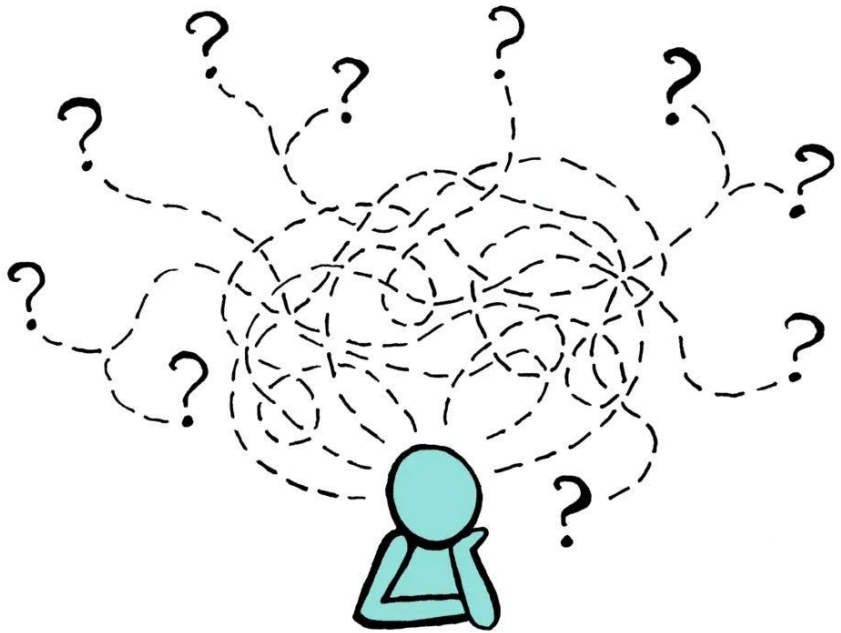
# Pose-based Tremor Classification for Parkinson's Disease Diagnosis from Video

Chengche Tsai, Alessandro Folloni

# What’s Parkinson’s Disease (PD)?

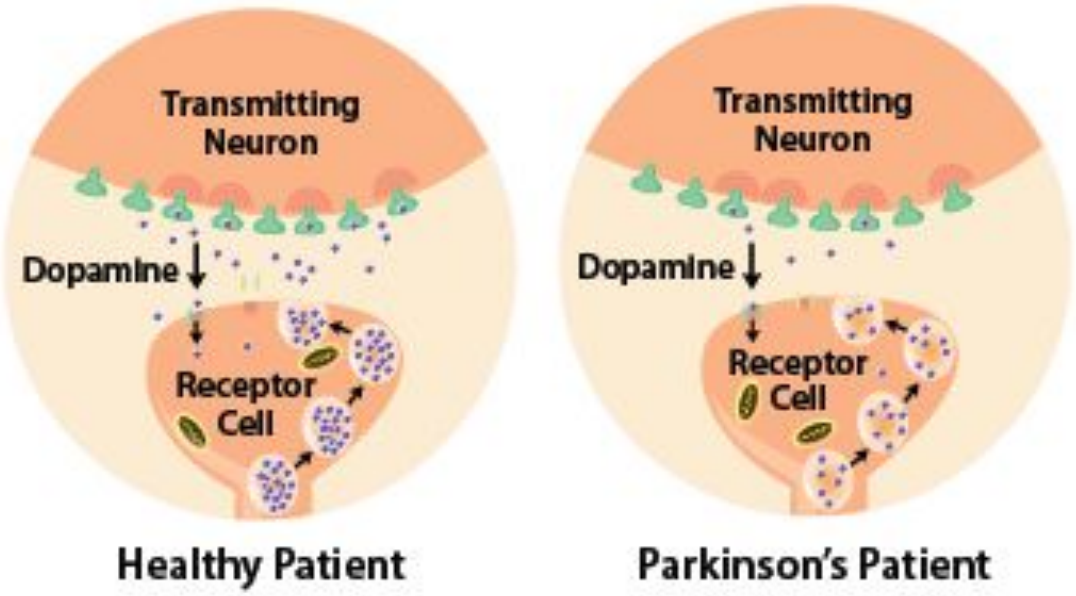


# What’s Parkinson’s Disease (PD)?

Some keywords:

