A Novel Graph Isomorphism Network For Hand Gesture Recognition With Leap Motion Controller

Rahma AMRI*, Nahla MAJDOUB BHIRI†, Hajer CHTIOUI ‡, Bassem SEDDIK §, Anouar BEN KHALIFA¶‖
*†‡§¶Université de Sousse, Ecole Nationale d'Ingénieurs de Sousse,
LATIS-Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisie;
‖Université de Jendouba, Institut National des Technologies et des Sciences du Kef, 7100, Le Kef, Tunisie;
§ ViZmerald SARL, Research and Development Department, Novation City, Sousse, Tunisia.
Email: * rahma.amri@eniso.u-sousse.tn, ¶‖anouar.benkhalifa@eniso.u-sousse.tn,

***Abstract*—The 3D hand skeletal data has received considerable amount of attention due to its potential uses case for Hand Gesture Recognition (HGR) systems. Additionaly, Graph Neural Network (GNN) have been widely employed for skeletal-based HGR. Yet still, traditional models frequently suffer from inefficient feature representation and generalizability. Thus, to overcome these limitations, we introduce Spatial Graph Isomorphism Neural Networks (S-GINs), which use GIN layers to improve the feature aggregation. First, we create a graph representing skeletal-based recordings. Following that, a spatial network convolution module learns the inherent topology of hand gestures from neighbor nodes and updates it using a multilayer perceptron. This method promotes classification accuracy over other spatial based graph models including graph attention network, graph sampeling and aggregating, and graph convolutional networks. We validate S-GINs employing three benchmark datasets: MMHGD, 2MLMD, and Multi-view Leap2 , both horizontal and vertical sub-datasets. Experimental findings show that our model outperforms state-of-the-art graph based methods, by elevating the accuracies to 72%, 81%, 85%, and 83% respectively .**

***Index Terms*—Hand gesture recognition, skeletal data, graph isomorphism networks, spatial features.**

I. Introduction

Hand Gesture Recognition (HGR) has become a prominent research topic in computer vision due to its vast range of applications over various domains [13]. It improves human-computer interaction by offering intuitive and touchless control mechanisms. Furthermore, it is widely used in smart home automation [2], allowing users to interact with connected equipment using natural hand gestures. In the gaming industry [16], HGR promotes player engagement by creating immersive and engaging experiences. Moreover, it enables seamless communication between pedestrians and human drivers, which improves driver safety [4]. HGR systems traditionally utilize various data modalities, including RGB video streams, depth sensing sequences and skeletal tarcking [19]. Skeletal data offers distinct advantages comapred to the other modalities by recording essential structure and movement data while being naturally resistant to common noise artifacts, background or obstructions [13]. Unlike RGB and depth-based methods, which can be influenced by lighting conditions, complicated backgrounds, or partial hand visibility, skeletal representations focus primarily on important joint movements, making them a more reliable and noise-resistant mode of comprehending hand gestures. Advancements in sensing technologies have enabled precise 3D skeletal hand data acquisition, a critical foundation for robust HGR systems. Devices like the Leap Motion Controller (LMC) [18] and Microsoft Kinect [17] excel in real-time skeletal tracking by mapping hand joint coordinates in three-dimensional space. To optimize these raw spatial data for analytical tasks, computational frameworks such as OpenPose [15], DetNet [19] and MediaPipe [12] have emerged as essential tools. These techniques refine hand pose estimation through high-accuracy joint localization and structured data organization, facilitating further applications such as graph-based modeling and gesture classification.

