Sensor Applications

Hierarchical Recurrent-Inception Residual Transformer (HRIRT) for Multi-Dimensional Hand Force Estimation using Force Myography Sensor

Muhammad Hamza Zafar[1*], Syed Kumayl Raza Moosavi[1], and Filippo Sanfilippo[1,2**]

[1] *Department of Engineering Sciences, University of Agder, 4879, Grimstad, Norway*
[2] *Department of Software Engineering, Kaunas University of Technology, 44029, Lithuania*
[*] *Graduate Student Member, IEEE*
[**] *Senior Member, IEEE*

Abstract—In this study, we present the Hierarchical Recurrent-Inception Residual Transformer (HRIRT), an innovative deep neural network architecture designed for accurate hand force estimation in Human-Robot Collaboration (HRC). The HRIRT combines recurrent layers, inception modules, residual connections, transformers, and Time2Vec feature engineering within a hierarchical framework to adeptly capture the complex spatiotemporal dynamics of hand force data. Our evaluation spans three dimensions of HRC—1D, 2D, and 3D hand force estimation—leveraging data from force myography sensors to train and test the model's performance. The HRIRT demonstrates exceptional accuracy and robustness, setting new benchmarks in hand force estimation across varied interaction scenarios. Specifically, the 1D interactions focus on linear force applications, while 2D and 3D interactions involve more complex spatial movements, showcasing the model's capability to generalize across different force interaction contexts. in 1D scenarios with the Kuka robot, HRIRT achieved a 93.76% R2 score, significantly outperforming TL-CDG and SCNN models. Similarly, in 2D and 3D force estimations, HRIRT demonstrated exceptional accuracy, with R2 scores of 94.25% and 91.61%, respectively, and maintained low error rates across RMSE, NMSE, and MAE metrics. These results underscore the HRIRT's state-of-the-art performance and interpretability, highlighting its potential as a powerful tool for real-time precise hand force estimation in diverse HRC applications.

Index Terms—Hand Force Estimation, Human-Robot Collaboration (HRC), Force Myography Sensors, Real-Time Estimation, Model Interpretability.

I. INTRODUCTION

In the context of collaborative industrial tasks, where human workers engage in activities like object handover or transportation, the utilisation of hand forces for interacting with machines is prevalent. Recent research has examined methods for estimating interactive forces through biological signals. One approach is surface electromyography (sEMG) which detects electrical activity in muscles. A newer technique is force myography (FMG), a noninvasive method using wearable devices that incorporates force sensing resistors to measure changes in resistance caused by pressure [1]. Specifically, an FMG band wrapped around the arm can estimate grasping forces or isometric contractions by detecting muscle activity [2]. Some studies have applied FMG signals to recognize intentional or random hand motions in human-robot collaborative tasks with manipulator robots. By detecting muscle contractions, FMG bands show promise for controlling robotic systems during physical human-robot interactions. Further research is still needed to refine FMG and determine optimal application for human-robot collaboration [3]. In these instances, human intention is identified through the implementation of a recurrent neural network (RNN), trained on intra-session data to enable the robot to avoid collisions during unintentional movements. Recent research has delved into classifying grasped objects based on intra-session FMG data during shared HRC tasks, focusing on recognising the intended tool (object) to be used by human workers [4]. Subsequent studies extended this exploration to inter-session grasping FMG data, demonstrating applicability to new users without prior training for HRC tasks [5].

Corresponding author: F. Sanfilippo (e-mail: filippo.sanfilippo@uia.no).

In the field of sEMG-based human-robot interaction, recent work has explored deep learning methods such as convolutional neural networks (CNNs) and long short-term memory networks (LSTMs) to estimate dynamic or static forces from sEMG signals [6]. Similarly, FMG biosignals have been utilised for applied force estimations in physical human-robot interaction (pHRI) activities [7]. Notably, studies investigating pHRI between participants and fixed linear robots employed intra-session data to train task-specific machine learning models, achieving real-time predictions of interactive forces [8]. Additionally, research involving inter-session FMG data in a planar workspace demonstrated improved force estimations during 2D pHRI studies. However, there is a notable gap in the literature concerning the application of FMG-based estimated forces in the control loop of human-robot collaborative tasks conducted in 3D [9]. This study addresses this gap by exploring FMG-based pHRI with a 7-degree-of-freedom Kuka robot, estimating grasping forces during dynamic motion. The study involves using a cylindrical gripper as the end-effector (EEF) for hand grasps, interacting in various directions within 1D, 2D, and 3D workspaces. The proposed Hierarchical Recurrent-Inception Residual Transformer (HRIRT) model in this study demonstrated moderate estimations of grasping forces during dynamic interactions in the 3D-HRC task.