- Progressive neurodegenerative disorder
- Parkison’s tremors

# What’s Parkinson’s Disease (PD)?



# Common medical approach

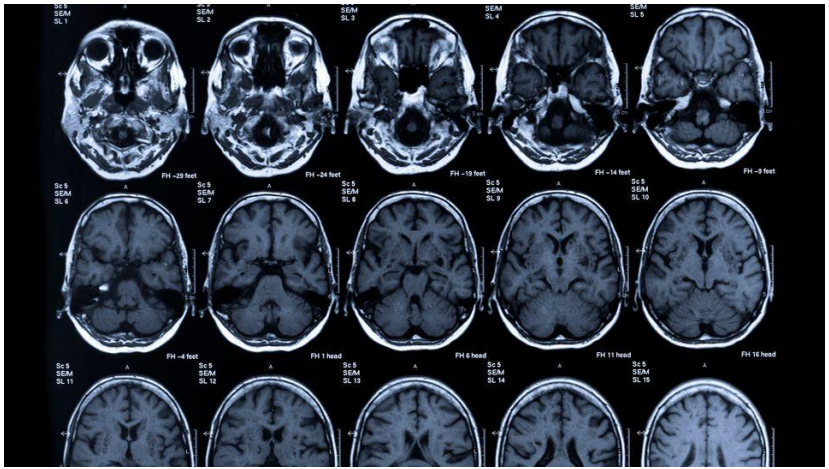
- Relies mostly on clinical experience (no two patients have the same symptoms or rate of progression)
- Diagnostic accuracy is 73-84%

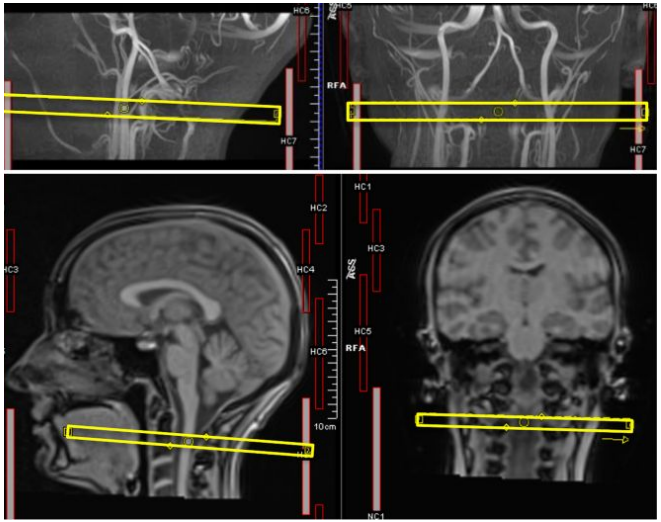
# Paper work

It focuses on the classification of Parkinson’s tremors (PT) through video analysis

# Past related works

- Some methods achieved impressive performance analyzing the neuroimaging, cerebrospinal fluid
- They analyze every possible symptom