Traditional deep learning models, such as Convolutional Neural Networks (CNNs) [13] and Recurrent Neural Networks [2], struggle to effectively model the complex joint dependencies in skeletal hand data because they primarily operate on Euclidean grid-based structures. However, Graph Neural Networks (GNNs) [11] provide a more suitable framework by representing hand skeletal as structured graphs, where joints are treated as nodes and their connections as edges. Several GNN-based approaches have been developed to encode skeletal hand sequences into spatial graphs, where joints are connected both within individual frames (spatial dependencies). However, existing methods have notable limitations. Graph Convolutional Networks (GCNs) [24] have limited expressive power and struggle to distinguish complex graph structures. Furthermore, Graph Attention Networks (GATs) models [3] with their static attention mechanism restricts adaptability, and their computational complexity makes them less practical for large-scale datasets. Moreover, GraphSAGE [10], which generates node embeddings through neighborhood sampling and aggregation. While it scales well, its performance depends heavily on the chosen aggregation function, which may lead to information loss in capturing hand dynamics. Accordingly, we notice that traditional GNNs often struggle to capture nuanced relational patterns due to their limited discriminative power, which is intrinsically bounded by the expressiveness of the Weisfeiler-Lehman (WL) graph isomorphism test [23]. In contrast, Graph Isomorphism Networks (GINs) are theoretically proven to match the WL test discriminative capacity, enabling them to learn injective aggregation functions and uniquely characterize graph substructures. For that reason, in this work we present the Spatial-Graph Isomorphism Network (S-GIN), a graph-based architecture designed to enhance structural expressiveness and spatial reasoning in hand gesture analysis. To demonstrate the effectiveness of our approach, we conduct extensive experiments on three large-scale datasets: 2MLMD [2], Multi-ViewLeap2 [7], and MultiModalHandGesture [14]. Our model consistently achieves state-of-the-art performance beyond all datasets, highlighting its robustness and generalizability. Thus, the main contributions of this paper are summarized as follows:

- We conducted a study of spatial graph-based methods for HGR, with a particular focus on their limitations.
- We proposed Spatial-Graph Isomorphism Networks (S-GINs) a novel graph-based architecture specifically designed to model topological dependencies and relational structures within skeletal hand data.
- We made an extensive evaluation of our model through three benchmark datasets to test our model generalizability and robustness.

The remainder of this paper is structured as follows. Section II provides a concise study of spatial-graph-based models for spatial feature extraction in skeletal gesture analysis. Section III introduces our S-GINs, detailing its architectural innovations, including spatial-geometric encoding. In Section IV, we validate S-GINs through extensive experiments on three benchmark datasets, comparative analysis with state-of-the-art methods, and a discussion of its robustness. Finally, Section V concludes this work.

II. Related work of HGR systems graph based methods

Skeletal-based HGR has emerged as a critical area in computer vision, particularly for applications requiring fine-grained motion analysis. While early approaches relied on CNNs, their limitations in modeling non-Euclidean skeletal relationships motivated the adoption of GNNs. Below, we study key advancements and persistent challenges in spatial-graph-based HGR. For instance, in [8], the authors suggested a multi-attention GCN embedded attention with channel, time and graph spatial dimension that can identify hand and body gesturs simultaneously. The spatial GCN is used to identify static hand movements to get the motion parameters of control. While the hand gestures were static, the spatial GCN outperformed the spatiotemporal GCN and achieved a 98.14% of recognition rate on the Microsoft Kinect and leap motion datasets. Similarly, Aiman *et al.* [1] proposed an Angle Based Hand Gestures Graph Convolution Neural Network (AHG-GCN), in order to connect the wrist with the tips and bases of each finger. This model included two new edge types to the graph. This allowed for the explicit capture of their spatial connections, which is essential for distinguishing between different hand movements. Furthermore, the angle formed with fingertip/finger-base joints is used to provide new characteristics for each skeletal joint. For the DHG 14/28 dataset, AHG-GCN obtained an accuracy of 90% and 88% for 14 and 28 gestures respectively, and for the SHREC 2017 dataset, an accuracy of 94.05% and 89.4% was respectively achieved for 14 and 28 gesture configurations. Graph topology controls feature aggregation in GCNs, making it essential for obtaining representative features. Chen *et al.* [6] identified an innovative Channel-wise Topology Refinement Graph Convolution (CTR-GC) for skeletal-based action recognition. This model aggregates joint data in many channels and dynamically learns various topologies. CTR-GC has strong correlation modeling capabilities by learning channel-wise topologies in a refinement manner. Experimental findings and quantitative research show that CTR-GC could represent data more accurately than other graph convolutions. By integrating temporal modeling modules with CTR-GC, they created the potent graph convolutional network known as CTR-GCN. Regarding three datasets, the CTR-GCN achieved respectively an accuracy of 92.4%, 90.6% and 96.5% regarding NTU RGB+D, NTU-RGB+D 120 and Northwestern-UCLA datasets. In addition, Zhou *et al.* [26] introduced a novel Covariance-based Graph Convolutional Network (CovGCN). They used graphs to depict signals from sEMG sensor. CovGCN automatically learns the channel-wise topologies in a refinement approach rather than modeling the signals individually with a fixed topology. The GCN model can adaptively learn the dynamic topologies for different hand motions thanks to the covariance-based topology refinement module. The model showed better performance on large datasets, Ninapro DB2 dataset and the UOW Dataset.