The contributions of this work are highlighted below:

- The development of the HRIRT model introduces a groundbreaking neural network architecture by blending recurrent layers, inception modules, residual connections, and transformers with Time2Vec feature engineering, significantly advancing force estimation techniques in HRC.
- HRIRT sets new benchmarks in 1D, 2D, and 3D hand force estimation, achieving unparalleled accuracy and robustness

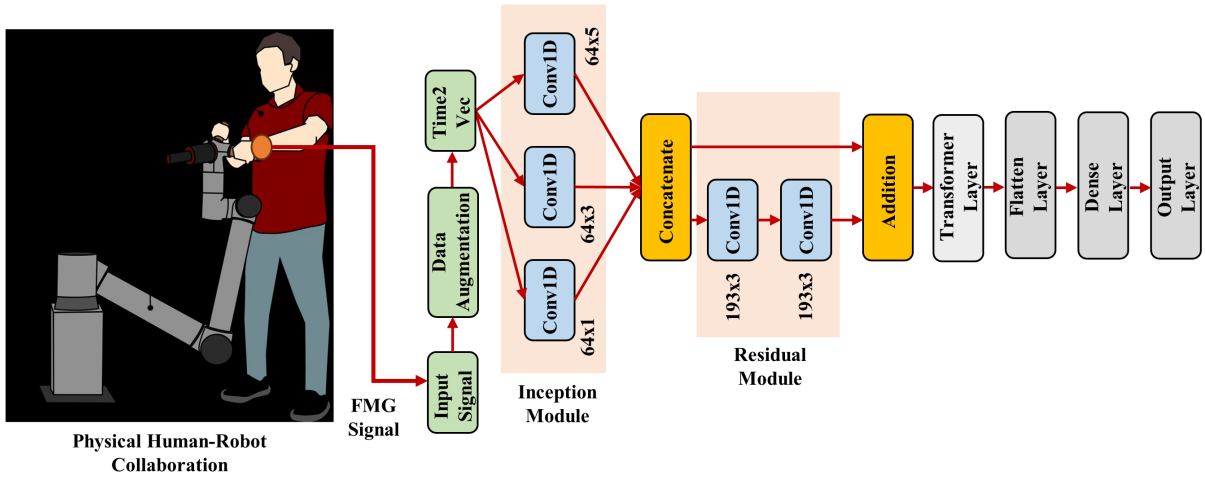

Fig. 1. Layer wise detailed architecture of the proposed Time2Vec based HRIRT Deep Neural Network Model.

with R2 scores of 93.76%, 94.25%, and 91.61% respectively, demonstrating its superior performance.

- Demonstrating remarkable generalization capabilities, HRIRT effectively handles varied HRC scenarios across different dimensions, showcasing its adaptability and potential in real-world applications.
- Through an in-depth interpretability analysis, HRIRT reveals the synergistic effects of its composite elements, particularly highlighting how Time2Vec feature engineering enriches model performance in complex HRC tasks.

II. DATA PRE-PROCESSING

The dataset used in this study includes FMG data obtained from pHRC settings in which a subject engaged with a manipulator: a serial manipulator (7-DoF Kuka robot). The FMG sensor is placed on forearm position when interacting with the Kuka robot [10]. A 6-axis force-torque (FT) sensor provided a matching actual force reading (N) for every row of FMG data. More detail of the dataset acqusition and sensors is available in supplementary file.

A. Feature Engineering

1) Data Augmentation: The mathematical model for augmenting the 32-channel FMG input feature space with average and variance as additional features:

Let the original 32-channel FMG feature space be:

$$X = [x_1, x_2, \ldots, x_{16}], \tag{1}$$

where x_i is the FMG signal from channel i.

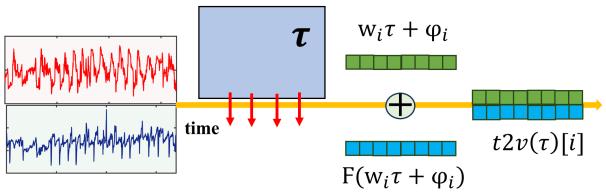

Fig. 2. Structure of Time2Vec Feature Engineering.

We construct a new feature space X^* as:

$$X^* = [X, \mu(X), \sigma(X)], \tag{2}$$

where $\mu(X)$ is a feature vector containing the mean/average signal value of each FMG channel and $\sigma(X)$ is a feature vector containing the variance of each FMG channel. So the final augmented feature space X^* concatenates the original 32-channel FMG data with average and variance feature.