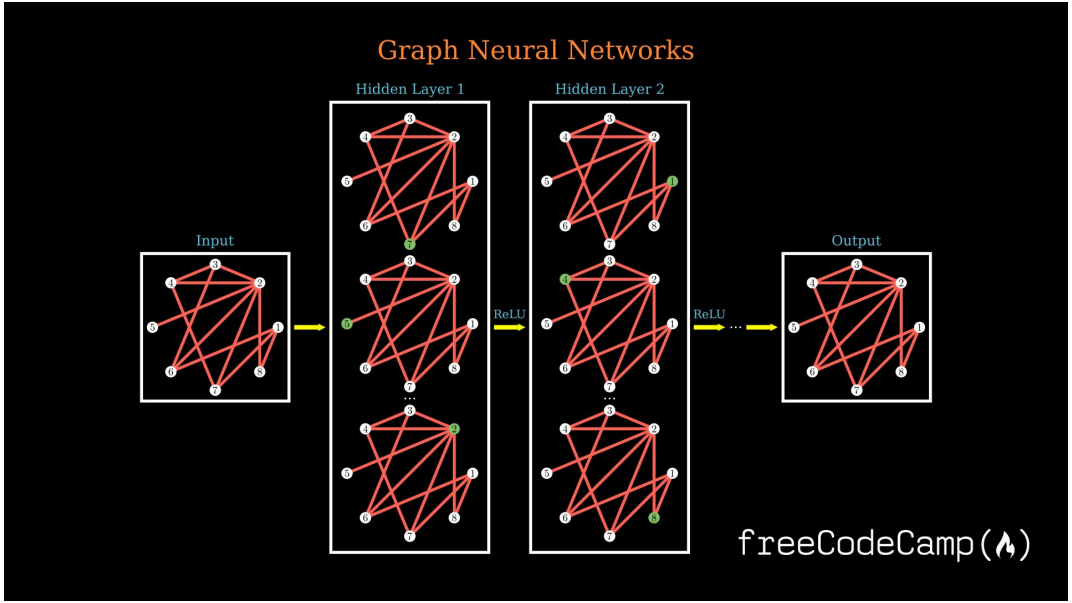
IT'S


FEDERAL RESERVE NOTE
HB 88478962 N
B2
THE UNITED STATES OF AMERICA
ONE HUNDRED DOLLARS

SIVE

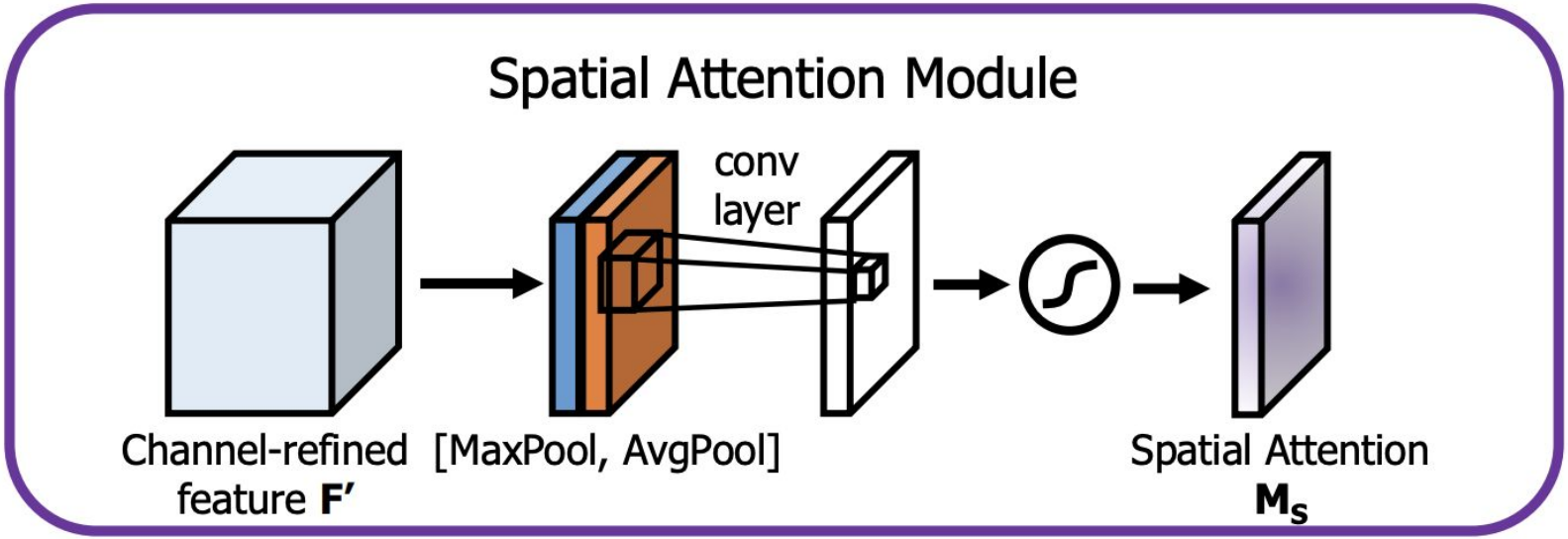
# Novelties of the paper

- Graph neural network to learn the spatial relationship between body joints from graph-structured data