Nevertheless, the GCNs shares the same weights for all neighborhoods of each nodes which poses a significant limitation in modeling complex and heterogeneous graph structures. To improve stability in fixed graph structures, recent works have integrated attention mechanisms. Guo *et al.* [9] suggested a dynamic static parallel network for dynamic body gestures, they added a graph attention module to improve the graph data fusion effect in the dynamic static network. The model was validated on the Hand-Gesture dataset and gained an accuracy of 98.84%. Moreover, Zhang *et al.* [25] proposed Human Skeletal Graph Attention Convolution Neural Network (HSGATC), which introduced graph attention mechanism to extract the spatial features of the human skeletal. This improved accuracy on NTU-RGB+D and Kinetics datasets but struggled with computational overhead. Furthermore, Xie *et al.* [21] advanced this by introducing a dimension-attention block and attention adjacency matrix, enhancing feature discriminability over skeletal dimensions. The model was tested on two datasets NTU-RGB+D and Kinetics and achieved better performance than the state-of-the-art works. Additionaly, Xing *et al.* [22] proposed a novel Hybrid Attention-Based Graph Neural Network (HA-GCN). The identification accuracy was clearly improved by a new adaptive spatial attention layer that

creates a new starting graph adjacency matrix connecting the head, hands, and feet and extends the local attention map to the global level based on relative distance and relative angle information. The HA-GCN was evaluated on two large datasets, NTU-RGB+D and Kinetic which achieved respec-

TABLE I

SUMMARY OF SPATIAL GRAPH-BASED METHODS USED IN LITERATURE FOR HAND GESTURE RECOGNITION (HGR)

Ref	Year	Modalities	Method	Dataset	Accuracy (%)
[25]	2020	Skeletal data	HSGAC	Kinectics	53.68
				NTU-RGB+D	90.30
[6]	2021	Skeletal data	CTR-GCN	NTU RGB+D	92.40
				NTU RGB+D 120	90.60
				Northwestern-UCLA	96.50
[22]	2022	Skeletal data	HA-GCN	Kinectics	61.60
				NTU-RGB+D	92.10
[5]	2022	YCrCg	GraphSAGE	Own dataset	91.70
[8]	2023	Skeletal data	Spatial GCN	Kinect	98.14
				Leap Motion	98.14
[1]	2023	Skeletal data	AHG-GCN	DHG 14/28	90.00 / 88.00
[26]	2025	sEMG data	CovGCN	Ninapro DB2	88.02
				UOW	84.69

tively an accuracy of 61.6% and 92.1%.

However, because of their computational cost and static attention mechanism, which limit their flexibility and make them less useful for large-scale datasets other graph-based method was applied.

In the same vien, Chen *et al.* [5] used another GNN model, the GraphSAGE, to recognize the graph-structured data of the gestures from images using YCrCg color space. Graph-SAGE was used to extract hand features and generated a homogenous graph model. The Adaboost algorithm was used to combine the two powerful classifiers of the random forest and the support vector machine into a cascade classifier through the cascade structure. The algorithm average detection accuracy in a test set was 91.70%.

To address the shortcomings of spatial-based GNNs that just employ local neighbor information and disregard global structural information, we propose the S-GINS technique that will be well developed in the next section.

