

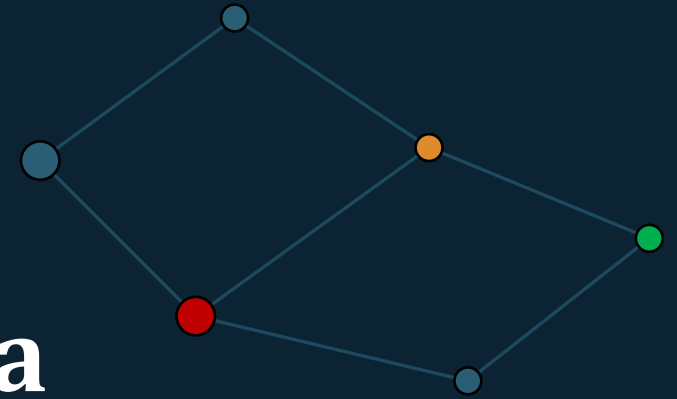
ELLIIT FOCUS PERIOD 26 HALMSTAD · SEMINAR TALK

Graph Neural Networks for Localizing Structured Sensor Data

From breathing motion to hand kinematics: sensors as nodes, coupling as edges, and explanations that say where the evidence lives.

Syed Muhammad Hamza Zaidi

KMD Lab · TACTIC Graduate School · Otto-von-Guericke University Magdeburg
syed.zaidi@ovgu.de



INTRODUCTION

I am



Syed Muhammad Hamza Zaidi

PhD researcher · Otto-von-Guericke University Magdeburg

- Temporal graph neural networks and XAI
- Physiological and biomechanical sensor time series
- Supervised by Prof. Myra Spiliopoulou and Prof. Benjamin Noack
- Background in Digital engineering



LinkedIn

THE WORK SO FAR

IEEE CBMS 2025

Graph neural networks for the localization of breathing abnormalities

IEEE CBMS 2026

Explainable finger kinematics decoding with temporal graph neural networks

XKDD @ ECML PKDD 2026

Counterfactual transition graphs: evaluating cross-class transition quality

XAI @ IJCAI 2026

Counterfactuals by Design: Residual Decomposition and Back-to-Normality Editing for Time Series

TACTIC Graduate School

- “Towards Co-Evolution in Human-Technology Interfaces”
- Interdisciplinary graduate school at OVGU Magdeburg
- Computer science, Tissue engineering, Medicine, Electrical engineering, Mechanical engineering and Neuro Scientists

