing gradients during training. To tackle this issue, He et al. [36] introduced the ResNet residual network in 2016, boasting 152 layers and securing victory in the 2015 ILSVRC tournament. The formulation for each residual block is as follows:

$$x^{i+1} = x^i + f(x^i, W_i) \tag{2}$$

The residual blocks are partitioned into two components, namely x^i, which represents a direct mapping, and $F(x^i, W_i)$, which represents the residual portion.

In our investigation, we implemented a residual architecture similar to the one found in the coding segment of the transformer model. This residual structure, inspired by the BiGRU network, harnesses the benefits associated with this specific configuration. Moreover, the BiGRU network can integrate normalization techniques. Layer normalization (LN) [37], a normalization method applied in recurrent neural networks, offers distinct advantages over BN. LN operates similarly to BN and can be expressed mathematically as follows:

$$\widehat{x}^i = \frac{x^i - E(x^i)}{\sqrt{var(x^i)}} \tag{3}$$

where x^i denotes the input vector in the i-th dimension and $\widehat{x}^i$ represents the output subsequent to LN.

This research introduces a novel approach termed ResBiGRU, which integrates residual structure with layer normalization in a BiGRU network. This combination is visually shown in Figure 6. The feature information (y) could be defined as recursive:

$$x_t^{f(i+1)} = LN(x_t^{f(i)} + GRU(x_t^{f(i)}, W_i)) \tag{4}$$

$$x_t^{b(i+1)} = LN(x_t^{b(i)} + GRU(x_t^{b(i)}, W_i)) \tag{5}$$

$$y_t = concat(x_t^f, x_t^b) \tag{6}$$

where *LN* denotes the normalization of layers. At the same time, the processing of input states in the GRU network is represented by *GRU*. The t-th moment in the time series is

denoted by the subscript t. The forward state is indicated by the superscript f, the reverse state by b, and the number of stacked layers by $(i + 1)$. The encoded information at time t (y_t) is generated by amalgamating the forward and backward states.
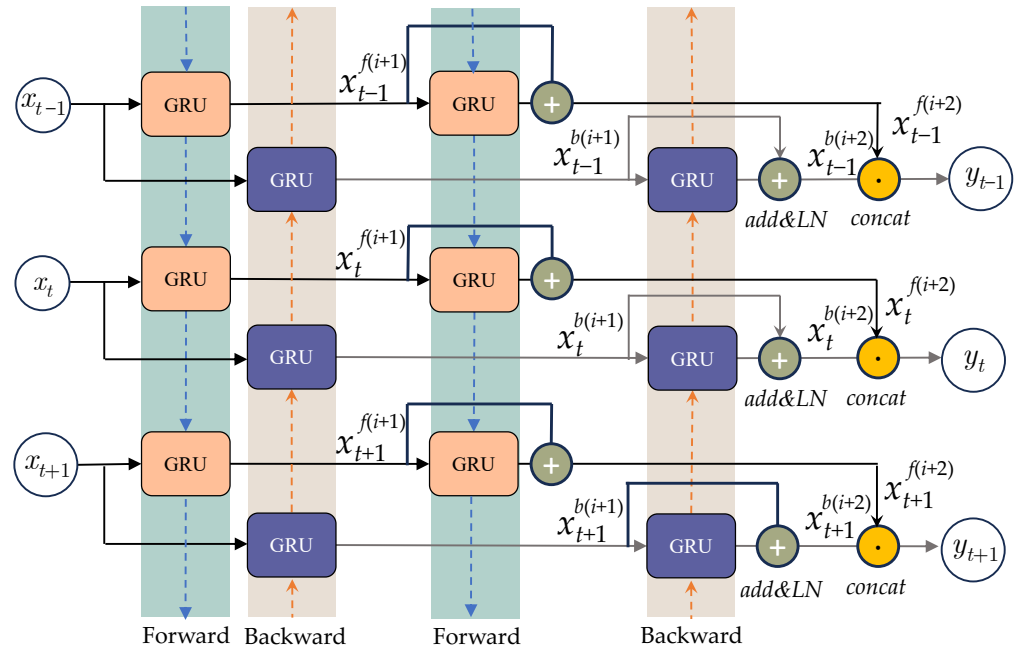
Figure 6. Structure of ResBiGRU.

3.5. Evaluation Metrics

In this research, we evaluated the CNN-ResBiGRU model's effectiveness by comparing it to baseline methods using key multi-class classification measures. These measures include overall accuracy, precision, recall, and F1-score, which offer crucial insights into the reliability of activity classification both within specific categories and across various datasets.

- In the context of the model, accuracy refers to the proportion of action occurrences that were accurately identified.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

The variables TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

- *Precision* is defined as the positive value of predictions, which is the ratio of correctly identified positive identifications (class × predicted) to the total number of positive identifications.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

- *Recall*, also known as sensitivity, quantifies the true-positive rate, the ratio of correctly predicted real positive cases (actual class) to the total number of occurrences.

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

- The *F1-score* is a statistic that combines accuracy and recall, providing a balanced measure.

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

Model performance evaluation encompasses precision, memory, and accuracy, allowing for a comprehensive assessment of overall corrections and the efficacy of identifying various types of activities while ensuring that classes with fewer instances are not overlooked. The *F1-score* consolidates these factors into a single, comprehensible number. Examining different models using these measures offers a more valuable understanding than only focusing on accuracy.