2) Time2Vec: Utilising the Time2Vec method on the hand force dataset is crucial for achieving precise hand force estimation. This method proves instrumental in comprehending the temporal patterns and interdependencies inherent in the hand force data. The dataset under examination encompasses various parameters, including muscle activity signals, hand kinematics, grasp type, object weight, and associated timestamps. By converting these timestamps into a suitable numerical representation, the temporal variable becomes an integrated, learned component of the model. This incorporation enables the model to grasp the sequential nature of the data and effectively capture prolonged dependencies.

An advantageous feature of employing the Time2Vec method in hand force estimation lies in its capability to handle periodic patterns. Hand force data exhibits repetitive patterns due to consistent variations in grasp type, object weight, and muscle fatigue over time. The Time2Vec method excels in encoding these periodic patterns, facilitating the model in learning the intricate relationships between time and hand force output at different temporal scales. Consequently, this leads to a substantial enhancement in estimation accuracy. The mathematical representation for Time2Vec is expressed as follows:

$$t2v(\tau)_i = \begin{cases} wi\tau + \phi_i, & \text{if } i = 0 \\ F(wi\tau + \phi_i), & \text{if } 1 \le i \le k \end{cases} \tag{3}$$

Here, F represents a periodic function, w_i and ϕ_i are variables subject to learning, $i = 0$ denotes the non-periodic period, and $1 \le i \le k$ pertains to the periodic period.

III. PROPOSED TECHNIQUE

The Hierarchical Recurrent-Inception Residual Transformer (HRIRT) model is meticulously designed for hand force estimation, leveraging Force Myography (FMG) sensor data. This comprehensive

model architecture intertwines various computational paradigms to process one-dimensional time-series data effectively.

A. Input Layer

At the heart of HRIRT's design is its hierarchical processing capability, which allows for the efficient handling of data across multiple scales. This is crucial for FMG sensor data, where the relevance of information can vary greatly across different temporal and spatial dimensions. The model's architecture is specifically tailored to gradually refine the sensor data through its layers, enhancing the model's ability to discern subtle patterns and relationships within the data. The HRIRT model starts with an input layer defined as follows:

$$X_{\text{in}} \in \mathbb{R}^{T \times C} \tag{4}$$

where T is the temporal sequence length, and C represents the number of sensor channels.

B. Inception Module

The inception module, with its parallel convolutional pathways, enables the model to capture a broad range of features from the input data. By utilizing different kernel sizes, HRIRT can simultaneously focus on fine-grained details and broader contextual information. This multipath approach ensures that the model is not constrained by a single field of view, enhancing its adaptability and sensitivity to diverse feature types. It is mathematically represented as:

$$Y_{\text{inception}} = \text{Concat}\left(\text{Conv}_{1\times1}(X_{\text{in}}), \text{Conv}_{3\times3}(X_{\text{in}}), \text{Conv}_{5\times5}(X_{\text{in}}), \text{MaxPool}(X_{\text{in}})\right) \tag{5}$$

This equation highlights the process of applying convolutional operations with distinct kernel sizes and a max-pooling operation on the input, followed by concatenation of the results.

C. Residual Block

The incorporation of residual blocks is a strategic choice to enhance the model's depth without compromising its stability. By introducing shortcut connections that bypass one or more layers, the model can mitigate the vanishing gradient problem, a common challenge in training deep neural networks. These residual connections facilitate the flow of gradients during backpropagation, allowing for deeper models without the risk of performance degradation.

$$Y_{\text{res}} = \text{ReLU}\left(X_{\text{in}} + \text{Conv}_{3\times3}\left(\text{ReLU}\left(\text{Conv}_{3\times3}(X_{\text{in}})\right)\right)\right) \tag{6}$$

This structure allows the model to learn residual features effectively, addressing the vanishing gradient problem.

D. Transformer Layer

The transformer layer represents a significant advancement in the model's ability to understand the temporal dynamics of FMG sensor data. Through self-attention mechanisms, HRIRT can weigh the importance of different parts of the input sequence, capturing long-range dependencies that are critical for accurate force estimation. This capability is especially important in scenarios where the relationship between sensor readings and hand force is complex and influenced by factors not immediately adjacent in the temporal sequence. This layer is designed as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{7}$$

$$Y_{\text{transformer}} = \text{LayerNorm}\left(X_{\text{in}} + \text{Attention}(X_{\text{in}}, X_{\text{in}}, X_{\text{in}})\right) \tag{8}$$

$$+ \text{LayerNorm}\left(X_{\text{in}} + \text{FFN}(Y_{\text{transformer}})\right) \tag{9}$$

where d_k represents the dimensionality of the key vectors. This layer emphasizes capturing long-range dependencies through self-attention.

