

Article

Myo Transformer Signal Classification for an Anthropomorphic Robotic Hand

Bolivar Núñez Montoya ¹, Edwin Valarezo Añazco ^{1,*}, Sara Guerrero ², Mauricio Valarezo-Añazco ¹, Daniela Espin-Ramos ³ and Carlos Jiménez Farfán ³

¹ Faculty of Engineering in Electricity and Computation (FIEC), Escuela Superior Politécnica del Litoral (ESPOL), Guayaquil 090112, Ecuador; bsnunez@espol.edu.ec (B.N.M.); mvalarez@espol.edu.ec (M.V.-A.)

² Faculty of Architecture and Design, Universidad Espíritu Santo, Samborondón 0901952, Ecuador; saguerrero@uees.edu.ec

³ Faculty of Mechanical Engineering and Production Sciences (FIMCP), Escuela Superior Politécnica del Litoral (ESPOL), Guayaquil 090112, Ecuador; danaespi@espol.edu.ec (D.E.-R.); cajifar@espol.edu.ec (C.J.F.)

* Correspondence: edgivala@espol.edu.ec

Abstract: The evolution of anthropomorphic robotic hands (ARH) in recent years has been sizable, employing control techniques based on machine learning classifiers for myoelectric signal processing. This work introduces an innovative multi-channel bio-signal transformer (MuCBiT) for surface electromyography (EMG) signal recognition and classification. The proposed MuCBiT is an artificial neural network based on fully connected layers and transformer architecture. The MuCBiT recognizes and classifies EMG signals sensed from electrodes patched over the arm's surface. The MuCBiT classifier was trained and validated using a collected dataset of four hand gestures across ten users. Despite the smaller size of the dataset, the MuCBiT achieved a prediction accuracy of 86.25%, outperforming traditional machine learning models and other transformer-based classifiers for EMG signal classification. This integrative transformer-based gesture recognition promises notable advancements for ARH development, underscoring prospective improvements in prosthetics and human-robot interaction.

Keywords: convolutional transformer; deep learning; bionic prosthetics; EMG data classification; anthropomorphic robotic hand



Citation: Núñez Montoya, B.; Valarezo Añazco, E.; Guerrero, S.; Valarezo-Añazco, M.; Espin-Ramos, D.; Jiménez Farfán, C. Myo Transformer Signal Classification for an Anthropomorphic Robotic Hand. *Prosthesis* **2023**, *5*, 1287–1300. <https://doi.org/10.3390/prosthesis5040088>

Academic Editors: Peter Kyberd and Alix Chadwell

Received: 17 September 2023

Revised: 11 November 2023

Accepted: 15 November 2023

Published: 28 November 2023



Copyright: © 2023 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

Recent years have witnessed significant advancements in the field of robotic prosthetics research. One of the most prominent challenges for researchers is designing anthropomorphic robotic hands that are capable of accurately replicating the appearance, movement, and control based on the bio-signals of human hands. In this context, using myoelectric sensors has emerged as a promising strategy to control robotic prostheses. Myoelectric sensors allow for the capture and analyzation of bio-electrical signals generated by muscle activity using sensors patched over the surface of the skin, avoiding invasive bio-signal sensing [1] and achieving the intuitive and natural control of the prosthesis.

In electromyography (EMG), invasive electrodes are characterized by needles inserted into specific muscle groups, aiming to replicate delicate movements. However, their intrusive nature limits the prosthesis use to a particular user and exposes them to risks such as infections due to unsanitary conditions or improper care [2]. Alternatively, strategically positioned surface electrodes on the dermis offer a non-invasive interface for sensing versatile and ergonomic EMG data [3]. Surface electrodes are particularly suitable for monitoring and controlling prostheses in everyday use and facilitate adaptability to various users, establishing themselves as a preferred option for more generalized applications [4].

Anthropomorphism plays a pivotal role in developing dexterous robotic prosthetic hands designed to assist individuals with different abilities. To achieve this goal, different designing approaches have been explored, for instance, the combination of distributed actuation [5], dual-mode torque actuation, and joint locking mechanism [6], and methods of flexing and tensioning with ropes (i.e., artificial tendons) [7] used to emulate human muscle action forces to predict the hand gesture strength in each finger movement [8]. These previous approaches have limitations regarding the angular mobility of the joints. On the other hand, a fully mechanical motor-driven drive is presented in [9], allowing for the full motion of the hand's fingers. Also, a design based on the bone structure of the hand was proposed to simulate how tendons provide firmness and degrees of mobility in each hand section [10]. In addition, 3D printing has been widely used in the development of anthropomorphic robotic hand designs due to its advantages in terms of fabrication time and printing materials. Among different materials available, Thermoplastic Polyurethane (TPU), developed by the authors of [11], has become a common choice in the production of prostheses. This is attributed to its impressive flexibility, exceptional durability, and resistance to breakage.

Machine learning (ML) algorithms have played essential roles in the control of robotic prosthetic hands. Remarkable examples include artificial neural networks (ANNs) [12], Support Vector Machines (SVMs) [13,14], Linear Discriminant Analysis (LDA) [13], Decision Trees (DT), multi-layer perceptron (MLP) [15], and Encoder-Decoder Temporal Convolutional Networks (ED-TCNs) [16]. These techniques aim to recognize and classify patterns within myoelectric signals. While prior research has yielded promising results, the effectiveness of classification varies based on the extracted features [17]. For instance, Phinyomark et al. [18] achieved an average classification rate of 82.57% using nine relevant features from an EMG dataset of 52 hand gestures with a sampling rate of 200 Hz and 16 electrodes. Nunez-Montoya et al. [15] reached a classification rate of 74.78% using 20 relevant features from an EMG dataset of four hand gestures with a sampling rate of 250 Hz and 12 electrodes (6 channels). Betthausen et al. [16] reached a classification rate of 72.10% using five relevant features from an EMG dataset of four hand gestures sensed with eight EMG sensors built in the Myo Armband at a sampling rate of 200 Hz. Chung et al. [12] reached a classification rate of 85.08% using an autoencoder for automatic feature extraction from an EMG dataset of five hand gestures with a Myo Armband. Wang et al. [13] reached a classification rate of 83% using five relevant features with a sampling rate of 1000 Hz and four electrodes (two channels) with their own EMG signal recording system. A different approach for EMG data collection focuses on human hand gestures that include the motion of all fingers to perform grasping positions. For instance, Yang et al. [14] reached a classification rate of 93.7% using one feature from an EMG dataset of eighteen hand gestures with a sampling rate of 100 Hz and 50 electrodes. Chen et al. [19] reached a classification rate of 93.1% using 5 relevant gestures from an EMG dataset of 25 hand gestures with a sampling rate of 2000 Hz and two electrodes (two channels). The field is evolving to encompass a broader range of hand gestures, including those relevant to prosthetic control and human-computer interaction [20]. Despite these achievements, traditional ML techniques, like those mentioned above, often face limitations in handling high-dimensionality and time series relations, which are inherent in myoelectric signals. Therefore, it may become useful to explore deeper architectures that can capture long-term dependencies and patterns.