4. Experiments and Results

This section investigates our research efforts to identify the most suitable deep learning models for S-HAR in categorizing gym exercises. We aim to create models capable of recognizing different gym activities accurately and efficiently. We center our study on the Myogym dataset, designed explicitly for tasks related to gym exercise recognition, containing diverse sensor data collected during various gym sessions. To gauge the effectiveness of the deep learning models, we rely on two widely used metrics in S-HAR for gym applications, namely accuracy and F1-score. These metrics comprehensively assess how well the models perform in accurately classifying gym exercise activities.

4.1. Experimental Setting

The training of deep learning models was accelerated in this study by using Google Colab Pro+ in conjunction with a Tesla V100-SXM2-16GB (Los Angeles, CA, USA) graphics processor module. The fundamental deep learning models, including CNN-ResBiGRU, were implemented using a Python 3.10.12 framework with TensorFlow and CUDA backends. The investigations specifically examined the Python packages listed below:

- The Numpy and Pandas libraries were used for the purpose of data management throughout the retrieval, processing, and analysis of sensor data.
- The outcomes of data exploration and model assessment were charted and presented using Matplotlib and Seaborn.
- In the experiments, the Scikit-learn model was used for the purposes of sampling and data creation.
- TensorFlow was used to generate and train deep learning models.

This study involved multiple experiments using the Myogym dataset to determine the best approach for recognizing gym exercises. To ensure the reliability and applicability of our findings, we employed a five-fold cross-validation technique. The experimental process was divided into five distinct scenarios, each focusing on different combinations of sensor data available in the Myogym dataset. In Scenario I, models were trained and tested using only accelerometer data, while Scenario II utilized only gyroscope data. Scenario III involved accelerometer and gyroscope data, referred to as IMU data, to assess the models' performance with a broader range of motion-related features. Scenario IV concentrated solely on EMG data, which captures muscle activation patterns during physical exercises. In the fifth scenario, IMU and EMG data were combined (IMU + EMG) to explore the potential benefits of integrating motion and muscle activity information for gym exercise recognition.

4.2. Experimental Results

Experiments were conducted to evaluate how well the proposed CNN-ResBiGRU model recognized activities and compare its interpretation with other deep learning methods.

The results from various deep learning models in recognizing gym exercise activities using only accelerometer data (Scenario I) are detailed in Table 2. The CNN-ResBiGRU model surpasses all others, boasting a peak accuracy of 95.67% and an F1-score of 88.08%, with a loss of 0.17. Coming in second place, the BiGRU model achieves an accuracy of 95.33%, an F1-score of 86.84%, and a loss of 0.29. BiLSTM and GRU perform similarly, with accuracies of 94.22% and 92.94%, F1-score of 84.08% and 80.09%, and losses of 0.32 and 0.35, respectively. The LSTM model attains an accuracy of 92.54%, an F1-score of 78.58%,

and a loss of 0.29. Among the deep learning models, CNN demonstrates the lowest performance, achieving an accuracy of 90.08%, an F1-score of 69.60%, and a loss of 0.48.

Table 2. Comparative evaluation of deep learning models for gym exercise recognition using accelerometer data (Scenario I).

Model	Recognition Performance		
	Accuracy	Loss	F1-score
CNN	90.08% (±0.44%)	0.48 (±0.03)	69.60% (±1.30%)
LSTM	92.54% (±0.15%)	0.29 (±0.01)	78.58% (±0.69%)
BiLSTM	94.22% (±0.27%)	0.32 (±0.01)	84.08% (±0.38%)
GRU	92.94% (±0.19%)	0.35 (±0.01)	80.09% (±1.02%)
BiGRU	95.33% (±0.27%)	0.29 (±0.01)	86.84% (±0.68%)
CNN-ResBiGRU	95.67% (±0.21%)	0.17 (±0.01)	88.08% (±0.56%)

Table 3 showcases the performance of various deep learning models in identifying gym exercise routines using only gyroscope data (Scenario II). The results are summarized below. The CNN-ResBiGRU model achieves an impressive accuracy of 95.58% and an F1-score of 87.02%, accompanied by a loss of 0.20. Notably, this performance mirrors the outcomes observed in the analysis of accelerometer data in Scenario I. The BiGRU model attains a precision of 93.00%, an F1-score of 78.02%, and a loss of 0.54. Similarly, BiLSTM and LSTM demonstrate comparable performance, with 92.15% and 90.96% accuracy, F1-score of 74.88% and 70.82%, and losses of 0.53 and 0.51, respectively. The GRU model achieves a precision rate of 90.89%, an F1-score of 70.84%, and a loss of 0.58. However, CNN exhibits the lowest performance among the deep learning models, with an accuracy of 82.94%, an F1-score of 42.16%, and a loss of 0.95.

Table 3. Comparative evaluation of deep learning models for gym exercise recognition using gyroscope data (Scenario II).

Model	Recognition Performance		
	Accuracy	Loss	F1-score
CNN	82.94% (±0.55%)	0.95 (±0.09)	42.16% (±1.27%)
LSTM	90.96% (±0.16%)	0.51 (±0.02)	70.82% (±0.52%)
BiLSTM	92.15% (±0.27%)	0.53 (±0.05)	74.88% (±1.33%)
GRU	90.89% (±0.39%)	0.58 (±0.01)	70.84% (±1.36%)
BiGRU	93.00% (±0.21%)	0.54 (±0.02)	78.02% (±0.55%)
CNN-ResBiGRU	95.58% (±0.18%)	0.20 (±0.01)	87.02% (±0.64%)

Table 4 presents the performance of various deep learning models in identifying gym exercise routines using both accelerometer and gyroscope data (Scenario III). The CNN-ResBiGRU model stands out, with an impressive accuracy of 96.96%, an F1-score of 91.78%, and a loss of 0.14. Notably, this performance surpasses Scenarios I and II, where only accelerometer or gyroscope data were utilized. The BiGRU model achieves an accuracy of 95.03%, an F1-score of 85.22%, and a loss of 0.35%. Similarly, BiLSTM and LSTM demonstrate comparable performance, with accuracies of 94.32% and 94.29%, F1-score of 82.82% and 82.92%, and losses of 0.35 and 0.30, respectively. The GRU model exhibits a mean accuracy of 94.15%, an F1-score of 82.60%, and a loss of 0.34. However, CNN displays the lowest performance among the deep learning models, with an accuracy of 89.13%, an F1-score of 64.72%, and a loss of 0.96.