E. Output Processing

The final stages involve flattening and dense layers, formulated for the output preparation:

$$Y_{\text{out}} = \text{ReLU}\left(\text{Dense}\left(\text{Flatten}(Y_{\text{transformer}})\right)\right) \tag{10}$$

This equation transitions the model's multidimensional output into a suitable format for prediction tasks.

IV. RESULTS AND DISCUSSION

In this section, we present a detailed comparative analysis of three prominent models—Hierarchical Recurrent Inception Residual Transformer (HRIRT), Transfer Learning with Cross-Domain Generalisation (TL-CDG) [10], and Stacked Convolutional Neural Network (SCNN)—in the context of multi-dimensional hand force estimation. The models have been evaluated across three distinct dimensions (1D, 2D, and 3D), and their performance has been

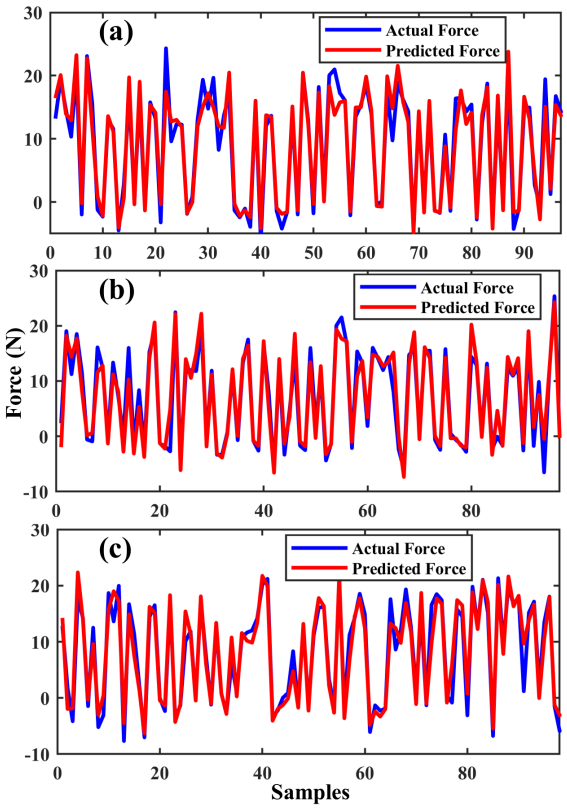

Fig. 3. Comparison of Actual vs Predicted Force by the proposed model for pHRC on Kuka robot in (a) 1D (b) 2D, and (c) 3D.

TABLE 1. Comparative analysis of different models for multi dimensional hand force estimation

Tech.	Data	R2	RMSE	NMSE	MAE
HRIRT	pHRC with Kuka (1D)	93.76%	2.206	0.064	1.512
TL-CDG		81.11%	8.901	0.161	2.334
SCNN		76.04%	9.862	0.332	3.065
HRIRT	pHRC with Kuka (2D)	94.25%	1.896	0.055	1.197
TL-CDG		80.56%	7.469	0.682	2.835
SCNN		76.63%	8.381	0.563	1.891
HRIRT	pHRC with Kuka (3D)	91.61%	2.697	0.071	1.868
TL-CDG		78.72%	9.288	0.314	2.677
SCNN		74.97%	10.396	0.854	3.755

rigorously assessed using key metrics such as the coefficient of determination (R2), root mean square error (RMSE), normalised mean squared error (NMSE), and mean absolute error (MAE). This comprehensive examination aims to shed light on the strengths and limitations of each model, providing valuable insights into their efficacy in capturing the complexities inherent in hand force prediction across different dimensions. The percentage comparisons and in-depth analyses presented here contribute to a nuanced understanding of the models' performance, facilitating informed decision-making in the selection of an optimal approach for multi-dimensional hand force estimation applications.