III. PROPOSED S-GINS MODEL

This section describes the architecture and implementation of the suggested model S-GINs. As illustrated in Fig 2, a novel framework designed to address the challenges of skeletal data. Our methodology comprises two stages: Graph construction from skeletal informations and a GIN architecture optimized for fine-grained gesture discrimination. To our knowledge, this is the first work to adapt GINs for LMC-based skeletal modeling, addressing its inherent scarcity and high-dimensional joint kinematics.

A. Graph generation

Hand skeletal information is produced by processing and analyzing the data collected by LMC. To significantly improve dynamic gesture detection, users can place their hands in the sensor's acquisition space to record 3D hand information inside a certain coordinate system. The 3D coordinates of several hand-arm joints, including the MCP, PIP, DIP, tip, wrist, velocity, direction, and palm center, are displayed in this data. To represent the hand as a graph, we selected four joints per finger, each described by their (x, y, z) coordinates. Specifically, for each finger, the joints are defined as: MCP Position, PIP Position, DIP Position, and Tip Position. This representation results in a total of 20 nodes per hand, corresponding to the joints of five fingers. Moreover, the edges of the graph of the graph are defined by connecting adjacent joints within each finger, a total of 21 edges per hand was generated. Fig 1 shows the adjacency matrix for a sample graph, highlighting the connections between joints. The graph is represented by an adjacency matrix A where:

$$\begin{cases} A_{ij} = 1 \text{ if there is connection between joints } i \text{ and } j. \\ A_{ij} = 0 \text{ otherwise .} \end{cases}$$

The bones and joints of the hand skeletal presents the edges and nodes of a graph, which is an anatomical representation.

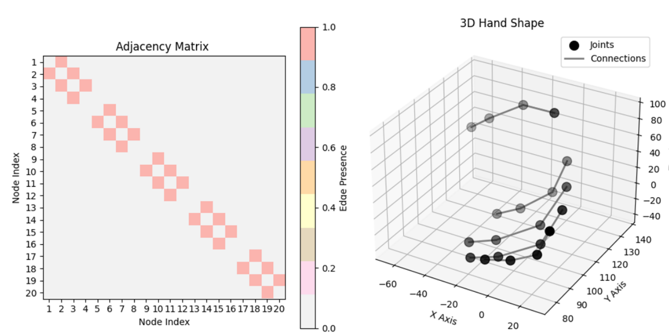

Fig. 1. 3D Hand graph visualization and its adjacency matrix

G= (N,E) Where E represents the set of edges and N represents the set of nodes. We summarize the main mathematical notations used in table II.