The results reveal that combining accelerometer and gyroscope data, commonly known as IMU data, improved the recognition performance of all deep learning models compared to using either type of sensor data alone. The CNN-ResBiGRU model incorporates spatial and temporal features from accelerometer and gyroscope data, leading to the highest accuracy and F1-score.

Table 4. Comparative evaluation of deep learning models for gym exercise recognition using accelerometer and gyroscope data (Scenario III).

Model	Recognition Performance		
	Accuracy	Loss	F1-score
CNN	89.13% (±0.42%)	0.96 (±0.10)	64.72% (±1.19%)
LSTM	94.29% (±0.15%)	0.30 (±0.02)	82.92% (±0.73%)
BiLSTM	94.32% (±0.07%)	0.35 (±0.01)	82.82% (±0.35%)
GRU	94.15% (±0.15%)	0.34 (±0.02)	82.60% (±0.29%)
BiGRU	95.03% (±0.27%)	0.35 (±0.02)	85.22% (±0.80%)
CNN-ResBiGRU	96.96% (±0.10%)	0.14 (±0.01)	91.78% (±0.42%)

Table 5 displays the effectiveness of various deep learning architectures in identifying gym workout activities solely through EMG data, as depicted in Scenario IV. The outcomes are outlined as follows: The CNN-ResBiGRU model illustrates a remarkable precision of 91.53%, an F1-score of 74.49%, and a calculated loss of 0.34. However, this model's efficacy diminishes notably when contrasted with its performance utilizing an accelerometer, gyroscope, or a fusion of both (IMU data) in Scenarios I, II, and III. In contrast, the LSTM model records an accuracy of 78.05%, an F1-score of 24.06%, and a loss of 1.25. Similarly, the BiLSTM model achieves an accuracy of 77.58%, an F1-score of 22.04%, and a loss of 1.83. The performance of GRU and BiGRU appears to be comparable, as indicated by their respective accuracies of 76.57% and 76.18%, accompanied by F1-score of 18.82% and 16.55% and losses of 1.37 and 2.14, respectively. Among the various examined deep learning architectures, the CNN model is the least effective, with an accuracy of 75.74%, an F1-score of 8.00%, and a recorded loss of 0.93.

The results suggest a significant decrease in performance across all deep learning models when EMG data alone are utilized for gym activity detection, compared to using accelerometer, gyroscope, or IMU data. This suggests that relying solely on EMG data might not provide adequate information for accurately identifying gym workouts.

Table 5. Comparative evaluation of deep learning models for gym exercise recognition using EMG data (Scenario IV).

Model	Recognition Performance		
	Accuracy	Loss	F1-score
CNN	75.74% (±0.51%)	0.93 (±0.02)	8.00% (±1.38%)
LSTM	78.05% (±0.40%)	1.25 (±0.04)	24.06% (±1.45%)
BiLSTM	77.58% (±0.34%)	1.83 (±0.07%)	22.04% (±0.98%)
GRU	76.57% (±0.85%)	1.37 (±0.03)	18.82% (±0.69%)
BiGRU	76.18% (±0.41%)	2.14 (±0.06)	16.55% (±0.45%)
CNN-ResBiGRU	91.53% (±0.60%)	0.34 (±0.02)	74.49% (±1.74%)

Table 6 displays the performance of several deep learning models in identifying gym workout activities using a fusion of IMU (accelerometer and gyroscope) and EMG data, as outlined in Scenario V. The results are summarized as follows: The CNN-ResBiGRU model achieves the highest accuracy of 97.29% and an F1-score of 92.68%, accompanied by a loss of 0.12. This performance surpasses its previous outcomes across all scenarios, highlighting the benefits of combining IMU and EMG data.

Meanwhile, the LSTM model records an accuracy of 91.79%, an F1-score of 73.00%, and a loss value of 0.33. Similarly, the BiLSTM model demonstrates a precision rate of 91.20%, an F1-score of 70.62%, and a loss of 0.50. Comparable performance is observed between GRU and BiGRU, with respective accuracies of 89.71% and 89.93%, F1-score of 64.66% and 65.91%, and losses of 0.46 and 0.65. The CNN model achieves an accuracy of 86.43%, an F1-score of 54.22%, and a recorded loss of 0.61, surpassing its performance in Scenario IV, which relied solely on EMG data.

The results indicate that combining IMU and EMG data leads to improved recognition performance across all deep learning models compared to using either type of data alone. The CNN-ResBiGRU model adeptly integrates spatial and temporal features from both IMU and EMG data, resulting in the highest accuracy and F1-score.

Table 6. Comparative evaluation of deep learning models for gym exercise recognition using IMU and EMG data (Scenario V).