Human and technology adapting to each other



The TACTIC PhD students and supervisors, Magdeburg · tactic.ovgu.de



Knowledge Management & Discovery Lab



KMD Lab, Faculty of Computer Science, OVGU Magdeburg · kmd.ovgu.de

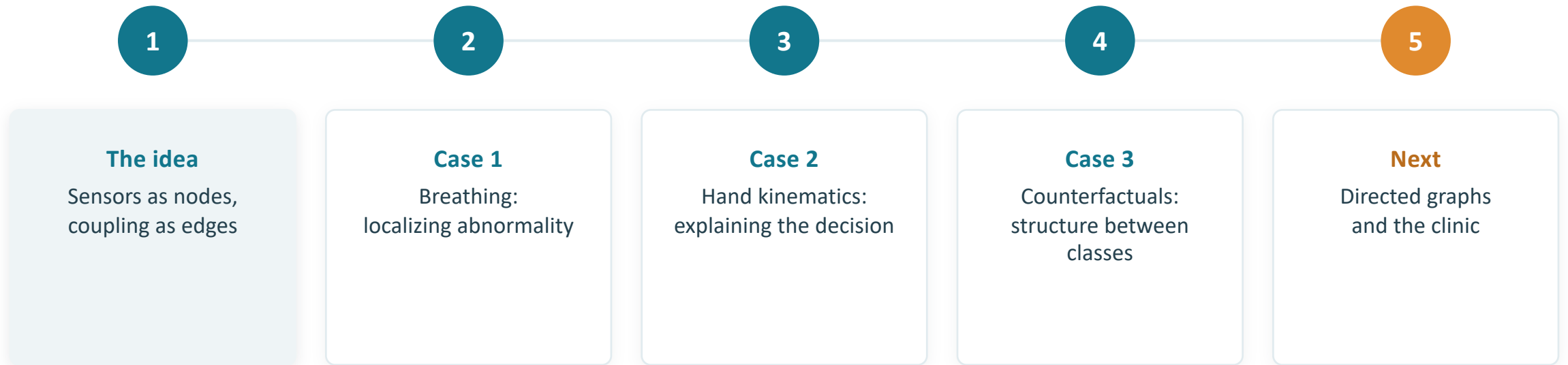


FACULTY OF
COMPUTER SCIENCE



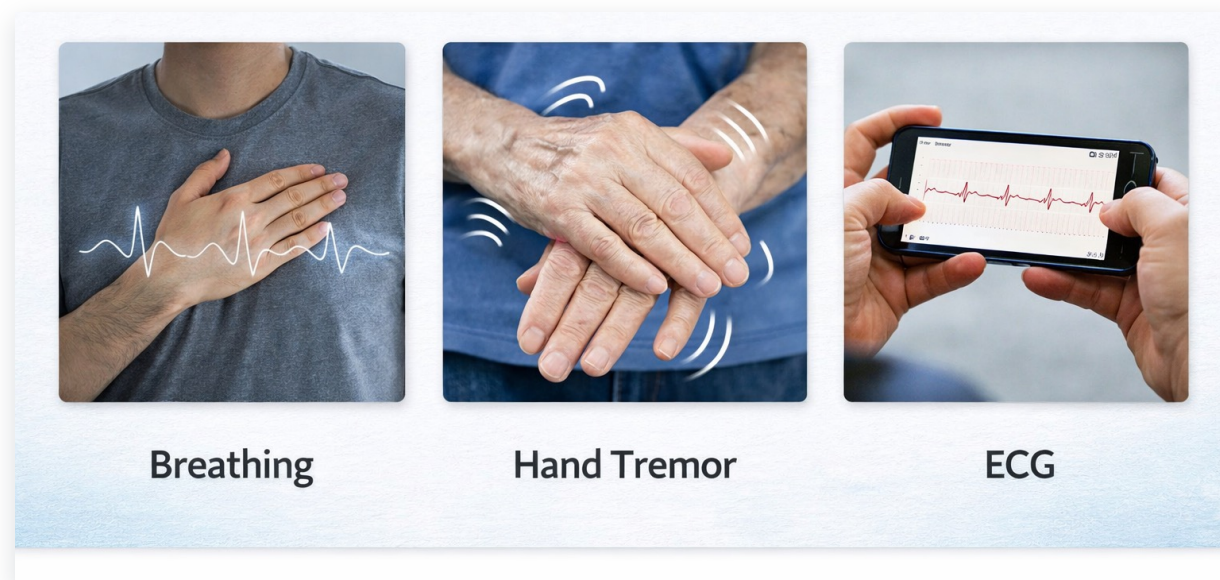
- Led by Prof. Dr. Myra Spiliopoulou
- Faculty of Computer Science, OVGU Magdeburg
- Mining evolving data: streams, sequences, trajectories
- Health analytics on sensor studies
- Interpretability and validation of learned models
- Multi-modal learning

Where we are going in the next minutes



One thread: moving from a model that says *that* something is wrong, to one that shows *where* the evidence sits and finally *what would have to change* for it to say something else.

Physiological function is distributed, not point like



Breathing, tremor and cardiac signals are all produced by many coupled sources.

- Breathing is produced by the whole thorax, region by region
- Hand function is produced by multiple finger joints moving together
- Wearables now give us many channels
- One channel can say “abnormal” it cannot say “where”

Three ways to look at a body, none of them enough

01

Single-point sensing

A chest strap, one wrist IMU.
Continuous and cheap, but one number:
no location, no structure.

02

Flat multi-channel models

Channels stacked into an unordered
vector. Anatomy is discarded, so the
model cannot generalise the notion of
“neighbour”.