Deep learning (DL) is a subfield of ML that emphasizes deep-layered models inspired by the brain structure (e.g., ANNs) and has demonstrated its effectiveness in the biomedical field for pattern recognition in images and time series bio-signals [21–24]. Transformer networks are DL models that have emerged as innovative architecture in natural language processing and computer vision [20]. Transformers are based on attention mechanisms and have consistently demonstrated outstanding performance in machine translation tasks, text generation, object recognition, and more [25,26]. Their ability to capture long-term relationships and contextual patterns in time series signals has revolutionized sequential

data processing and opened new opportunities in diverse research domains, for instance, Human Activity Recognition (HAR) [27] and its applicability in recognizing specific hand gesture actions based on EMG data. In the state of the art, Xiao et al. [28] implemented a self-attention mechanism using a Two-stream Transformer Network (TTN) to model temporal-spatial dependencies in multimodal HAR. Furthermore, there has been a growing interest in the development of Transformer models for the time series of biomedical data, such as electrocardiogram (ECG) signals [29–31] and EMG signals using Vision Transformer (ViT) models [32,33].

In this paper, we introduce the multi-channel bio-signal transformer (MuCBiT), an innovative approach to recognize and classify hand gestures sensed through EMG electrodes placed on strategic areas of the arm's surface, aiming to replicate the human hand motions in an anthropomorphic robotic hand (ARH). A printable ARH is used to evaluate the MuCBiT. Artificial tendons operate the printable ARH to mimic the contraction and extension of human hand muscles. To evaluate the proposed bio-signal classification, we utilize a collected database and compare its performance to prior research [15] and other transformer models implemented in previous research articles. The database is split into training, validation, and testing datasets. The MuCBiT model achieves 86.25% with the validation dataset and 86.78% with the testing dataset.

2. Materials and Methods

Figure 1 shows an overview of the implemented system to control the ARH. The EMG DAQ section involves sensing EMG data using a Cyton Board, which captures the myoelectrical signals of the user through six channels placed on their forearm. These signals are then transmitted to a server via serial communication using the Lab Streaming Layer (LSL) application. LSL is compatible with Python and facilitates data processing. The signal classification stage encodes the EMG signals into four hand gestures using the MuCBiT model. Finally, the predicted result is sent to an ESP32 servo-controller board, which controls the servo motors using multi-channel output. It applies a Single-Input Single-Output (SISO) positional control approach to modulate the PWM signals and execute the corresponding hand gesture in the ARH.

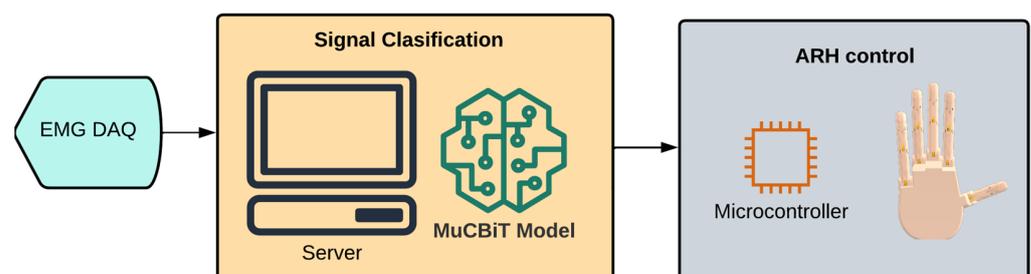


Figure 1. Overview of the proposed system.

2.1. Anthropomorphic Robotic Hand

The printable ARH used in this work was our design based on the work of [15], where a mechanical system driven by servo motors was used. Our design was inspired by the muscles' flexion and extension system responsible for the finger movements of the human hand. The human anatomy is shown in Figure 2. The movements of the index, middle, ring, and little fingers are controlled by both the muscles in the upper and lower regions of the forearm. The thumb is controlled by smaller muscle groups in different arm areas because of its unique dexterity. Our design imitates this complex coordination of muscles, allowing the robotic hand to replicate human-like movements more accurately.

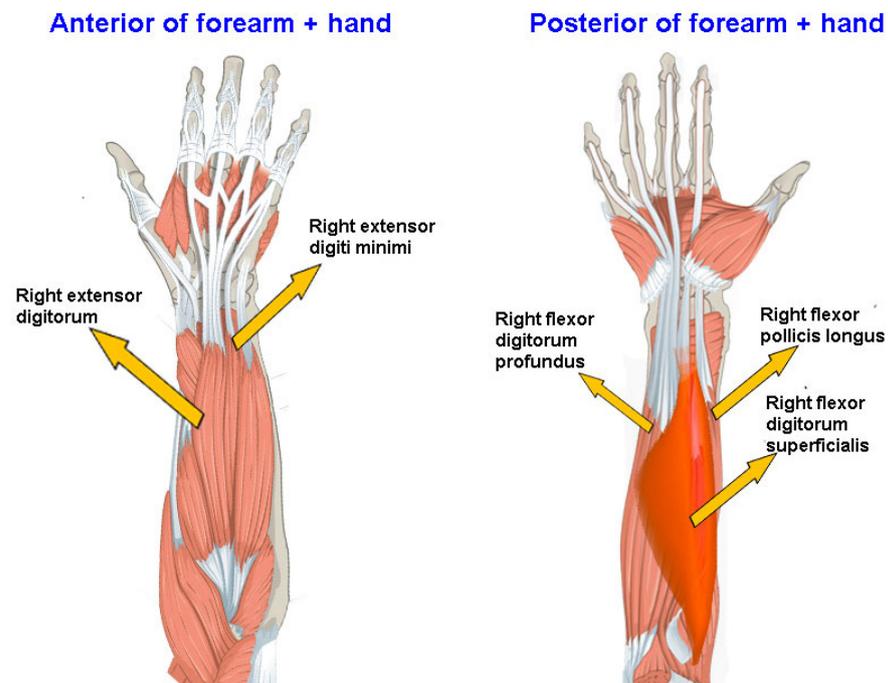


Figure 2. Main muscles that drive human fingers.

A rope system was implemented to control the fingers of the anthropomorphic robotic hand (ARH), motivated by the biomechanics of flexing and tensing the muscles of human arms [34]. Also, to reduce the number of required actuators, we considered the relations between fingers when they move, as noted in [35]. These interfinger relations can be appreciated in Table 1.

Table 1. Relation between the movements of the fingers (“-” symbolizes independence and “x” symbolizes dependence).

	Index	Middle	Ring	Little
Index	x	x	-	-
Middle	x	x	x	-
Ring	-	-	x	x
Little	-	-	x	x

Rows are the target finger motions; columns are the relations between the motions.

The ARH will feature a 3D-printed membrane using TPU, which provides flexibility and emulates natural forces in hand motion, such as friction. The proposed design can be seen in Figure 3.