Model	Recognition Performance		
	Accuracy	Loss	F1-score
CNN	86.43% (±0.54%)	0.61 (±0.07)	54.22% (±2.97%)
LSTM	91.79% (±0.42%)	0.33 (±0.01)	73.00% (±1.55%)
BiLSTM	91.20% (±0.31%)	0.50 (±0.04)	70.62% (±0.92%)
GRU	89.71% (±0.28%)	0.46 (±0.02)	64.66% (±0.73%)
BiGRU	89.93% (±0.41%)	0.6 5(±0.03)	65.91% (±1.88%)
CNN-ResBiGRU	97.29% (±0.20%)	0.12 (±0.01)	92.68% (±0.59%)

5. Discussion

This part presents a detailed examination of the experimental results obtained from our research on deep learning models for the identification of gym exercises using various sensor modalities. We initiated the analysis by comparing the effectiveness of our proposed CNN-ResBiGRU model with state-of-the-art techniques, highlighting its exceptional accuracy and F1-score under different conditions. First, we conducted comprehensive ablation tests to assess the contributions of individual components within the CNN-ResBiGRU architecture, thereby gaining valuable insights into each module's efficacy. Additionally, we explored how different sensor modalities impact recognition performance, investigating the synergistic interplay between IMU and EMG data and their integration to enhance accuracy. Lastly, we scrutinized the practical implications of our research, showcasing potential applications of the proposed gym activity detection system in real-world scenarios such as fitness monitoring, personalized training, and injury prevention.

5.1. Ablation Studies

In the domain of neural networks, researchers often use ablation studies to comprehend the individual contributions of various components inside a model to its overall performance [38]. This method eliminates or alters elements and observes the subsequent effect. The suggested CNN-ResBiGRU model was subjected to three ablation trials using this methodology. For each instance, we modified certain blocks or layers to evaluate their impact on the correctness of the model [39,40]. Through a comprehensive analysis of the findings from these trials, we successfully discovered the configuration that exhibited the highest recognition performance level, enhancing our model's overall efficacy.

The F1-score is a vital indicator for examining model variants. This measure safeguards against the potential for inflated assessment scores resulting from selecting the dominant class in unbalanced datasets. On the contrary, it provides accurate and detailed information on the model's effectiveness in managing various classroom tasks [41].

5.1.1. Impact of Convolution Blocks

A series of ablation investigations were performed on the Myogym dataset utilizing IMU and EMG sensor data to investigate the impact of using convolution blocks on the effectiveness of our model.

The findings of ablation research examining the influence of convolutional blocks on the identification performance of the CNN-ResBiGRU model are shown in Table 7. This investigation used several sensor modalities from the Myogym dataset. The primary outcomes are outlined as follows: The CNN-ResBiGRU model, when using convolutional blocks, demonstrates superior performance compared to the baseline model that does not use convolutional blocks across all sensor modalities. The CNN-ResBiGRU model has an

F1-score of 88.08% when only using accelerometer data, in contrast to the model without convolutional blocks, which produces an F1-score of 68.90%. The CNN-ResBiGRU model produces an F1-score of 87.02% for gyroscope data, while the baseline model without convolutional blocks derives an F1-score of 69.29%. The CNN-ResBiGRU model achieves an F1-score of 91.78% when including accelerometer and gyroscope data (IMU), in contrast to the model without convolutional blocks, which achieves an F1-score of 80.93%. By only using EMG data, the CNN-ResBiGRU model demonstrates a substantial improvement in the F1-score, reaching 74.49%. In contrast, the baseline model, which lacks convolutional blocks, only achieves a score of 29.92%. The CNN-ResBiGRU model achieves an F1-score of 92.68% when combining IMU and EMG data (IMU+EMG), whereas the model without convolutional blocks has a lower F1-score of 75.99%.

Table 7. Impact of convolutional blocks.

Model	F1-score (%)				
	Accelerometer	Gyroscope	IMU	EMG	IMU + EMG
The proposed model without convolution blocks	68.90%	69.29%	80.93%	29.92%	75.99%
CNN-ResBiGRU	88.08%	87.02%	91.78%	74.49%	92.68%

This study's discoveries provide evidence for the efficacy of integrating convolutional blocks into the CNN-ResBiGRU model in collecting spatial data from diverse sensor modalities. Convolutional blocks facilitate the acquisition of distinctive patterns within the sensor data, enhancing the model's identification capabilities in various settings. The influence of convolutional blocks is notably significant in the context of EMG data analysis, as shown by the CNN-ResBiGRU model's notable enhancement in F1-score compared to the benchmark model with no convolutional blocks.

5.1.2. Impact of the ResBiGRU Blocks

To further evaluate the convolutional block's capacity to collect spatial characteristics, we ran an additional ablation study. We compared our whole model with a modified version that did not include the residual BiGRU block. The revised iteration functioned as the reference point.

The ablation study examining the impact of the ResBiGRU block on the recognition performance of the CNN-ResBiGRU model is detailed in Table 8, utilizing various sensor modalities from the Myogym dataset. Overall, including the ResBiGRU block consistently enhances the CNN-ResBiGRU model's performance across all sensor modalities compared to the baseline model without it. Specifically, when using accelerometer data, the CNN-ResBiGRU model achieves an F1-score of 88.08% versus 82.10% for the baseline model without the ResBiGRU block. Similarly, for gyroscope data, the CNN-ResBiGRU model achieves an F1-score of 87.02% compared to 83.78% for the baseline model. When incorporating accelerometer and gyroscope data (IMU), the CNN-ResBiGRU model achieves an F1-score of 91.78% compared to 91.49% for the baseline. Using only EMG data significantly improves the F1-score to 74.49% for the CNN-ResBiGRU model, while the baseline model achieves only 54.21%. Combining IMU and EMG data (IMU+EMG) yields an F1-score of 92.68% for the CNN-ResBiGRU model, whereas the model without the ResBiGRU block achieves a slightly lower F1-score of 92.40%.