03

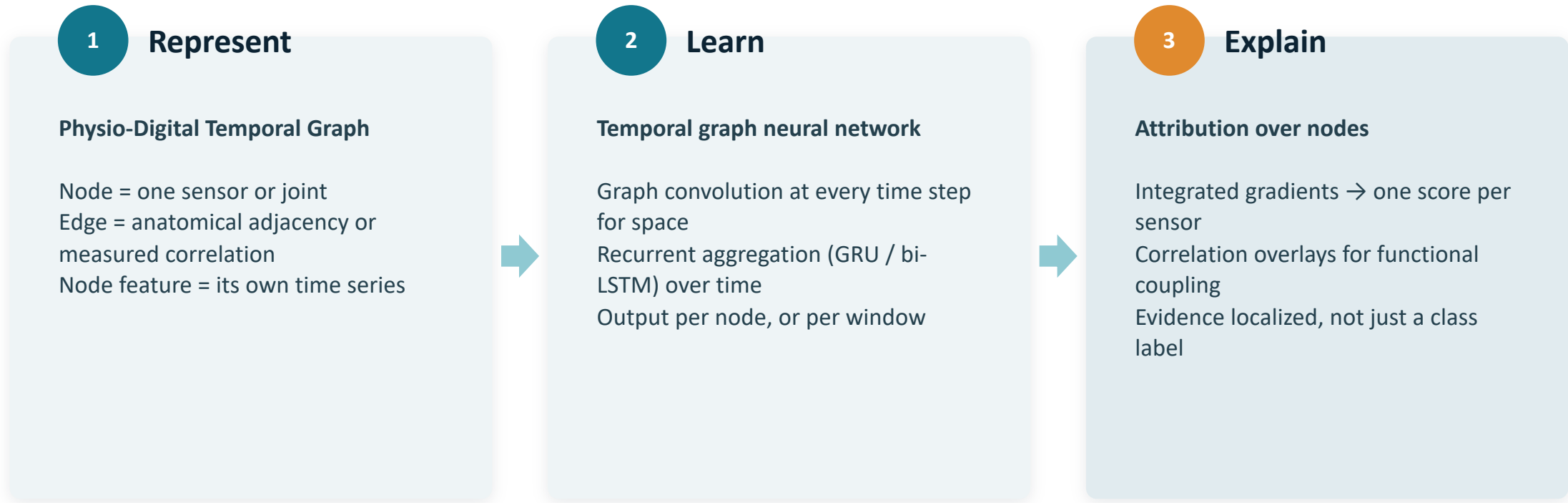
Imaging (X-ray, CT)

Precise localization, but radiation-based,
hospital-only and never continuous.

What a clinician actually asks for

Abnormality localized to a region of thorax or a finger joint and accompanied by a reason

One recipe, three case studies



The influence factor. A local change at one node does not stay local. it propagates to its neighbours. The graph is what makes that propagation measurable, and what turns a prediction into a localization.

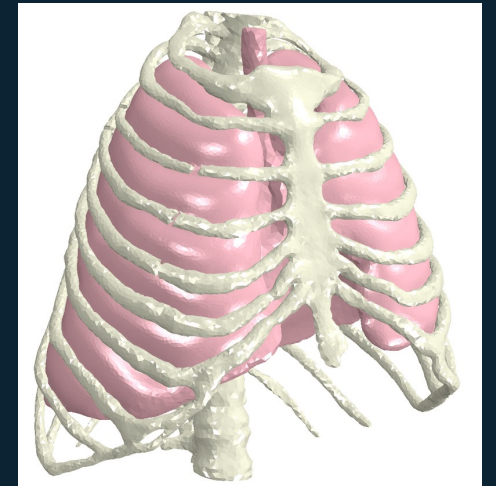
CASE STUDY 1 · CBMS 2025

Localizing breathing abnormalities

A wearable sensor network on the thorax, modelled as a graph, that says which region is breathing badly and how badly.