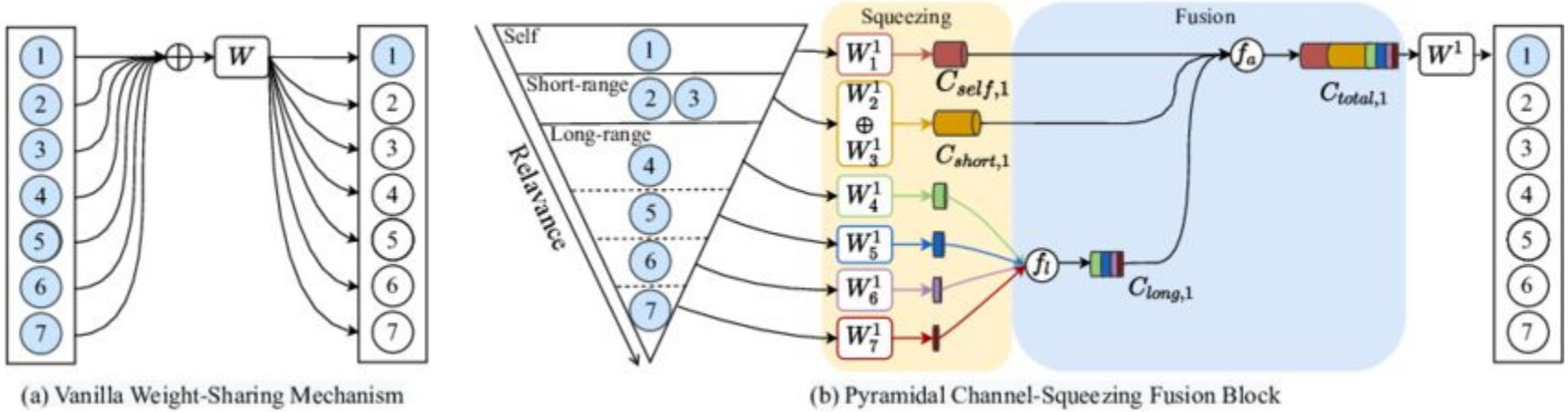
# Novelties of the paper

- Spatial attention module



# Novelties of the paper

- Pyramidal Channel-Squeezing-Fusion Block (PCSF)




(a) Vanilla Weight-Sharing Mechanism

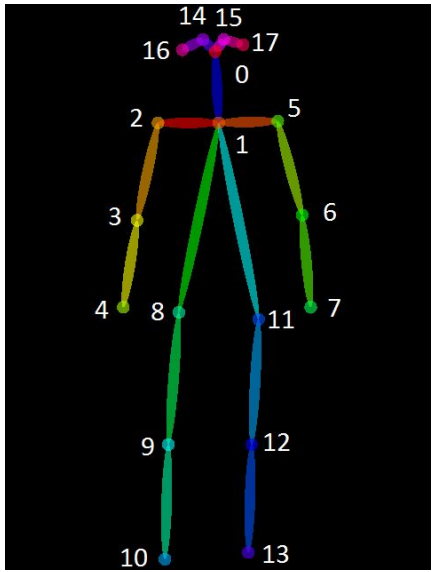
(b) Pyramidal Channel-Squeezing Fusion Block

# Method

- Input: video recordings of each participant sitting in a chair in a normal upright position with various poses

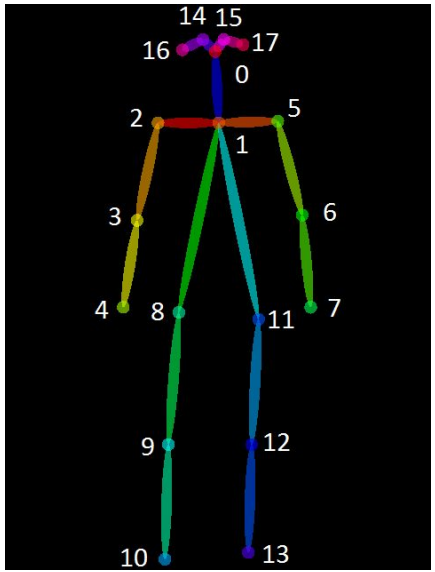
# Method

We extract the human joint position features from the RGB video by OpenPose algorithm



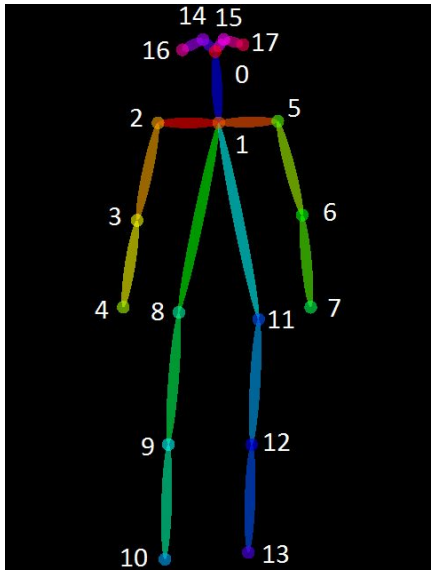
# Method

Each video frame is fed to OpenPose (no 3D estimation, 2D is robust and efficient for the task)