**Table 8.** Impact of ResBiGRU blocks.

Model	F1-score (%)				
	Accelerometer	Gyroscope	IMU	EMG	IMU + EMG
The proposed model without convolution blocks	82.10%	83.78%	91.49%	54.21%	92.40%
CNN-ResBiGRU	88.08%	87.02%	91.78%	74.49%	92.68%

These findings underscore the importance of the ResBiGRU block within the CNN-ResBiGRU model, as it efficiently captures temporal dependencies and long-term contextual information directly from the sensor data. Integrating the ResBiGRU block enhances the model's ability to accurately represent the sequential characteristics of sensor signals, leading to improved identification performance across different contexts. Particularly in EMG data analysis, the CNN-ResBiGRU model demonstrates a significant increase in F1-score compared to the baseline model without the ResBiGRU block, indicating its effectiveness in capturing temporal patterns inherent in EMG data.

The ablation analysis provides compelling evidence that the ResBiGRU block plays a crucial role in the CNN-ResBiGRU model, contributing significantly to its outstanding implementation in tasks related to recognizing gym exercises across diverse sensor modalities.

5.2. Impact of Different Types of Sensors

The effectiveness of multiple deep learning models for gym activity identification utilizing different sensor modalities is shown in Tables 2–6. The CNN-ResBiGRU model performs better than other baseline models in all circumstances. This is attributed to its hybrid design, which efficiently captures both spatial and temporal characteristics from sensor input. In Scenario V, integrating IMU (accelerometer and gyroscope) and EMG data yields optimal results. CNN-ResBiGRU achieves a remarkable accuracy of 97.29% and an F1-score of 92.68%. These results underscore the valuable complementary information these sensor modalities offer in accurately recognizing gym exercises, as shown in Figure 7.
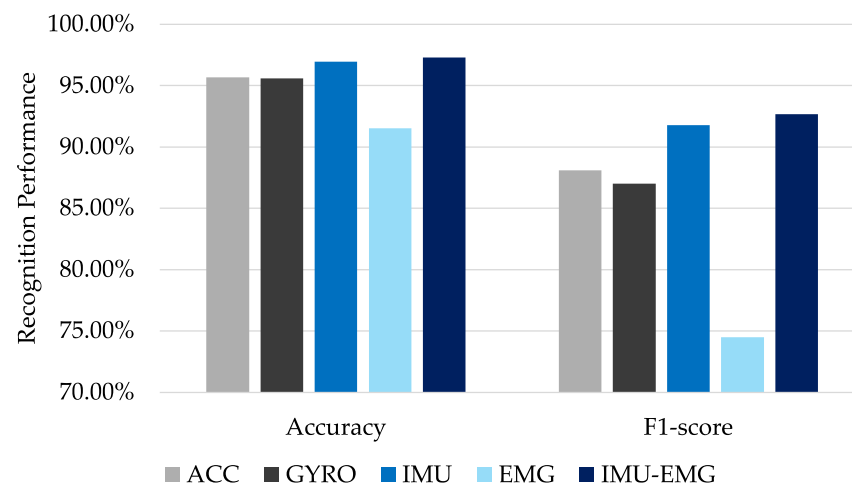
Figure 7. Comparative results of the proposed CNN-ResBiGRU using different types of sensor data.

5.3. Comparison with State-of-the-Art Models

In order to emphasize the originality and effectiveness of our CNN-ResBiGRU model, we performed a comparative assessment of its performance against other cutting-edge models often used for time-series classification tasks, including InceptionTime [42], Deep-ConvLSTM [43], and ResNet [44]. InceptionTime is a deep learning framework that draws inspiration from the Inception-v4 model, which is well known for its outstanding performance in various time-series categorization tasks. This model has several convolutional layers with varied kernel sizes and mixes their outputs to collect characteristics at different time scales. The DeepConvLSTM model is a well-recognized architecture that integrates convolutional layers with LSTM cells. This allows it to effectively collect spatial and temporal features from time-series data. ResNet, first developed for image classification, has been modified for time-series classification by substituting 2D convolutions with 1D convolutions. The incorporation of residual connections effectively tackles the issue of the vanishing gradient problem, enabling the successful training of networks with increased depth.

In order to guarantee an equitable evaluation, we included these cutting-edge models and trained them on the same Myogym dataset, employing the five-fold cross-validation

procedure outlined in Section 3.3.4. The effectiveness assessment of our CNN-ResBiGRU model in comparison with InceptionTime, DeepConvLSTM, and ResNet on the Myogym dataset is shown in Table 9.

Table 9. Results of comparison with state-of-the-art models.