Syed M. H. Zaidi, Myra Spiliopoulou

IEEE CBMS 2025, pp. 630–635



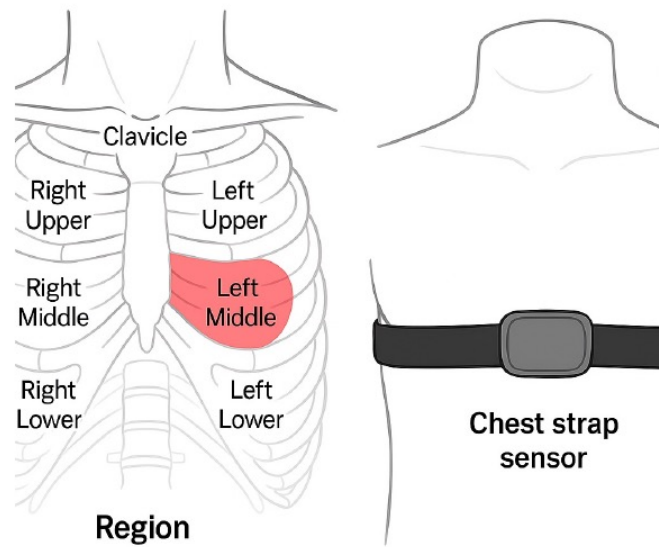
Why non-invasive breathing monitoring?

- Respiratory problems are a major global health burden
- Rib fractures, implants , post-operative recovery: the location matters clinically
- Thorax performance drops as the affected segment grows
- X-ray and CT show where but are radiation-based and periodic
- A vest is safe, cheap and can run at home for weeks

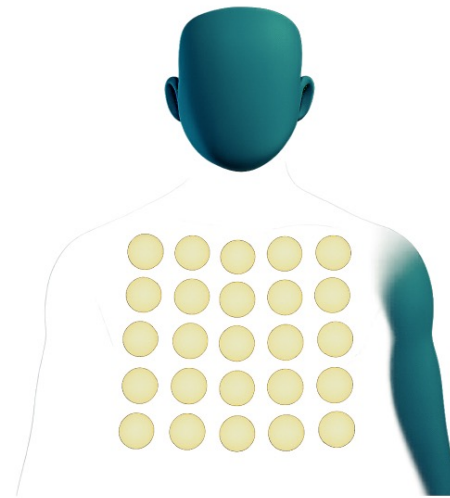
Goal: detect an abnormal breathing pattern and pinpoint its location on the chest at the same time, from wearable motion alone.



What makes this hard



A single strap sensor cannot provide localization on the thorax.

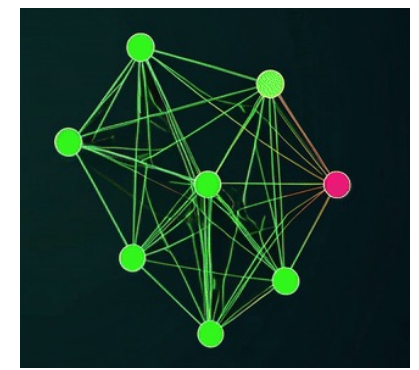


A wearable array covers the thorax, could localize

- Chest motion is regional: every area moves differently
- Traditional sensors are single-point
- Imaging cannot be continuous
- Almost no labelled clinical recordings exist

Hard constraint: no patient data. Everything here is a controlled simulation and I will come back to that.

Three questions, one model



RQ1

Detection

Can a wearable sensor network detect abnormal breathing motion at all?

RQ2

Localization

Can the same model say which region of the thorax is responsible?