# Method

We extract 18 OpenPose-skeleton format landmarks with 2D coordinates (x,y) and a confidence score c (for the reliability of the estimation)



# Method

We will use the seven upper body landmarks because PT usually occur in the upper part of the body

# SPAPNet

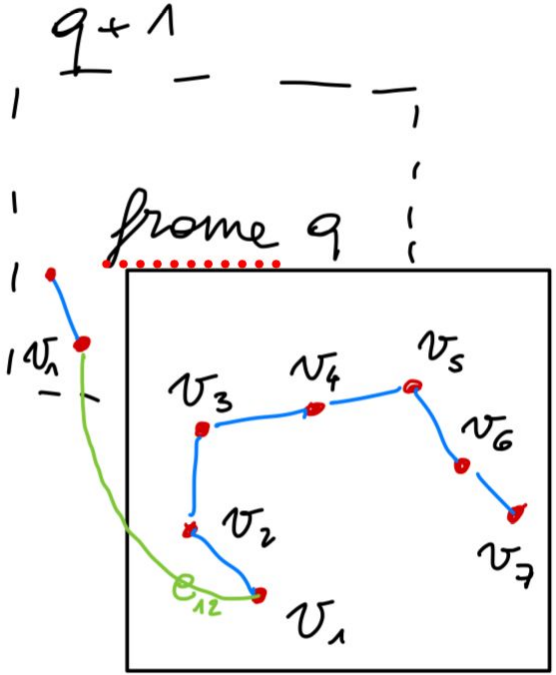
Namely, Spatial Pyramidal Attention Parkinson’s tremor classification Network:

- Graph Neural Network
- Spatial Attention Mechanism
- PCSF block

# Graph Neural Network (GNN)

Pose graph G=(V,E)

- {V=$v_{pq}$} denotes the joints positions
- $v_{pq}$ represents the p-th joint at q-th frame
- E is the edge set and includes (1) the intra-skeleton connection between the human joints in the same frame and (2) the inter-frame connections between consecutive frames



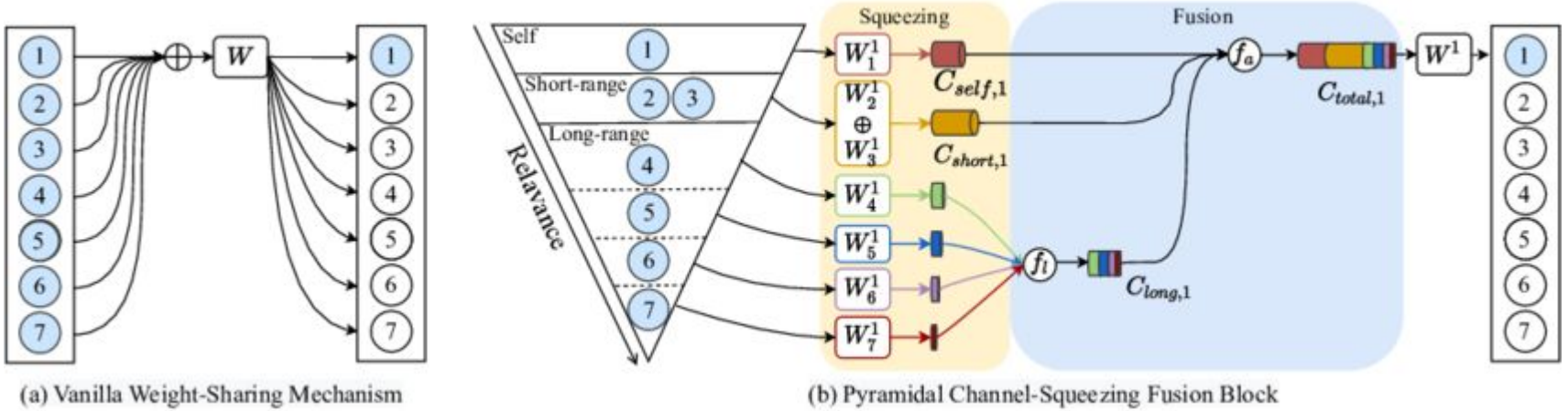
# Spatial attention mechanism

- To improve the PT classification performance
- It interprets the important joints that the network considers at each frame by attention weights and their temporal aggregation

$$\mathbf{h_i} = \sigma\left(\sum_{j \in \mathcal{N}^i} \mathbf{W}_j^i \mathbf{x}_j \hat{a}_{ij}\right)$$