Model	Accuracy (%)				
	Accelerometer	Gyroscope	IMU	EMG	IMU + EMG
InceptionTime [42]	91.56%	90.42%	77.50%	77.50%	86.84%
DeepConvLSTM [43]	89.29%	84.31%	88.62%	78.30%	90.46%
ResNet [44]	79.24%	84.04%	95.95%	77.45%	92.85%
CNN-ResBiGRU	95.69%	95.58%	96.96%	91.53%	97.29%

The data shown in Table 9 illustrate the exceptional efficacy of our CNN-ResBiGRU model in comparison to cutting-edge models such as InceptionTime, DeepConvLSTM, and ResNet when considering different combinations of sensor modalities. CNN-ResBiGRU routinely achieves superior performance compared to these models in situations that include accelerometer, gyroscope, IMU, EMG, and the combination of IMU and EMG data fusion. Our model efficiently preserves spatial and temporal interdependence in the sensor data by integrating convolutional layers and residual BiGRU blocks. This is achieved by the architecture's design, which incorporates residual connections and layer normalization. Furthermore, incorporating IMU and EMG data offers further information, enabling our model to acquire additional distinctive characteristics for precisely identifying gym exercises. The findings demonstrate the originality and efficiency of CNN-ResBiGRU in using multimodal sensor data to achieve reliable and precise categorization of gym workouts. This establishes it as a promising method in this field.

5.4. Practical Applications of Gym Exercise Recognition

The present study examines the recognition capabilities of deep learning networks in the context of gym activity identification. The superior outcomes achieved by the CNN-ResBiGRU model demonstrate its potential for implementation in practice.

Gym exercise identification systems have applications that go beyond just classifying activities. These applications promise to improve personal fitness by observing and tracking medical compliance, ensuring safety, and assessing exercise intensity. By offering precise and comprehensive workout records, these systems can enable users to make choices based on their data, establish attainable objectives, and visually represent their progress over time. This phenomenon serves as a source of motivation for people to maintain their dedication to their fitness endeavors but also obviates the requirement for inaccurate manual data input. In addition, the use of remote monitoring capabilities allows healthcare experts and fitness instructors to impartially track the compliance of patients and trainees with recommended workout programs, eliminating the need for physical presence. Exercise recognition systems can identify muscle failure and rapidly notify monitors when persons engage in heavy-weight training, possibly reducing the risk of severe injuries. Additionally, by examining exercise execution quality, these systems can provide vital information to enhance performance, reduce the likelihood of damage, and guarantee that people get the utmost advantages from their workouts. Although evaluating the quality of exercise is a complicated area of study, its practical use has great potential to transform the field of fitness technology in gyms.

5.5. Constraints and Prospective Advancements

The present investigation provides valuable perspectives on the effectiveness of the CNN-ResBiGRU model in identifying physical exercises performed in a gymnasium setting utilizing multimodal sensor data. Nonetheless, it is imperative to acknowledge the constraints of our endeavor and pinpoint potential avenues for future exploration.

Our work is built upon the Myogym dataset, a comprehensive and widely used dataset in the industry. While this dataset is a standard for comparison, it may not fully represent the range of gym activities and user demographics in real-world scenarios. To address this, further investigation might examine the applicability of our results by assessing the suggested CNN-ResBiGRU model on supplementary datasets, thereby enhancing the validity of our findings.

Another constraint is the exclusive emphasis on data obtained from a solitary wearable device, namely the Myo armband, which is worn on the right forearm. Although this device offers comprehensive IMU and EMG data, the effectiveness of the suggested model might be affected by the precise positioning and setup of the sensors. Subsequent research might explore the influence of other sensor placements on identification precision, such as the wrist, upper arm, or lower body. In addition, investigating the use of numerous wearable devices or including data from other sensory modalities, such as heart rate or breathing, can improve the strength and flexibility of the identification system.

6. Conclusions

This study presents empirical evidence supporting the effectiveness of a newly developed hybrid deep learning model, CNN-ResBiGRU, in accurately identifying gym exercise activities using multimodal EMG and IMU sensor data. The model integrates a CNN, residual connections, and BiGRU to capture spatial and temporal characteristics from the sensor data. Extensive experiments on the Myogym dataset, comprising data from 10 individuals performing 30 different gym activities, validated the superiority of the CNN-ResBiGRU model. The most promising results were obtained by combining IMU and EMG data from the Myo armband, leveraging 14 sensor channels (six for IMU and eight for EMG data). When integrating IMU and EMG data, the model achieves an accuracy of 97.29% and an F1-score of 92.68%, surpassing baseline models in various sensor modality scenarios.

Rigorous ablation investigations provide insights into the contributions of components within the CNN-ResBiGRU architecture. Convolutional blocks capture spatial patterns, while the ResBiGRU block represents temporal relationships and long-term contextual information, which is particularly significant for EMG data. Examining multiple sensor modalities highlights the synergistic relationship between IMU and EMG data in gym activity detection, resulting in optimal performance by leveraging diverse information sources.