2.2. Database

The database used for this research was obtained from previous experiments [15]. It includes data gathered from ten participants, evenly divided between genders, ranging in age from 18 to 50 years. Data acquisition was facilitated using a Python script, which recorded the streaming EMG signal from the six channels transmitted by the Open-BCI’s Cyton board at a sampling rate of 250 Hz. Each of the four hand gestures was repeated 20 times by all participants, with a two-second interval between each repetition, as illustrated in Figure 4. The presentation order of these hand gestures was randomized within each loop for user performance. Finally, the acquired data from the experiments was stored in *.csv format for further analysis.

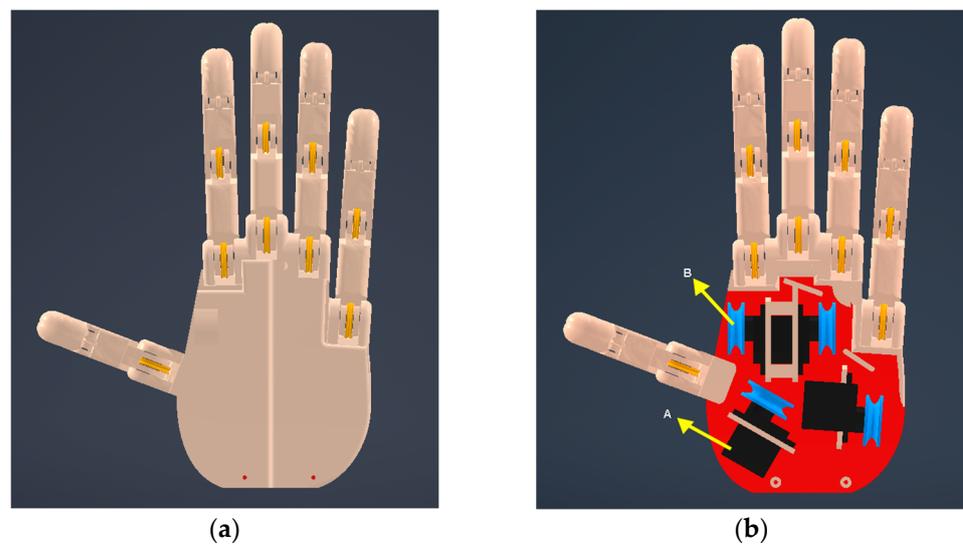


Figure 3. Three-dimensional design of ARH: (a) assembled design; (b) internal structure of ARH, where A represents servomotors in black, B represents pulleys in the blue.

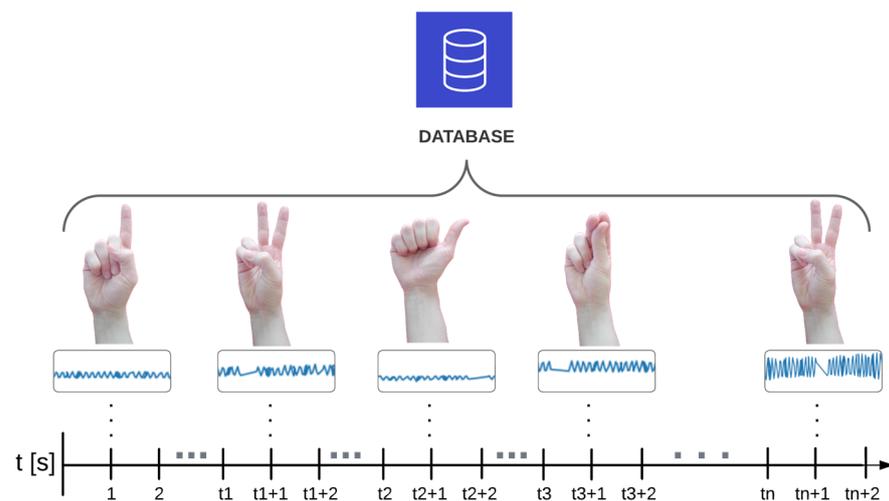


Figure 4. EMG data acquisition protocol.

We selected these four specific hand gestures due to the activation of the muscles we analyzed in this study, as shown in Figure 2. Each gesture was chosen because it involves an ordered and synchronized motion of each finger. Also, some of these hand motions were used in other studies [12–15,19,33]. This approach enabled us to conduct a comprehensive examination of the muscle activation patterns associated with various finger movements, enhancing the overall scope of our study.

Processing Data

The collected EMG time series signals were not filtered to preserve the nature of the data. The data were organized into a time-dependent tensor to strengthen our MuCBiT model to capture the relevant information. Figure 5 shows the set of EMG signals of the six channels as a function of time. A specific time frame was taken as an example, which allowed us to visualize how the EMG signals varied over time in different channels.

This tensor representation of the data allowed us to capture the temporal and spatial information of the EMG signals, which is essential to analyze and classify hand gestures.

The variability in motion speed among the participants resulted in variations in the lengths of collected motions. Therefore, we performed data filling to standardize the data

length to 512 myoelectric data points recorded in each movement. This standardized the length of the time sequence and ensured consistency for data analysis and processing.

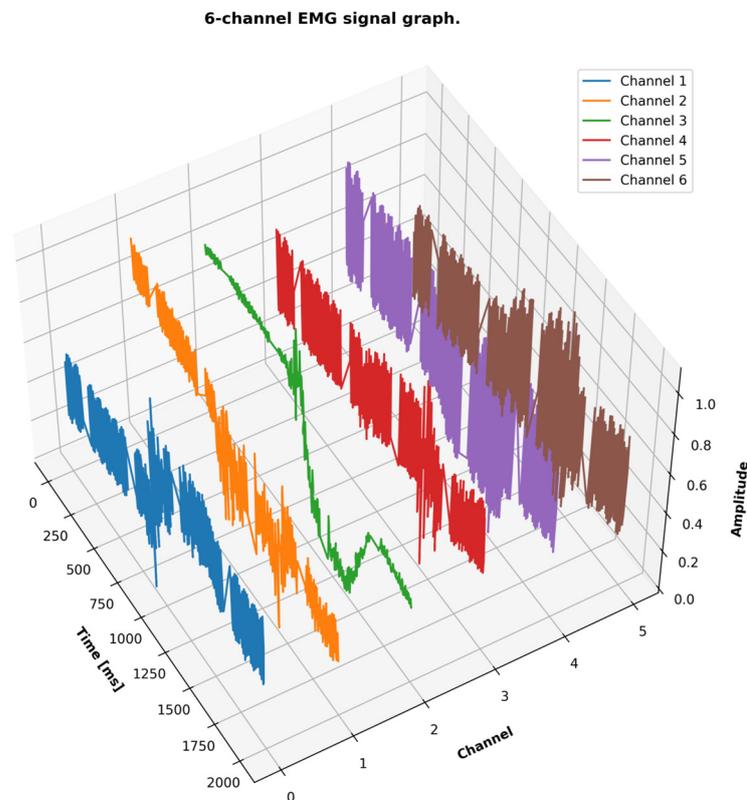


Figure 5. Graph channels vs. time of an executed movement.