# PCSF

- Extension of the spatial attention module
- Novel lightweight inverted pyramid architecture
- Good for extracting PT information and filter noise




(a) Vanilla Weight-Sharing Mechanism

(b) Pyramidal Channel-Squeezing Fusion Block

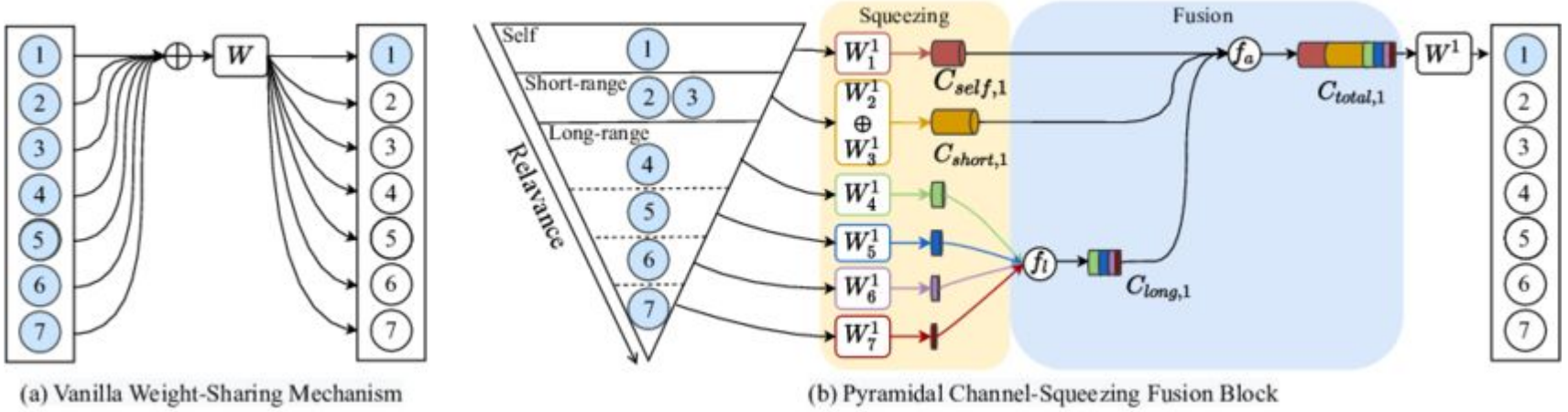
# Motivation behind

- Information gain decreseas exponentially with increasing distance between graph nodes

- Clinical evidence shows that PT occurs only on one side of the PD patient’s upper body

# PCSF

- Final attention weight for target joint-1 is learned from the relevant joints (due to adjacency mostly)




(a) Vanilla Weight-Sharing Mechanism

(b) Pyramidal Channel-Squeezing Fusion Block

# Channel-squeezing block

Farther assumptions:

- Node m to be the target node
- Node k to be the relevant node to m
- Shortest path between m and k is k-a

$$C_{out,k} = b \times C_{in}, \quad |k - m| \leq 2$$

$$C_{out,k} = d^{|k-m|} C_{in}, \quad |k - m| > 2$$

# Channel-fusion block

To fuse the relevancy information for the node m:

1. we fuse long-range features from less-related joints by f_L
2. then fuse all features by f_A