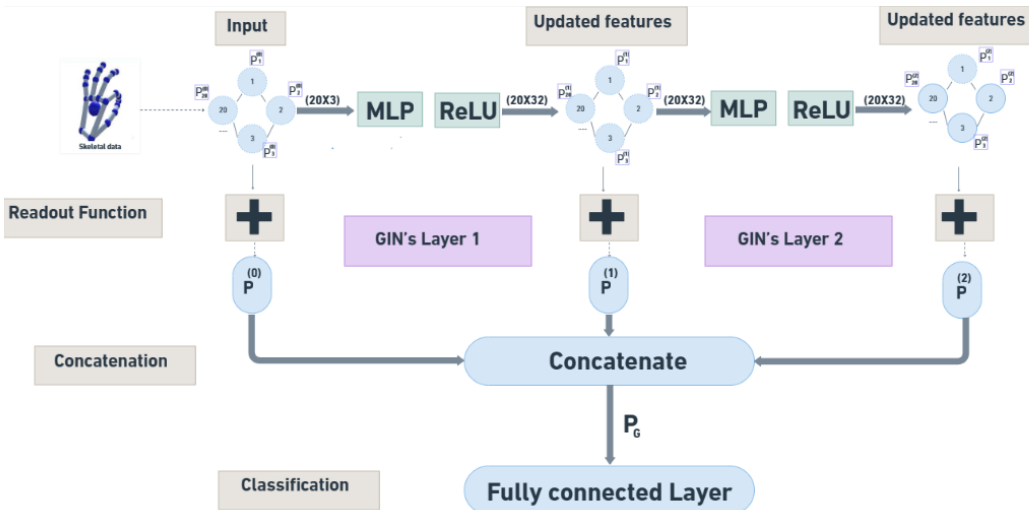

Fig. 2. Proposed S-GINs framework

TABLE II
SUMMARY OF MATHEMATICAL NOTATIONS IN GRAPH NEURAL NETWORKS

Notation	Definition
$p_v^{(k)}$	Feature vector of node v at layer k.
$a_v^{(k)}$	Aggregated feature vector at node v for layer k.
$N(v)$	Neighborhood set of node v.
p_G	Graph-level feature representation.
$\epsilon^{(k)}$	Learnable parameter in Graph Isomorphism Network.
$\mathrm{MLP}^{(k)}$	Multi-layer perceptron at layer k.
$\mathcal{C}^{(k)}$	Feature dimension at layer k.
$p_G^{(k)}$	Summed node features at layer k.
$\sum_{i=0}^{N} p_i^{(k)}$	Summation of node features at layer k.
$\text{concat}\left(\{p_G^{(k)}\} \mid k = 0, 1, \ldots, K\right)$	Concatenation of summed features from all layers.

B. GIN architecture

GIN was proposed by Xu *et al.* [23] as a special model of spatial based GNN. It has a strong expressive power, which refer to the ability to capter and represent the local patterns and saptial denpendencies in graph structed data. The network uses the total of the node characteristics to implement the aggregate and combine functions as detailed in Eq 1:

$$p_v^{(k)} = \mathrm{MLP}^{(k)} \left((1 + \epsilon^{(k)}) \cdot p_v^{(k-1)} + \sum_{u \in N(v)} p_u^{(k-1)} \right) \in \mathbb{R}^{C^{(k)}} \tag{1}$$

Where $p_v^{(k)}$ is the feature vector of node v at layer k.
Looking at the representation from the previous layer $p_v^{(k-1)}$ and then adding them, we notice the fact that adding a tiny variation to one, epsilon, which is a learnable parameter and updated during training, regulates the weight of the node's own characteristics during the update phase. In our case the dimension of the input vector was three which represent the 3D coordinates (x,y,z) of each node. This aggregation mechanism of the S-GINs can capture local and global structural information so improving the expressivness in gesture classification task. This representaion is fed into a non linear function Multi-Layer-Perceptron (MLP) which could be a deep neural network with at least two layers. The usage of MLP allows to capture more complex iteractions between node and its neighbors and avoid oversoomthing of spatial features. Considering that the dimension of the feature vector after updating will increase and the dimension is fixed during defining the model and its layers.
For the graph representation for $p_G^{(k)}$, we have a readout function for each one layer, which is in our case a sum function as show in Eq 2, and then we concatenated them and this will be a repsentation of the entire graph p_G as noted in Eq 3. The discriminative power of GIN matches the WL test, enabling it to extract complex spatial features that effectively differentiate graphs classification results.

For the classification we used a fully connected, we used a fully connected layer that map the features to desired output dimension.

$$p_G^{(k)} = \sum \left(p_0^{(k)}, p_1^{(k)}, \ldots, p_N^{(k)} \right) \tag{2}$$

$$p_G = \text{concatenate} \left(\{p_G^{(k)}\} \mid k = 0, 1, \ldots, K \right) \tag{3}$$

IV. EXPERIMENTS

A. Datasets and Setup

In this section we cover experimental evaluations on three large scale datasets and publicly available.

- Multi-Modal Leap Motion Dataset (MMHGD)[1]: This dataset was created by the Universidad Politécnica de Madrid. It contains 15 diffrent one-hand gestures divided into 16 diffrent hand pose captured by the LMC device. Hand gestures were performed by 25 different subject (8 women and 17 men). Every gesture has 20 repetitions

[1] https://www.gti.ssr.upm.es/data/MultiModalHandGesture-dataset

per subject, for each instance, both infrared images and skeletal information were collected.