2.3. Mio-Transformer Network

The proposed MuCBiT is based on a ViT. The ViT is a variation of the original Transformer utilized to perform computer vision tasks [30,33,36,37]. The ViT network architecture comprises an encoder to extract relations between the time series signals and a decoder to produce an output based on the relations and features extracted in the encoder. The proposed transformer-based architecture for multi-channel biological signal analysis is depicted in Figure 6. The MuCBiT uses the key transformer component, the Multi-Head Self-Attention (MHSA) mechanism from the encoder. This advanced encoding mechanism allows for relations or dependencies to be found between the time series signals and the multi-channel.

2.3.1. Embedding Patches

The embedding patches split the data into patches, and then projected them into a high-dimensional space through an embedding layer. Classification tokens (CLS tokens) were added at the beginning of the input sequence, and they were embedded with positional information before being passed through the transformer [38,39].

Unlike in other transformer models oriented towards natural language processing (NLP), the CLS token does not represent any actual word in the sequence. It acts as a reference point for the classification task, assimilating the acquired features related to the class.

EMG signals are an ordered sequence of data; keeping this order is critical to correctly interpret the sequence and recognize patterns in the EMG signals of the subject [16]. Sinusoidal and cosine functions at different frequencies were employed as the positional embedding (PE), according to Equation (1),

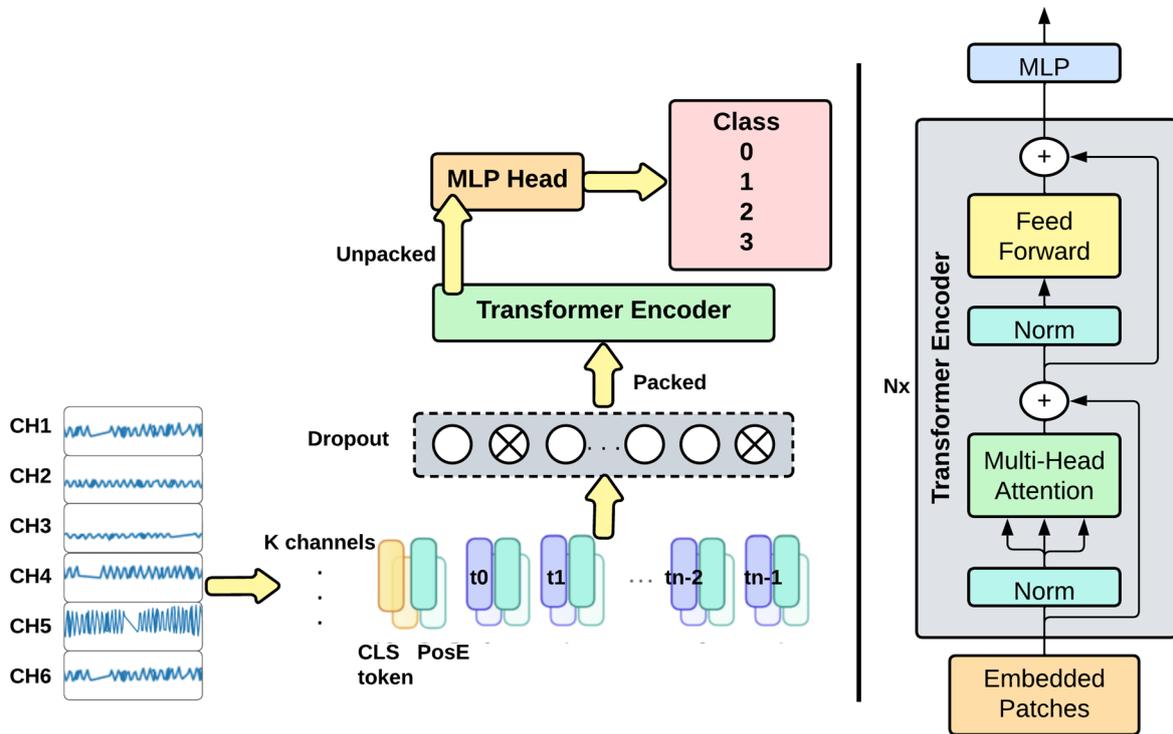


Figure 6. Architecture of the MuCBiT model.

$$PE \begin{cases} P_{i,2t} = \sin\left(\frac{pos}{10,000^{2t/d}}\right) \\ P_{i,2t+1} = \cos\left(\frac{pos}{10,000^{2t/d}}\right) \end{cases} \quad (1)$$

where *pos* is the position of an object in the input sequence, and *d* is the embedding output dimension [40]. This approach helps to capture the temporal relationships within the data, making the model more sensitive to the temporal ordering of the EMG signals and enhancing its prediction accuracy.

2.3.2. Transformer Encoder

Our proposed encoder architecture is derived from the standard structure [40], using a normalization layer that precedes each MHSA layer and feedforward layer. This architecture ensures that the data are normalized before any subsequent operations.

Figure 7 shows the architecture of our MHSA. The MHSA module integrates a dropout mechanism applied to the attention vector post-softmax. This enhances the model’s robustness and its ability to generalize from the training data. The dropout is set to zero with some attention weights during training to prevent the model from being over-reliant on any single attention position on any individual attention position. Thus, its ability to adapt to new (i.e., unseen) data strengthens.

The computation of the attention function is explained in Equation (2) [38]:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_K}}\right) \quad (2)$$

where d_K represents the dimensionality of key values. *V*, *K*, and *Q* are the value, key, and query, respectively, according to the original transformer model [40].

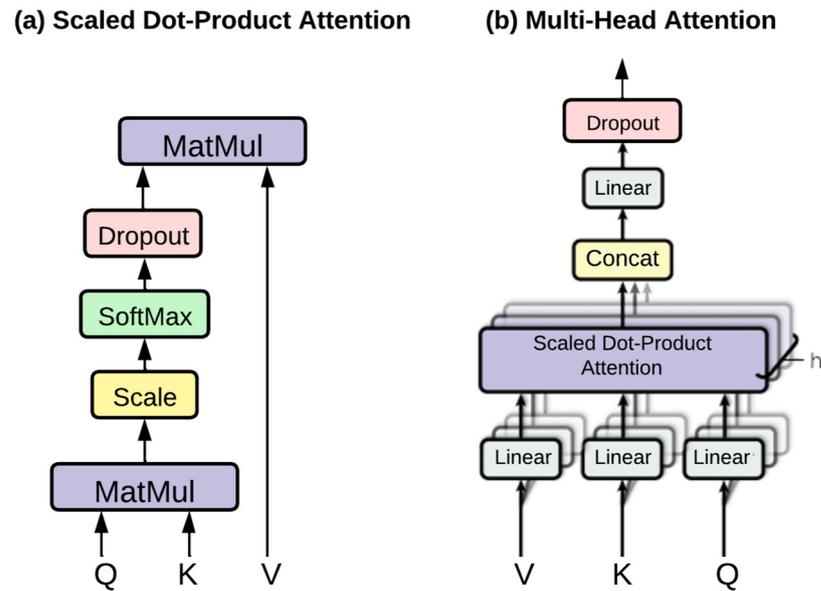


Figure 7. (a) Scaled dot-product attention. (b) Multi-head attention layer architecture.