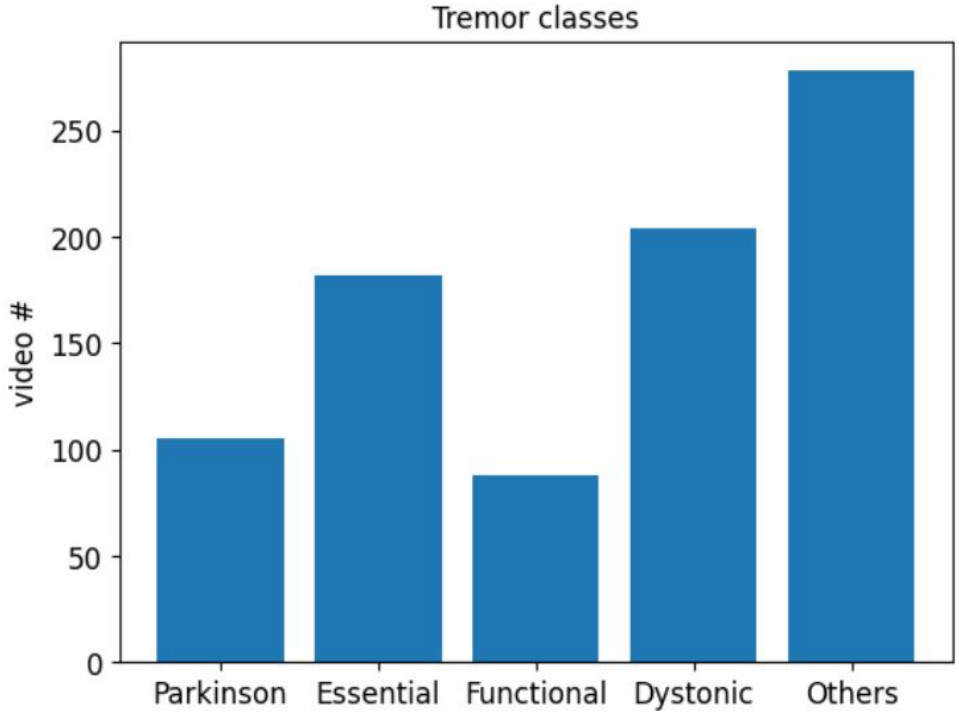
$$\mathbf{h_m} = f_a[\mathbf{h_{slef}}, \mathbf{h_{short}}, f_l(\mathbf{h_{long,p}})\mathbf{W^m}$$

# Experiment

# Dataset

- 917 videos from 55 participants
  - Sitting in a chair and
  - Performing a set of tasks

- Length: 18 - 112 seconds



# Clips

- Each clip is about 100 frames (~3.5 sec)
- Participant is not occluded by the interaction with the clinician
- Tremor types in clips are weakly labeled based on their parent videos.

# Before classification: tremor frequency

- Frequency dissociation in (PT, ET): simultaneous tremors at two limbs different frequencies (> 0.1 Hz)

- 

| Type | **Frequency (Hz)** |
|---|---|
| **Parkinsonian** | 3-5, rhythmic |
| **Essential** | 4-12 |
| **Functional** | 6-11 |
| **Dystonic** | < 7, variable |

→ Frequency & shaking pattern require temporal understanding

# Tasks

- Binary classification (PT vs non-PT labels)

- Five class classification (PT, ET, FT, DT, and No tremor)

# Result-1

sighr 11/21/2023, 9:27:42..

- Spatiotemporal model
- 4000 citations

| | Method | Binary Classification | | | |
|---|---|---|---|---|---|
| | | AC | SE | SP | F1 |
| | CNN-LSTM [28] | 81.0 | n/a | 79.0 | 80.0 |
| | LSTM [28] | 80.0 | n/a | 79.0 | 79.0 |
| | SVM-1 [28] | 53.0 | n/a | 63.0 | 55.0 |
| | ST-GCN [31] | 87.7 ± 3.8 | 88.3 ± 5.3 | 87.4 ± 3.1 | 87.0 ± 4.4 |
| | CNN-Conv1D | 81.6 ± 5.7 | 83.4 ± 9.1 | 80.7 ± 4.4 | 80.3 ± 6.0 |
| | Decision Tree | 74.5 ± 4.7 | 73.4 ± 5.7 | 75.8 ± 4.0 | 73.6 ± 4.6 |
| | SVM | 64.3 ± 5.4 | 62.2 ± 7.5 | 66.7 ± 4.6 | 63.1 ± 7.1 |
| Ours | SPAPNet - full | **90.9 ± 3.4** | **90.7 ± 5.0** | **91.3 ± 2.3** | **90.6 ± 3.7** |
| Ours | w/o PCSF | 88.4 ± 4.5 | 90.4 ± 6.9 | 87.0 ± 3.7 | 87.5 ± 5.2 |
| Ours | w/o Attention | 82.6 ± 5.3 | 82.7 ± 6.0 | 82.8 ± 5.1 | 81.3 ± 6.8 |