A. Comparative Analysis

The comparative analysis in Table 1 reveals insightful performance metrics for the three models—Hierarchical Recurrent Inception Residual Transformer (HRIRT), Transfer Learning with Cross-Domain Generalisation (TL-CDG), and Stacked Convolutional Neural Network (SCNN)—across different dimensions of hand force estimation. Notably, HRIRT consistently outperforms the other models in terms of the coefficient of determination (R2), demonstrating its superior ability to explain the variance in the data. In the 1D, 2D, and 3D dimensions, HRIRT achieves R2 values of 93.76%, 94.25%, and 91.61%, respectively, showcasing its robust predictive capabilities. In contrast, TL-CDG and SCNN exhibit lower R2 values across all dimensions, with TL-CDG reaching a maximum of 81.11% in 1D, 80.56% in 2D, and 78.72% in 3D. SCNN, while competitive, lags slightly behind with maximum R2 values of 76.04%, 76.63%, and 74.97% in the respective dimensions. This indicates that HRIRT excels in capturing the intricate patterns in hand force data, leading to more accurate predictions compared to its counterparts.

A detailed examination of the root mean square error (RMSE) further emphasises the HRIRT's superiority in minimising prediction errors. In all dimensions, HRIRT consistently achieves lower RMSE values (2.206, 1.896, and 2.697) compared to TL-CDG (8.901, 7.469, and 9.288) and SCNN (9.862, 8.381, and 10.396). The percentage comparison underscores the HRIRT's remarkable performance, showcasing a substantial reduction in prediction errors across diverse dimensions. These findings collectively suggest that the HRIRT not only excels in capturing the underlying patterns in the data but also demonstrates superior accuracy in predicting hand force values, positioning it as a formidable model for multi-dimensional hand force estimation tasks.

V. CONCLUSION

The paper introduces HRIRT, a novel deep neural network designed for nuanced hand force estimation in Human-Robot Collaboration (HRC). With a hierarchical structure incorporating recurrent layers, inception modules, residual connections, transformers, and Time2Vec feature engineering, HRIRT demonstrates versatility across 1D, 2D, and 3D force estimation dimensions, outperforming existing models on a force myography sensor dataset. Achieving state-of-the-art accuracy with R2 values exceeding 93%, HRIRT's generalization and interpretability highlight its efficacy in diverse force interaction contexts. Overall, HRIRT emerges as a potent tool for real-time and precise hand force estimation, advancing HRC applications in robotics, prosthetics, and human-machine interfaces.

ACKNOWLEDGEMENT

This work is supported by Artificial Intelligence, Biomechatronics and Collaborative Robotics Group at Top Research Center Mechatronics (TRCM), University of Agder, Grimstad, Norway.

REFERENCES

[1] A. Prakash, N. Sharma, and S. Sharma, "Novel force myography sensor to measure muscle contractions for controlling hand prostheses," *Instrumentation Science & Technology*, vol. 48, no. 1, pp. 43–62, 2020.

[2] X. Jiang, L.-K. Merhi, and C. Menon, "Force exertion affects grasp classification using force myography," *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 2, pp. 219–226, 2017.

[3] H. Zhou, C. Tawk, and G. Alici, "A multipurpose human–machine interface via 3d-printed pressure-based force myography," *IEEE Transactions on Industrial Informatics*, 2024.

[4] N. D. Kahanowich and A. Sintov, "Robust classification of grasped objects in intuitive human-robot collaboration using a wearable force-myography device," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1192–1199, 2021.

[5] E. Bamani, N. D. Kahanowich, I. Ben-David, and A. Sintov, "Robust multi-user in-hand object recognition in human-robot collaboration using a wearable force-myography device," *IEEE Robotics and Automation Letters*, vol. 7, no. 1, pp. 104–111, 2021.

[6] T. Zhang, H. Chu, Y. Zou, and H. Sun, "A robust electromyography signals-based interaction interface for human-robot collaboration in 3d operation scenarios," *Expert Systems with Applications*, vol. 238, p. 122 003, 2024.

[7] Z. Chen, H. Wang, H. Chen, and T. Wei, "Continuous motion finger joint angle estimation utilizing hybrid semg-fmg modality driven transformer-based deep learning model," *Biomedical Signal Processing and Control*, vol. 85, p. 105 030, 2023.

[8] U. Zakia and C. Menon, "Estimating exerted hand force via force myography to interact with a biaxial stage in real-time by learning human intentions: A preliminary investigation," *Sensors*, vol. 20, no. 7, p. 2104, 2020.

[9] U. Zakia and C. Menon, "Toward long-term fmg model-based estimation of applied hand force in dynamic motion during human–robot interactions," *IEEE Transactions on Human-Machine Systems*, vol. 51, no. 4, pp. 310–323, 2021.

[10] U. Zakia and C. Menon, "Human–robot collaboration in 3d via force myography based interactive force estimations using cross-domain generalization," *IEEE Access*, vol. 10, pp. 35 835–35 845, 2022.