The multi-layer perceptron is a set of feedforward layers in line. The feedforward layers use the Gaussian Error Linear Units (GELU) activation function [41]. The GELU captures complex patterns within the EMG data more effectively. This section produces classification labels based on the high-level representations generated by the encoder, thereby enabling the system to make informed decisions about motion classification.

Furthermore, an extra dropout regularization is employed within the feedforward layer to mitigate overfitting and enhance the ability to generalize the model.

This regularization strategy prevents the model from becoming overly dependent on individual neurons and ensures a balance between bias and variance.

2.3.3. Model Configuration

The configuration of the MuCBiT model is defined by a set of key parameters that control various aspects of its architecture and behavior in training. The key parameters include the following:

- **seq_len:** This parameter defines the length of the input sequence to the model, which, in this case, is set to 512 sample points. This means that the model processes sequences of 512 data points at a time.
- **channels:** The number of input channels refers to the six channels used to sense the EMG signal data.
- **patch_size:** This parameter controls the sizes of the patches used in the model, which are set to 4, 8, 16, and 32.
- **num_classes:** The number of distinct classes the model is designed to classify is set to 4, which depends on the number of different hand gestures you want to identify in the EMG signals.
- **dim:** The dimension of the embedding space into which the input sequences are transformed. The embedding space was trained with 1024 and 2048 dimensions.
- **depth:** The depth of the transformer network, i.e., the number of transformer layers in the model. The model was tested with 6, 8, and 10 transformer layers.
- **heads:** The number of multi-head attention layers, set to 8, 12, and 16 layers in this case.
- **dim_head:** The dimension of each MHSA of the transformer layers. In this case, each head has a dimension of 32 or 64.
- **mlp_dim:** The dimension of the MLP layer, which was set to 1024 and 2048.
- **dropout:** The dropout rate applied to the transformer encoder layers was set to 0.05.
- **emb_dropout:** The dropout rate applied to the embedded patch was set to 0.05.

2.4. Training Details

The database consisted of 800 samples of the four hand gestures. We performed a training and validation split with an 8:2 ratio (i.e., 640 samples for training and 160 for validation). Additionally, a testing dataset of 400 samples was generated using EMG signals from four women and one man.

During the training process, different hyperparameter settings were configured and evaluated; the number of epochs was initially set to 100 and increased to 120, and batch sizes of 5, 10, and 30 were tested. Learning rates of 1×10^{-5} and 1×10^{-4} were used alongside an AdamW optimizer, which is a variant of the conventional Adam optimizer with weight decay rates of 0.01 and 1×10^{-4} .

2.5. Framework

In this study, EMG data are acquired without undergoing any data preprocessing, allowing the EMG data acquisition (DAQ) device to directly transmit the sensed data to a local workstation for gesture recognition.

3. Results

3.1. Parameters Tuning

The proposed model offers various configuration parameters that affect the classification accuracy. To identify the optimal settings for our model, we conducted a two-stage optimization process. In the first stage, we evaluated the correlation of each parameter with the accuracy of the model while keeping a fixed learning rate of 1×10^{-4} and a weight decay of 1×10^{-4} for 100 epochs. Subsequently, in the second stage, we performed a grid search, considering variations in the number of epochs, learning rates, and weight decay based on the best parameters identified in the first stage. This iterative approach allowed us to refine the configuration of the model for the best outcome.

Table 2 summarizes our experimentation results from the first stage, showing the top 10 configurations while considering accuracy as the evaluation metric. The highest prediction accuracy of 85% was obtained with a batch of 10, a patch size of 4, a projection dimension of 1024 in the data sequence, a depth of 10 transformer layers, 16 attention heads with 64 neurons each, and an MPL dimension of 1024.

Table 2. Table of the top 10 configurations based on accuracy with the validation dataset.

Batch	Patch Size	Dim	Depth	Heads	Dim Head	MPL Dim	Accuracy
10	4	1024	6	8	32	2048	70.00
10	8	1024	6	8	32	1024	72.50
10	16	2048	8	12	32	2048	73.75
5	4	1024	6	8	32	1024	75.00
30	16	1024	8	12	64	2048	76.87
10	8	1024	10	16	32	1024	79.00
5	4	1024	10	16	32	2048	80.00
30	4	1024	10	16	64	1024	80.55
5	4	1024	6	8	32	1024	81.87
10	4	1024	10	16	64	1024	85.00

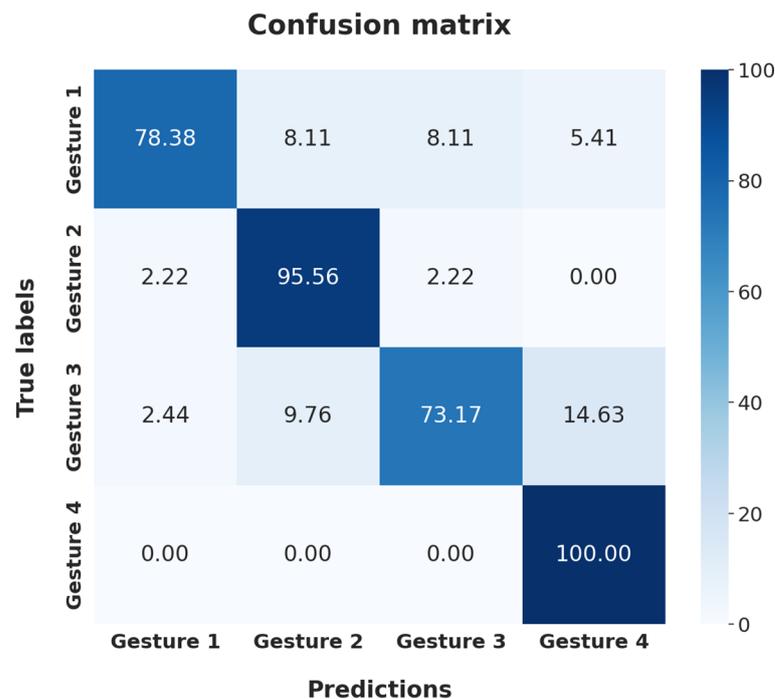
Once the MuCBiT configuration was set, different configurations of hyperparameters, such as the learning rate, weight decay, and training epochs, were fine-tuned to enhance the classification accuracy. Table 3 shows that the model reached a classification performance of 86.25% using a learning rate of 1×10^{-5} , a weight decay of 0.01, and 120 epochs.