The study's findings demonstrate the practical utility of the CNN-ResBiGRU model in fitness monitoring, tailored training, and injury prevention. Providing precise exercise histories empowers users to make informed decisions, set achievable goals, and visually track progress. The model's computational efficiency enables real-time deployment on embedded devices, making it suitable for remote monitoring and feedback systems.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Müller, P.N.; Müller, A.J.; Achenbach, P.; Göbel, S. IMU-Based Fitness Activity Recognition Using CNNs for Time Series Classification. *Sensors* **2024**, *24*, 742. [CrossRef]
2. Mekruksavanich, S.; Jitpattanakul, A. A Deep Learning Network with Aggregation Residual Transformation for Human Activity Recognition Using Inertial and Stretch Sensors. *Computers* **2023**, *12*, 141. [CrossRef]
3. Patalas-Maliszewska, J.; Pajak, I.; Krutz, P.; Pajak, G.; Rehm, M.; Schlegel, H.; Dix, M. Inertial Sensor-Based Sport Activity Advisory System Using Machine Learning Algorithms. *Sensors* **2023**, *23*, 1137. [CrossRef]
4. Concha-Pérez, E.; Gonzalez-Hernandez, H.G.; Reyes-Avendaño, J.A. Physical Exertion Recognition Using Surface Electromyography and Inertial Measurements for Occupational Ergonomics. *Sensors* **2023**, *23*, 9100. [CrossRef]
5. Mahyari, A.; Pirolli, P.; LeBlanc, J.A. Real-Time Learning from an Expert in Deep Recommendation Systems with Application to mHealth for Physical Exercises. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 4281–4290. [CrossRef]
6. Morshed, M.G.; Sultana, T.; Alam, A.; Lee, Y.K. Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities. *Sensors* **2023**, *23*, 2182. [CrossRef]
7. Barbosa, W.A.; Leite, C.D.F.C.; Reis, C.H.O.; Machado, A.F.; Bullo, V.; Gobbo, S.; Bergamin, M.; Lima-Leopoldo, A.P.; Vancini, R.L.; Baker, J.S.; et al. Effect of Supervised and Unsupervised Exercise Training in Outdoor Gym on the Lifestyle of Elderly People. *Int. J. Environ. Res. Public Health* **2023**, *20*, 7022. [CrossRef]
8. Hussain, A.; Zafar, K.; Baig, A.R.; Almakki, R.; AlSuwaidan, L.; Khan, S. Sensor-Based Gym Physical Exercise Recognition: Data Acquisition and Experiments. *Sensors* **2022**, *22*, 2489. [CrossRef]
9. Pathan, N.S.; Talukdar, M.T.F.; Quamruzzaman, M.; Fattah, S.A. A Machine Learning based Human Activity Recognition during Physical Exercise using Wavelet Packet Transform of PPG and Inertial Sensors data. In Proceedings of the 2019 4th International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, 20–22 December 2019; pp. 1–5. [CrossRef]
10. Li, J.H.; Tian, L.; Wang, H.; An, Y.; Wang, K.; Yu, L. Segmentation and Recognition of Basic and Transitional Activities for Continuous Physical Human Activity. *IEEE Access* **2019**, *7*, 42565–42576. [CrossRef]
11. Bouchabou, D.; Nguyen, S.M.; Lohr, C.; LeDuc, B.; Kanellos, I. A Survey of Human Activity Recognition in Smart Homes Based on IoT Sensors Algorithms: Taxonomies, Challenges, and Opportunities with Deep Learning. *Sensors* **2021**, *21*, 6037. [CrossRef]
12. Aquino, G.; Costa, M.G.F.; Filho, C.F.F.C. Explaining and Visualizing Embeddings of One-Dimensional Convolutional Models in Human Activity Recognition Tasks. *Sensors* **2023**, *23*, 4409. [CrossRef]
13. Mekruksavanich, S.; Jitpattanakul, A. LSTM Networks Using Smartphone Data for Sensor-Based Human Activity Recognition in Smart Homes. *Sensors* **2021**, *21*, 1636. [CrossRef]
14. Mekruksavanich, S.; Jitpattanakul, A. Deep Convolutional Neural Network with RNNs for Complex Activity Recognition Using Wrist-Worn Wearable Sensor Data. *Electronics* **2021**, *10*, 1685. [CrossRef]
15. Webber, M.; Rojas, R.F. Human Activity Recognition With Accelerometer and Gyroscope: A Data Fusion Approach. *IEEE Sens. J.* **2021**, *21*, 16979–16989. [CrossRef]
16. Masum, A.K.M.; Bahadur, E.H.; Shan-A-Alahi, A.; Uz Zaman Chowdhury, M.A.; Uddin, M.R.; Al Noman, A. Human Activity Recognition Using Accelerometer, Gyroscope and Magnetometer Sensors: Deep Neural Network Approaches. In Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 6–8 July 2019; pp. 1–6. [CrossRef]
17. Ashry, S.; Gomaa, W.; Abdu-Aguye, M.G.; El-borae, N. Improved IMU-based Human Activity Recognition using Hierarchical HMM Dissimilarity. In Proceedings of the 17th International Conference on Informatics in Control, Automation and Robotics—ICINCO, Online, 7–9 July 2020; INSTICC, SciTePress: Setúbal, Portugal , 2020; pp. 702–709. [CrossRef]
18. Nurhanim, K.; Elamvazuthi, I.; Izhar, L.; Capi, G.; Su, S. EMG Signals Classification on Human Activity Recognition using Machine Learning Algorithm. In Proceedings of the 2021 8th NAFOSTED Conference on Information and Computer Science (NICS), Hanoi, Vietnam, 21–22 December 2021; pp. 369–373. [CrossRef]
19. Zia ur Rehman, M.; Waris, A.; Gilani, S.O.; Jochumsen, M.; Niazi, I.K.; Jamil, M.; Farina, D.; Kamavuako, E.N. Multiday EMG-Based Classification of Hand Motions with Deep Learning Techniques. *Sensors* **2018**, *18*, 2497. [CrossRef]
20. Ding, Z.; Yang, C.; Tian, Z.; Yi, C.; Fu, Y.; Jiang, F. sEMG-Based Gesture Recognition with Convolution Neural Networks. *Sustainability* **2018**, *10*, 1865. [CrossRef]
21. Faust, O.; Hagiwara, Y.; Hong, T.J.; Lih, O.S.; Acharya, U.R. Deep learning for healthcare applications based on physiological signals: A review. *Comput. Methods Programs Biomed.* **2018**, *161*, 1–13. [CrossRef]
22. Lee, K.H.; Min, J.Y.; Byun, S. Electromyogram-Based Classification of Hand and Finger Gestures Using Artificial Neural Networks. *Sensors* **2022**, *22*, 225. [CrossRef]
23. Wang, J.; Sun, S.; Sun, Y. A Muscle Fatigue Classification Model Based on LSTM and Improved Wavelet Packet Threshold. *Sensors* **2021**, *21*, 6369. [CrossRef]
24. Xiong, D.; Zhang, D.; Zhao, X.; Zhao, Y. Deep Learning for EMG-based Human-Machine Interaction: A Review. *IEEE/CAA J. Autom. Sin.* **2021**, *8*, 512–533. [CrossRef]