- Multi Model Leap Motion Dataset (2MLMD)[2]: It contains skeletal and depth data of 30 hand gestures that are split into 6 static gestures, 16 dynamic gestures with one hand and 8 dynamic gestures with two hands. All gestures were performed by 135 participants. Every gesture has 5 repetitions per subject.
- Multi-view Hand Pose Dataset (ML2HP)[3]: Two LMCs positioned horizontally and vertically were used in a multi-view recording configuration to record the hand poses. Consequently, this dataset was divided into two smaller datasets, ML2HP-H and ML2HP-V. Each one has 21 people with 17 distinct positions. 1000 frames of skeletal and infrared data are included in each instance.

For each dataset, we used only the skeletal modality. In our experiments, we set the training epoch to 100. Further, our model was trained with the ADAM optimizer and CrossEntropyLoss as the loss function. The learning rate was set to 0.001 and a weight decay of $1\mathrm{e}{-3}$. The batch size was adjusted to 32 for all datasets. We split the datasets into 80% for training and 20% for testing and evaluation. All experiments were performed on NVIDIA GeForce RTX 3060 GPU.

B. Results and discussion

- ***Quantitative results***

Results are shown in Table III, which compares the performance of different spatial-graph-based approach through the mentioned datasets. Our evaluation using accuracy revealed that the S-GINs significantly outperform GCNs, GAT, and GraphSAGE. Where S-GINs achieved an accuracy of 72%, 81%, 85% and 83% respectively on MMHGD, 2MLMD, ML2HP-H and ML2HP-V datatsets. The strength of S-GINs lies in their capacity to distinguish different graph structures by mapping them to distinct representations in the embedding space. This implies addressing the challenging graph isomorphism problem, where isomorphic graphs should be mapped to the same representation, and non-isomorphic graphs to different representations. This ability explains the effectiveness of S-GINs when dealing with different types of gestures, both static and dynamic, as it maps static graphs into similar embeddings and dynamic graphs into distinct representations.

TABLE III
PERFORMANCE ANALYSIS ON DIFFERENT DATASETS OF S-GINS

Method	MMHGD	2MLMD	ML2HP-H	ML2HP-V
GCN [11]	63%	72%	37%	36%
GAT [20]	58%	50%	79%	72%
GraphSAGE [10]	69%	79%	34%	40%
GIN [23]	72%	81%	85%	83%

- ***Qualitative results***

To discuss more the obtained results we used the confusion matrix as a metric to evaluate our proposed model as shown in Fig 3. For the 2MLMD dataset, the confusion matrix (b) demonstrates that S-GINs achieved high accuracy for most gestures. However, S-GINs struggle with classifying reciprocal dynamic gestures (19, 20) and (21, 22), indicating that these gestures share similar patterns that are challenging for the model to distinguish. For static gestures, confusion matrix (a), (c), (d) show a dominant diagonal, indicating that the S-GINs model classifies most samples correctly but the model still struggles to accurately classify gestures performed with one hand in the MMHGD, ML2HP-H, and ML2HP-V datasets (such as ”like” and ”dislike”). Thus, gesture similarity poses a significant challenge in HGR systems.

V. CONCLUSION

In this study, we addressed the problem of over-smoothing in graph based models faced with state-of-the-art spatial GNN such as GCN, GAT and GraphSAGE. S-GINs presents a novel method for skeletal-based HGR methods. The key contribution of our approach lies in its ability to differentiate non-isomorphic graphs using a dynamic distinguishing factor aggregation technique adapted to each method. The aggregation mechanism used in S-GINs did lead to capture local and global structure features compared to previous frameworks, resulting in a better accuracy level.
In this work, our focus was only on capturing spatial dependencies within hand gesture data using S-GINs. However, gestures inherently evolve over time, making temporal modeling crucial for improving recognition performance. As future work, we aim to integrate a temporal component into our framework, allowing it to aggregate temporal dynamics more effectively.