Table 3. Hyperparameter tuning results.

N° Epochs	Weight Decay	Learning Rate	Validation Accuracy
100	1×10^{-4}	1×10^{-4}	85.00%
100	1×10^{-4}	1×10^{-5}	83.12%
100	0.01	1×10^{-4}	81.87%
100	0.01	1×10^{-5}	80.62%
120	1×10^{-4}	1×10^{-4}	83.75%
120	1×10^{-4}	1×10^{-5}	81.25%
120	0.01	1×10^{-4}	84.00%
120	0.01	1×10^{-5}	86.25%

3.2. MuCBiT Classification Results

Figure 8 displays the confusion matrix obtained following the 5-fold cross-validation process. In this matrix, the columns represent the predicted hand gestures, while the rows contain the ground truths. On the main diagonal, a minimum precision of 73.17% is observed, corresponding to gesture 3. It is followed by gesture 1, with a score of 78.38%. This indicates that the Transformer model confuses these two hand gestures. In contrast, gestures 2 and 4 show scores of 95.56% and 100%, respectively.

**Figure 8.** Confusion matrix of 4 hand gestures in percentage (%).

In Table 4, the validation accuracy of the baseline models and previous implementations is presented. The baseline models include machine learning algorithms such as KNN, polynomial, and MLP Classifier as well as DL architecture such as ED-TCN and ViT models. The previous implementation includes the works of De Godoy et al. [32] and Montazerin et al. [33], where hand gesture classification is performed using a transformer-based algorithm. As shown in Table 4, the proposed MuCBiT performs slightly better in terms of the classification accuracy. Despite not being a direct comparison with previous work due to the use of different hand gestures, Table 4 shows the usefulness of the proposed MuCBiT.

Table 4. Validation accuracy of baseline ML and DL-based models and the proposed MuCBiT.

Model	Validation Accuracy	N° Hand Gestures
KNN	72.28%	4
Polynomial	80.98%	4
MLPC	84.78%	4
MuCBiT	86.25%	4
ED-TCN	72.10%	7
ViT-HGR ¹	84.62%	66
TMC-ViT ²	89.60%	17

¹ Vision Transformer-based hand gesture recognition [33]. ² Temporal multi-channel vision transformers [32].

3.3. Classification Results: Testing Dataset

Table 5 shows the results of the model evaluated on the testing dataset. This process led us to determine the adaptability of the model to new users. Our model attained an accuracy rate of 86.78%, a precision rate of 87.50%, a recall rate of 86.78%, an F1 Score of 86.52%, and an Area Under Curve Receiver Operating Characteristic (AUCROC) of 95.98%.

Table 5. Performance results of the MuCBiT with the testing dataset.

Accuracy	Precision	Recall	F1 Score	AUCROC
0.8678	0.8750	0.8678	0.8652	0.9598

The values on the cell are in the range between 0 and 1.

4. Discussion

Our experiment's findings indicate a negative correlation between the increment of the batch and patch sizes and the accuracy achieved. Conversely, an increment in the number of Transformer and attention layers exhibited a positive correlation with the accuracy of our proposed model, suggesting that these parameters played crucial roles in capturing the complex patterns present in the myoelectric signals.

The F1 Score provides a harmonic means of precision and recall. The MuCBiT obtained an F1 Score of 86.52%, underlining the successful balance between precision and sensitivity. Moreover, the AUROC of 95.98% shows the capability of the model to distinguish between the four classes of hand gestures.

The architecture of the MuCBiT effectively extracts and identifies the features of EMG signals, capturing and mapping the contextual dependencies and long-term relationships that are inherent in EMG signals. Table 4 shows that all of the Transformer-based models outperformed the baseline classifiers based on ML and basic DL architectures such as MLP Classifier and ED-TCN. Regarding the previous implementation, our MuCBiT slightly outperformed the work by Montazerin et al. [33]. De Godoy et al. [32] and Montazerin et al. [33] used convolutional layers for local feature extraction and discerned relevant patterns within the signals. In contrast, the MuCBiT employs a patch-based decomposition of the signals, followed by their subsequent projection into a high-dimensional space. Additionally, our proposed approach provides accurate and robust hand gesture prediction capabilities, making it a promising tool for recognizing and classifying EMG signals based on EMG signals.

In contrast to our prior experiments published in [15], three preliminary filters were applied in extra preprocessing steps. The filters are a notch filter at 60 Hz, a fifth-order Butterworth high-pass filter, and low-pass filters between 10 and 100 Hz. The current study does not employ filters. This unfiltered learning enables and forces the model to search deeper into the raw and pure features of the signal, potentially enhancing the robustness and accuracy of gesture recognition. Finally, the result obtained using our MuCBiT model outperformed the MLP utilized in our previous work [15], which scored an accuracy of 84.78%.

Limitations and Future Directions

The implementation of pre-trained Transformer-based models such as Bidirectional Encoder Representations from Transformers (BERTs) [42], GPT (with the left-to-right decoder) [43], BART [44], and T5 [45], alongside the MuCBiT model, holds the potential to revolutionize the analysis of EMG signals. These transformer-based models were designed to predict sequences from preceding sequences [46]. Then, a fusion with our MuCBiT might allow for a deeper understanding of the complex dynamics of EMG data. The fusion of these models, each equipped with unique attention mechanisms, coupled with MuCBiT's domain-specific insights tailored for gesture detection, holds the potential to deliver promising outcomes, enhancing both accuracy and robustness.

The MuCBiT model stands out for its exceptional versatility within bio-signal analysis. It has emerged as a robust, generalized model in this field, primarily due to its capacity to maintain stability when confronted with novel input data such as the testing dataset. Its ability to address variability in bio-signal data makes it a promising contender for broader applications. Considering the pivotal aspects of domain diversity and model adaptability, we aim to explore the potential extension of the efficacy of our model into the realm of pattern recognition within different bio-signals, such as ECG or EEG.

The principal limitation of our proposed model lies in the size of the dataset. A more extensive and diverse dataset with a wide range of hand gestures from users without and with disabilities will ensure an extensive evaluation of the model's robustness. Then, further enhancements will be explored by training the MuCBiT model with a larger dataset.

5. Conclusions

In this work, we present a new design of a transformer-based deep learning model called MuCBiT, which successfully classifies raw EMG signals between four classes of hand gestures with a precision of 86.25% with the validation dataset and an accuracy rate of 86.78% with the testing dataset. This percentage of successful recognition and classification demonstrates the feasibility of using the model to identify EMG signals from new users accurately. Furthermore, since completely raw EMG signals are directly input into the MuCBiT, we can conclude that our proposed model is able to learn from the innate nature of the EMG bio-signals.

Integrating the transformer-based model into the control system of the ARH showcases the potential for advanced prosthetics and human–robot interaction.