# Result-2

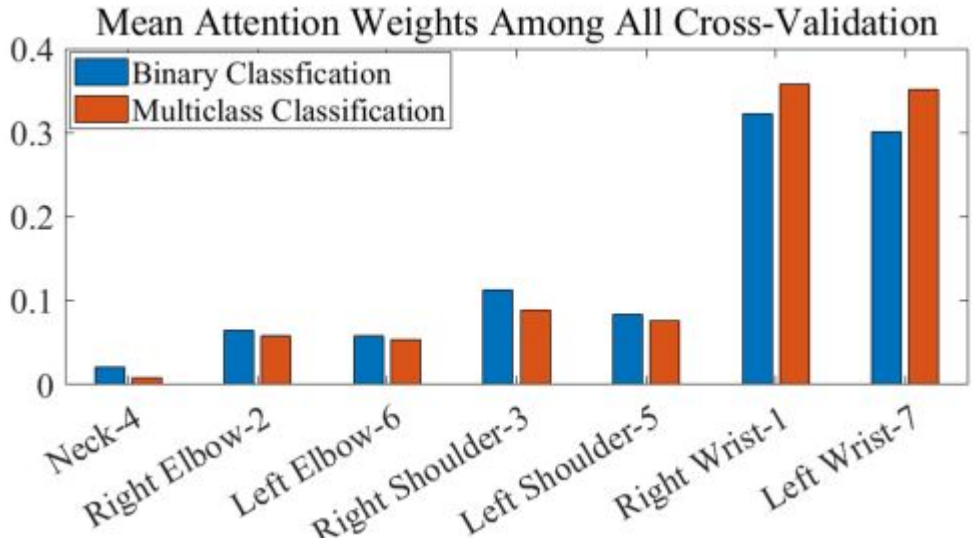
| Method | | AC | SE | SP | F1 |
|---|---|---|---|---|---|
| | | Multiclass Classification | | | |
| ST-GCN [31] | | 70.3 ± 6.9 | 69.5 ± 6.4 | 90.7 ± 5.4 | 67.9 ± 6.7 |
| CNN-Conv1D | | 63.1 ± 6.5 | 59.5 ± 5.6 | 90.8 ± 7.4 | 61.9 ± 8.3 |
| Decision Tree | | 54.3 ± 5.7 | 49.0 ± 7.3 | 92.3 ± 5.4 | 55.5 ± 6.5 |
| SVM | | 47.6 ± 6.4 | 45.7 ± 6.9 | 91.6 ± 6.1 | 52.1 ± 7.2 |
| Ours | SPAPNet - full | **73.3 ± 6.8** | **72.8 ± 5.1** | **92.3 ± 4.1** | **70.7 ± 6.5** |
| | w/o PCSF | 69.1 ± 6.9 | 69.9 ± 4.0 | 88.2 ± 4.6 | 65.7 ± 7.1 |
| | w/o Attention | 65.9 ± 6.8 | 64.2 ± 5.5 | 90.4 ± 7.9 | 65.0 ± 7.9 |

# Ablation

| | | | Binary Classification | | |
|---|---|---|---|---|---|
| | | | Multiclass Classification | | |
| GNN | | ST-GCN [31] | 70.3 ± 6.9 | 69.5 ± 6.4 | 90.7 ± 5.4 | 67.9 ± 6.7 |
| CNN | | CNN-Conv1D | 63.1 ± 6.5 | 59.5 ± 5.6 | 90.8 ± 7.4 | 61.9 ± 8.3 |
| | | Decision Tree | 54.3 ± 5.7 | 49.0 ± 7.3 | 92.3 ± 5.4 | 55.5 ± 6.5 |
| | | SVM | 47.6 ± 6.4 | 45.7 ± 6.9 | 91.6 ± 6.1 | 52.1 ± 7.2 |
| GNN | Ours | SPAPNet - full | **73.3 ± 6.8** | **72.8 ± 5.1** | **92.3 ± 4.1** | **70.7 ± 6.5** |
| | | w/o PCSF | 69.1 ± 6.9 | 69.9 ± 4.0 | 88.2 ± 4.6 | 65.7 ± 7.1 |
| | | w/o Attention | 65.9 ± 6.8 | 64.2 ± 5.5 | 90.4 ± 7.9 | 65.0 ± 7.9 |

PCSF: “nearby joints matter the most” concatenation

# Visualize attention




# Personal thoughts

- If the traditional CV field is too competitive and you want to take a break, there are some low-hanging fruit outside of CS, like medicine.

- Their PCSF layer doesn't make too much sense in medicine. Nearby joints don't matter the most for diagnosing tremor.