[2]https://sites.google.com/site/benkhalifaanouar1/6-datasets
[3]https://edatos.consorciomadrono.es

REFERENCES

[1] Umme Aiman and T. Ahmad. Angle based hand gesture recognition using graph convolutional network. *Computer Animation and Virtual Worlds*, 35, 2023.

[2] N. M. Bhiri, S. Ameur, I. Jegham, et al. 2mlmd: Multi-modal leap motion dataset for home automation hand gesture recognition systems. *Arabian Journal for Science and Engineering*, 2024.

[3] Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks?, 2022.

[4] Xiang Chang, Zihe Chen, Xiaoyan Dong, Yuxin Cai, Tingmin Yan, Haolin Cai, Zherui Zhou, Guyue Zhou, and Jiangtao Gong. “it must be gesturing towards me”: Gesture-based interaction between autonomous vehicles and pedestrians. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, page 1–25. ACM, May 2024.

[5] H. Chen, H. Zhao, and B. Qi et al. Gesture recognition of graph convolutional neural network based on spatial domain. *Neural Computing and Applications*, 35:2157–2167, 2023.

[6] Yuxin Chen, Ziqi Zhang, Chunfeng Yuan, Bing Li, Ying Deng, and Weiming Hu. Channel-wise topology refinement graph convolution for skeleton-based action recognition, 2021.

[7] Manuel Gil-Martín, Marco Raoul Marini, Rubén San-Segundo, and Luigi Cinque. Dual leap motion controller 2: A robust dataset for multi-view hand pose recognition. *Scientific Data*, 11, 10 2024.

[8] Xiaofeng Guo, Qing Zhu, Yaonan Wang, and Yang Mo. A hierarchical attention gcn network for body-hand gesture recognition. pages 799–804, 11 2023.

[9] Xiaofeng Guo, Qing Zhu, Yaonan Wang, Yang Mo, Jingtao Sun, Kai Zeng, and Mingtao Feng. A fast-response dynamic-static parallel attention gcn network for body–hand gesture recognition in hri. *IEEE Transactions on Industrial Electronics*, 71:5993–6004, 2024.

[10] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs, 2018.