Author Contributions: Conceptualization, B.N.M., E.V.A. and M.V.-A.; data curation, B.N.M., E.V.A. and M.V.-A.; investigation, S.G., C.J.F. and D.E.-R.; methodology, B.N.M., E.V.A. and M.V.-A.; software, B.N.M., E.V.A., C.J.F. and D.E.-R.; validation, B.N.M. and E.V.A.; writing—original draft, B.N.M., E.V.A. and S.G.; writing—review and editing, B.N.M., E.V.A., S.G., C.J.F. and D.E.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical approval was not required for this manuscript due to this is a non-interventional study.

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

Data Availability Statement: The data presented in this study are openly available in IEEE Data Port at <https://dx.doi.org/10.21227/ear9-rt77> (accessed on 7 July 2023) and at <https://dx.doi.org/10.21227/ndjb-x470> (accessed on 14 July 2023).

Acknowledgments: We acknowledge the Faculty of Engineering in Electricity and Computation (FIEC) for providing the equipment to train and validate the models used in this research.

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

References

1. Campanini, I.; Disselhorst-Klug, C.; Rymer, W.Z.; Merletti, R. Surface EMG in Clinical Assessment and Neurorehabilitation: Barriers Limiting Its Use. *Front. Neurol.* **2020**, *11*, 4–7. [[CrossRef](#)]
2. Ambikapathy, B.; Krishnamurthy, K. Analysis of Electromyograms Recorded Using Invasive and Noninvasive Electrodes: A Study Based on Entropy and Lyapunov Exponents Estimated Using Artificial Neural Networks. *J. Ambient Intell. Humaniz. Comput.* **2018**, 1–9. [[CrossRef](#)]
3. Merletti, R.; Rainoldi, A.; Farina, D. Surface Electromyography for Noninvasive Characterization of Muscle. *Exerc. Sport Sci. Rev.* **2001**, *29*, 20–25. [[CrossRef](#)] [[PubMed](#)]
4. Choi, C.; Micera, S.; Carpaneto, J.; Kim, J. Development and Quantitative Performance Evaluation of a Noninvasive EMG Computer Interface. *IEEE Trans. Biomed. Eng.* **2009**, *56*, 188–191. [[CrossRef](#)] [[PubMed](#)]
5. Lee, D.-H.; Park, J.-H.; Park, S.-W.; Baeg, M.-H.; Bae, J.-H. KITECH-Hand: A Highly Dexterous and Modularized Robotic Hand. *IEEE ASME Trans. Mechatron.* **2017**, *22*, 876–887. [[CrossRef](#)]
6. Shin, Y.J.; Kim, S.; Kim, K.-S. Design of Prosthetic Robot Hand with High Performances Based on Novel Actuation Principles. *IFAC Proc. Vol.* **2013**, *46*, 313–318. [[CrossRef](#)]
7. Zhou, J.; Chen, X.; Chang, U.; Lu, J.-T.; Leung, C.C.Y.; Chen, Y.; Hu, Y.; Wang, Z. A Soft-Robotic Approach to Anthropomorphic Robotic Hand Dexterity. *IEEE Access* **2019**, *7*, 101483–101495. [[CrossRef](#)]
8. Zhang, J.; Zhao, Y.; Shone, F.; Li, Z.; Frangi, A.F.; Xie, S.Q.; Zhang, Z.-Q. Physics-Informed Deep Learning for Musculoskeletal Modeling: Predicting Muscle Forces and Joint Kinematics from Surface EMG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2023**, *31*, 484–493. [[CrossRef](#)]
9. Unanyan, N.N.; Belov, A.A. Design of Upper Limb Prosthesis Using Real-Time Motion Detection Method Based on EMG Signal Processing. *Biomed. Signal Process. Control* **2021**, *70*, 103062. [[CrossRef](#)]
10. Xu, Z.; Todorov, E. Design of a Highly Biomimetic Anthropomorphic Robotic Hand towards Artificial Limb Regeneration. *IEEE Int. Conf. Robot. Autom.* **2016**, 2016, 3485–3492.
11. Mohammadi, A.; Lavranos, J.; Zhou, H.; Mutlu, R.; Alici, G.; Tan, Y.; Choong, P.; Oetomo, D. A Practical 3D-Printed Soft Robotic Prosthetic Hand with Multi-Articulating Capabilities. *PLoS ONE* **2020**, *15*, e0232766. [[CrossRef](#)]
12. Chung, E.A.; Benalcázar, M.E. Real-Time Hand Gesture Recognition Model Using Deep Learning Techniques and EMG Signals. In Proceedings of the 2019 27th European Signal Processing Conference (EUSIPCO), A Coruna, Spain, 2–6 September 2019. [[CrossRef](#)]
13. Wang, J.; Tang, L.; Bronlund, J.E. Pattern Recognition-Based Real Time Myoelectric System for Robotic Hand Control. In Proceedings of the 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), Xi'an, China, 19–21 June 2019; pp. 1598–1605. [[CrossRef](#)]
14. Yang, D.; Zhao, J.; Gu, Y.; Wang, X.; Li, N.; Jiang, L.; Liu, H.; Huang, H.; Zhao, D. An Anthropomorphic Robot Hand Developed Based on Underactuated Mechanism and Controlled by EMG Signals. *J. Bionic. Eng.* **2009**, *6*, 255–263. [[CrossRef](#)]
15. Nunez-Montoya, B.; Valarezo Anazco, M.; Saravia-Avila, A.; Loayza, F.R.; Valarezo Anazco, E.; Teran, E. Supervised Machine Learning Applied to Non-Invasive EMG Signal Classification for an Anthropomorphic Robotic Hand. In Proceedings of the 2022 IEEE ANDESCON: Technology and Innovation for Andean Industry, ANDESCON 2022, Barranquilla, Colombia, 16–19 November 2022.
16. Betthausen, J.L.; Krall, J.T.; Bannowsky, S.G.; Lévy, G.; Kaliki, R.R.; Fifer, M.S.; Thakor, N.V. Stable Responsive EMG Sequence Prediction and Adaptive Reinforcement with Temporal Convolutional Networks. *IEEE Trans. Biomed. Eng.* **2020**, *67*, 1707–1717. [[CrossRef](#)]
17. Stashuk, D. EMG Signal Decomposition: How Can It Be Accomplished and Used? *J. Electromyogr. Kinesiol.* **2001**, *11*, 151–173. [[CrossRef](#)]
18. Phinyomark, A.; Khushaba, R.N.; Scheme, E. Feature Extraction and Selection for Myoelectric Control Based on Wearable EMG Sensors. *Sensors* **2018**, *18*, 1615. [[CrossRef](#)] [[PubMed](#)]
19. Chen, X.; Zhang, X.; Zhao, Z.Y.; Yang, J.H.; Lantz, V.; Wang, K.Q. Multiple Hand Gesture Recognition Based on Surface EMG Signal. In Proceedings of the 2007 1st International Conference on Bioinformatics and Biomedical Engineering, Wuhan, China, 6–8 July 2007; pp. 506–509. [[CrossRef](#)]
20. Xin, W.; Liu, R.; Liu, Y.; Chen, Y.; Yu, W.; Miao, Q. Transformer for Skeleton-Based Action Recognition: A Review of Recent Advances. *Neurocomputing* **2023**, *537*, 164–186. [[CrossRef](#)]
21. Putri, F.T.; Caesarendra, W.; Królczyk, G.; Glowacz, A.; Prawibowo, H.; Ismail, R.; Indrawati, R.T. Human Walking Gait Classification Utilizing an Artificial Neural Network for the Ergonomics Study of Lower Limb Prosthetics. *Prosthesis* **2023**, *5*, 647–665. [[CrossRef](#)]
22. Riaz, Z.; Khan, B.; Abdullah, S.; Khan, S.; Islam, M.S. Lung Tumor Image Segmentation from Computer Tomography Images Using MobileNetV2 and Transfer Learning. *Bioengineering* **2023**, *10*, 981. [[CrossRef](#)]
23. McGeown, J.P.; Pedersen, M.; Hume, P.A.; Theadom, A.; Kara, S.; Russell, B. A Novel Method to Assist Clinical Management of Mild Traumatic Brain Injury by Classifying Patient Subgroups Using Wearable Sensors and Exertion Testing: A Pilot Study. *Biomechanics* **2023**, *3*, 20. [[CrossRef](#)]

24. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.; et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* **2016**, *316*, 2402–2410. [[CrossRef](#)]
25. Liu, Y.; Huang, W.; Jiang, S.; Zhao, B.; Wang, S.; Wang, S.; Zhang, Y. TransTM: A Device-Free Method Based on Time-Streaming Multiscale Transformer for Human Activity Recognition. *Def. Technol.* **2023**. [[CrossRef](#)]
26. Li, B.; Cui, W.; Wang, W.; Zhang, L.; Chen, Z.; Wu, M. Two-Stream Convolution Augmented Transformer for Human Activity Recognition. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 286–293. [[CrossRef](#)]
27. Che, C.; Zhang, P.; Zhu, M.; Qu, Y.; Jin, B. Constrained Transformer Network for ECG Signal Processing and Arrhythmia Classification. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 184. [[CrossRef](#)] [[PubMed](#)]
28. Xiao, S.; Wang, S.; Huang, Z.; Wang, Y.; Jiang, H. Two-Stream Transformer Network for Sensor-Based Human Activity Recognition. *Neurocomputing* **2022**, *512*, 253–268. [[CrossRef](#)]
29. Dong, Y.; Zhang, M.; Qiu, L.; Wang, L.; Yu, Y. An Arrhythmia Classification Model Based on Vision Transformer with Deformable Attention. *Micromachines* **2023**, *14*, 1155. [[CrossRef](#)] [[PubMed](#)]
30. Shukla, N.; Pandey, A.; Shukla, A.P.; Neupane, S.C. ECG-ViT: A Transformer-Based ECG Classifier for Energy-Constraint Wearable Devices. *J. Sens.* **2022**, *2022*, 2449956. [[CrossRef](#)]
31. Hu, R.; Chen, J.; Zhou, L. A Transformer-Based Deep Neural Network for Arrhythmia Detection Using Continuous ECG Signals. *Comput. Biol. Med.* **2022**, *144*, 105325. [[CrossRef](#)]
32. de Godoy, R.; Lahr, G.; Dwivedi, A.; Reis, T.; Polegato, P.; Becker, M.; Caurin, G.; Liarokapis, M. Electromyography-Based, Robust Hand Motion Classification Employing Temporal Multi-Channel Vision Transformers. *IEEE Robot. Autom. Lett.* **2022**, *7*, 10200–10207. [[CrossRef](#)]
33. Montazerin, M.; Zabihi, S.; Rahimian, E.; Mohammadi, A.; Naderkhani, F. ViT-HGR: Vision Transformer-Based Hand Gesture Recognition from High Density Surface EMG Signals. *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS* **2022**, *2022*, 5115–5119. [[CrossRef](#)]
34. The BioDigital Human Platform Interactive 3D Anatomy—Upper Limb. Available online: https://human.biodigital.com/view?id=production/maleAdult/male_region_upper_limb_18&lang=en (accessed on 5 November 2023).
35. Van Beek, N.; Stegeman, D.F.; Jonkers, I.; de Korte, C.L.; Veeger, D.; Maas, H. Single Finger Movements in the Aging Hand: Changes in Finger Independence, Muscle Activation Patterns and Tendon Displacement in Older Adults. *Exp. Brain Res.* **2019**, *237*, 1141–1154. [[CrossRef](#)] [[PubMed](#)]
36. Mehta, S.; Rastegari, M. Mobilevit: Light-Weight, General-Purpose, and Mobile-Friendly Vision Transformer. *arXiv* **2021**, arXiv:2110.02178. [[CrossRef](#)]
37. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. In Proceedings of the ICLR 2021—9th International Conference on Learning Representations, Virtual Event, Austria, 3–7 May 2021. [[CrossRef](#)]
38. Nalmpantis, C.; Vrakas, D. Signal2Vec: Time Series Embedding Representation. In *Engineering Applications of Neural Networks. EANN 2019. Communications in Computer and Information Science*; Springer: Cham, Switzerland, 2019; Volume 1000, pp. 80–90. ISBN 978-3-030-20257-6. [[CrossRef](#)]
39. Uribarri, G.; Mindlin, G.B. Dynamical Time Series Embeddings in Recurrent Neural Networks. *Chaos Solitons Fractals* **2022**, *154*, 111612. [[CrossRef](#)]
40. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2023**, arXiv:1706.03762.
41. Hendrycks, D.; Gimpel, K. Gaussian Error Linear Units (GELUs). *arXiv* **2023**, arXiv:1606.08415.
42. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the NAACL HLT 2019—2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies—Proceedings of the Conference, Minneapolis, MN, USA, 2–7 June 2018; Volume 1, pp. 4171–4186. [[CrossRef](#)]
43. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving Language Understanding by Generative Pre-Training. Technical Report, OpenAI. 2018. Available online: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf (accessed on 5 November 2023).
44. Lewis, M.; Liu, Y.; Goyal, N.; Ghazvininejad, M.; Mohamed, A.; Levy, O.; Stoyanov, V.; Zettlemoyer, L. BART: Denoising Sequence-to-Sequence Pre-Training for Natural Language Generation, Translation, and Comprehension. *Proc. Annu. Meet. Assoc. Comput. Linguist.* **2019**, *58*, 7871–7880. [[CrossRef](#)]
45. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *J. Mach. Learn. Res.* **2019**, *21*, 5485–5551.
46. Hu, Y.; Ye, K.; Kim, H.; Lu, N. BERT-PIN: A BERT-Based Framework for Recovering Missing Data Segments in Time-Series Load Profiles. *arXiv* **2023**, arXiv:2310.17742.

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.