25. Elamvazuthi, I.; Duy, N.; Ali, Z.; Su, S.; Khan, M.A.; Parasuraman, S. Electromyography (EMG) based Classification of Neuromuscular Disorders using Multi-Layer Perceptron. *Procedia Comput. Sci.* **2015**, *76*, 223–228. [CrossRef]

26. Cai, S.; Chen, Y.; Huang, S.; Wu, Y.; Zheng, H.; Li, X.; Xie, L. SVM-Based Classification of sEMG Signals for Upper-Limb Self-Rehabilitation Training. *Front. Neurorobot.* **2019**, *13*, 31. [CrossRef]

27. Di Nardo, F.; Morbidoni, C.; Cucchiarelli, A.; Fioretti, S. Influence of EMG-signal processing and experimental set-up on prediction of gait events by neural network. *Biomed. Signal Process. Control* **2021**, *63*, 102232. [CrossRef]

28. Nazmi, N.; Abdul Rahman, M.A.; Yamamoto, S.I.; Ahmad, S.A. Walking gait event detection based on electromyography signals using artificial neural network. *Biomed. Signal Process. Control* **2019**, *47*, 334–343. [CrossRef]

29. Koskimäki, H.; Siirtola, P.; Röning, J. MyoGym: Introducing an open gym data set for activity recognition collected using myo armband. In Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, New York, NY, USA, 11–15 September 2017; UbiComp '17; pp. 537–546. [CrossRef]

30. Jung, P.G.; Lim, G.; Kim, S.; Kong, K. A Wearable Gesture Recognition Device for Detecting Muscular Activities Based on Air-Pressure Sensors. *IEEE Trans. Ind. Inform.* **2015**, *11*, 485–494. [CrossRef]

31. Crema, C.; Depari, A.; Flammini, A.; Sisinni, E.; Haslwanter, T.; Salzmann, S. IMU-based solution for automatic detection and classification of exercises in the fitness scenario. In Proceedings of the 2017 IEEE Sensors Applications Symposium (SAS), Glassboro, NJ, USA, 13–15 March 2017; pp. 1–6. [CrossRef]

32. Pernek, I.; Kurillo, G.; Stiglic, G.; Bajcsy, R. Recognizing the intensity of strength training exercises with wearable sensors. *J. Biomed. Inform.* **2015**, *58*, 145–155. [CrossRef]

33. Hochreiter, S. The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **1998**, *6*, 107–116. [CrossRef]

34. Cho, K.; van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. In Proceedings of the SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, Doha, Qatar, 25 October 2014; pp. 103–111. [CrossRef]

35. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. In Proceedings of the NIPS 2014 Workshop on Deep Learning, Montreal, QC, Canada, 13 December 2014.

36. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]

37. Aşuroğlu, T.; Açici, K.; Erdaş, ç.B.; Oğul, H. Texture of Activities: Exploiting Local Binary Patterns for Accelerometer Data Analysis. In Proceedings of the 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Naples, Italy, 28 November–1 December 2016; pp. 135–138. [CrossRef]

38. Montaha, S.; Azam, S.; Rafid, A.K.M.R.H.; Ghosh, P.; Hasan, M.Z.; Jonkman, M.; De Boer, F. BreastNet18: A High Accuracy Fine-Tuned VGG16 Model Evaluated Using Ablation Study for Diagnosing Breast Cancer from Enhanced Mammography Images. *Biology* **2021**, *10*, 1347. [CrossRef]

39. de Vente, C.; Boulogne, L.H.; Venkadesh, K.V.; Sital, C.; Lessmann, N.; Jacobs, C.; Sánchez, C.I.; van Ginneken, B. Improving Automated COVID-19 Grading with Convolutional Neural Networks in Computed Tomography Scans: An Ablation Study. *arXiv* **2020**, arXiv:2009.09725.

40. Meyes, R.; Lu, M.; de Puiseau, C.W.; Meisen, T. Ablation Studies in Artificial Neural Networks. *arXiv* **2019**, arXiv:1901.08644.

41. Ojiako, K.; Farrahi, K. MLPs Are All You Need for Human Activity Recognition. *Appl. Sci.* **2023**, *13*, 11154. [CrossRef]

42. Ismail Fawaz, H.; Lucas, B.; Forestier, G.; Pelletier, C.; Schmidt, D.F.; Weber, J.; Webb, G.I.; Idoumghar, L.; Muller, P.A.; Petitjean, F. InceptionTime: Finding AlexNet for time series classification. *Data Min. Knowl. Discov.* **2020**, *34*, 1936–1962. [CrossRef]

43. Ordóñez, F.J.; Roggen, D. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* **2016**, *16*, 115. [CrossRef] [PubMed]

44. Wang, Z.; Yan, W.; Oates, T. Time series classification from scratch with deep neural networks: A strong baseline. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 1578–1585. [CrossRef]