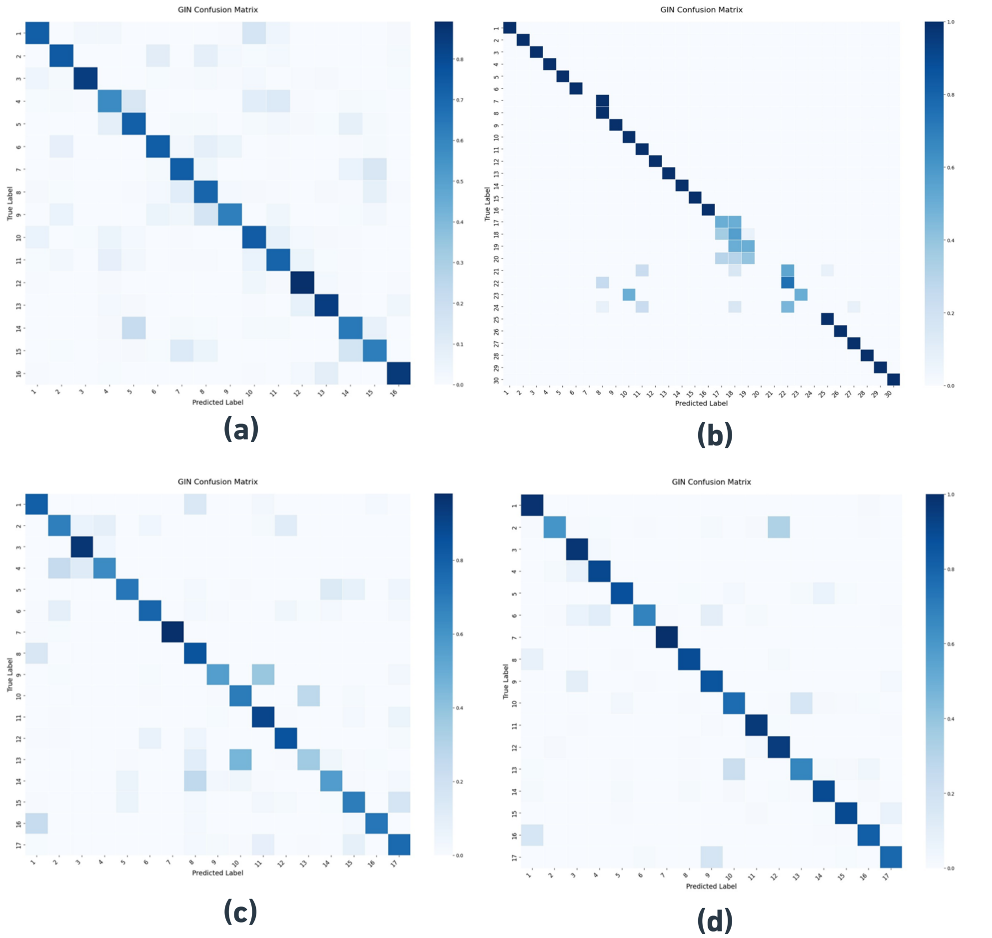

Fig. 3. Confusion matrix of S-GINs : (a) on MMHGD,(b) on 2MLMD, (c) on ML2HP-H, (d) on ML2HP-V

[11] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks, 2017.

[12] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. Mediapipe: A framework for building perception pipelines, 2019.

[13] Nahla Majdoub Bhiri, Safa Ameur, Ihsen Alouani, Mohamed Ali Mahjoub, and Anouar Ben Khalifa. Hand gesture recognition with focus on leap motion: An overview, real world challenges and future directions. *Expert Systems with Applications*, 226:120125, 2023.

[14] Tomás Mantecón, Carlos R Del-Blanco, Fernando Jaureguizar, and Narciso García. A real-time gesture recognition system using near-infrared imagery. *PloS one*, 14(10):e0223320, 2019.

[15] Sen Qiao, Yilin Wang, and Jian Li. Real-time human gesture grading based on openpose. In *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–6, 2017.

[16] Prema Selvam, Deena Gnanasekaran, Hemalatha D., Aruna K B, and Hashini S. Gaming using different hand gestures using artificial neural network. *EAI Endorsed Transactions on Internet of Things*, 02 2024.

[17] Jan Smisek, Michal Jancosek, and Tomas Pajdla. 3d with kinect. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 1154–1160, 2011.

[18] K Sneha and R Kayalvizhi. Mid-air gesture based multi-finger control system for paralyzed patients using leap motion. In *2023 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–6, 2023.

[19] Tsung Han Tsai, Yi Jhen Luo, and Wei Chung Wan. A skeleton-based dynamic hand gesture recognition for home appliance control system. In *IEEE International Symposium on Circuits and Systems (ISCAS 2022)*, pages 3265–3268. IEEE, 2022.

[20] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks, 2018.

[21] Jun Xie, Qiguang Miao, Ruyi Liu, Wentian Xin, Lei Tang, Sheng Zhong, and Xuesong Gao. Attention adjacency matrix based graph convolutional networks for skeleton-based action recognition. *Neurocomputing*, 440:230–239, 2021.

[22] Hao Xing and Darius Burschka. Skeletal human action recognition using hybrid attention based graph convolutional network. *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 3333–3340, 2022.

[23] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks?, 2019.

[24] Xinzheng Xu, Xiaoyang Zhao, Meng Wei, and Zhongnian Li. A comprehensive review of graph convolutional networks: approaches and applications. *Electronic Research Archive*, 31:4185–4215, 05 2023.

[25] Deyuan Zhang, Hongwei Gao, Hailong Dai, and Xiangbin Shi. Human skeleton graph attention convolutional for video action recognition. In *2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT)*, pages 183–187, 2020.

[26] Hao Zhou, Hoang Thanh Le, Shen Zhang, Son Lam Phung, and Gursel Alici. Hand gesture recognition from surface electromyography signals with graph convolutional network and attention mechanisms. *IEEE Sensors Journal*, 25(5):9081–9092, 2025.