RQ3

Severity

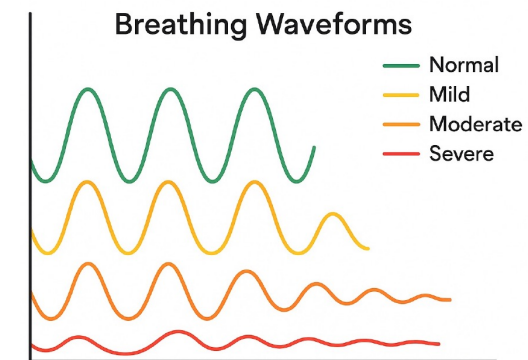
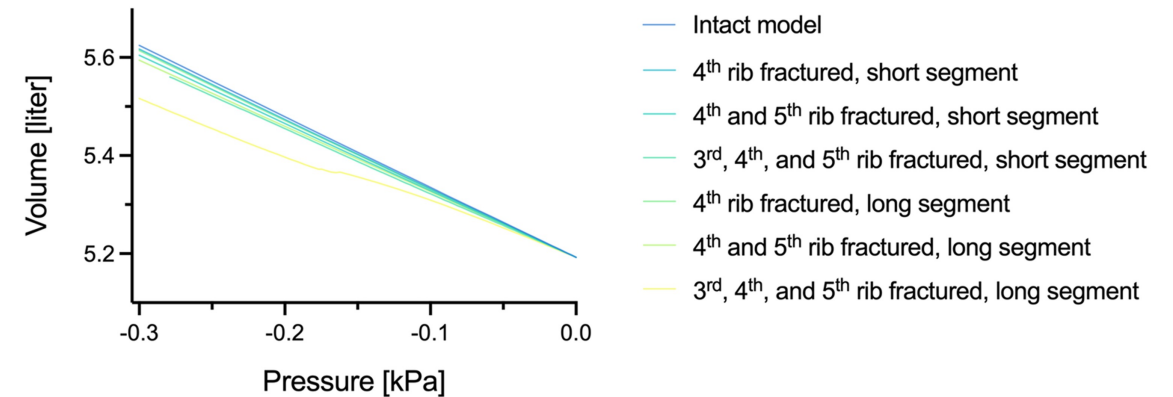
Can it grade how severe the abnormality is, per region, in one pass?

The design decision that answers all three at once: make the prediction node-level, so every sensor gets its own severity label.

A controlled synthetic dataset

- 5-minute abnormal breathing signals at 50 Hz, PAMAP2 as reference
- Severity levels: normal, mild, moderate, severe
- An “injury” is injected at a chosen sensor location
- Neighbours are perturbed too, scaled by an influence factor
- Shallow breaths, rapid breathing and Gaussian noise added
- Rates kept realistic: ~12–20 /min normal, up to ~50 /min abnormal

The influence factor is the ground truth for coupling: we know exactly how far the disturbance should spread.



Amplitude collapses and rhythm destabilises as severity increases.

The sensor network becomes a graph

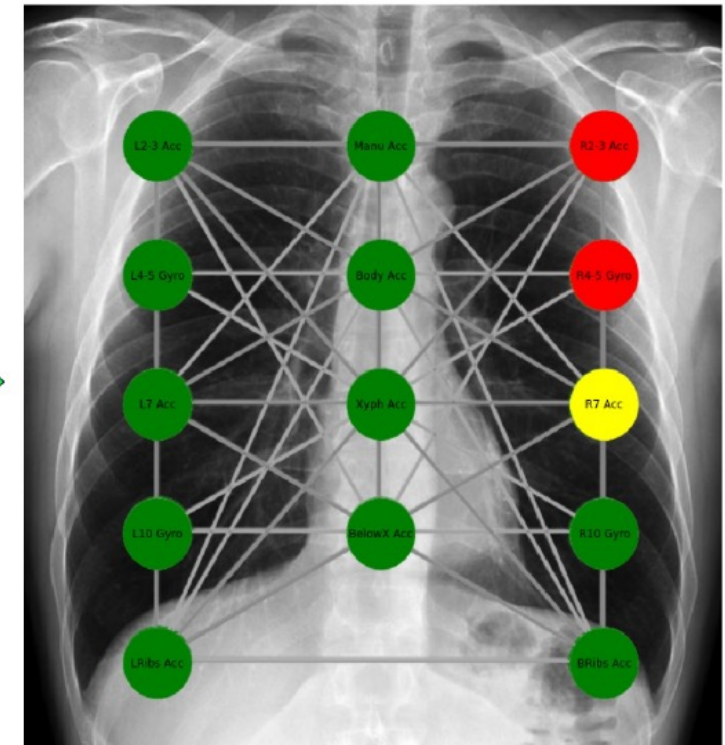
- 14 IMUs covering the thorax each sensor is one node
- Node feature = that sensor's own time series
- An edge exists only if both conditions hold

Spatial proximity

Connect sensors closer than a distance threshold τ_s on the body.

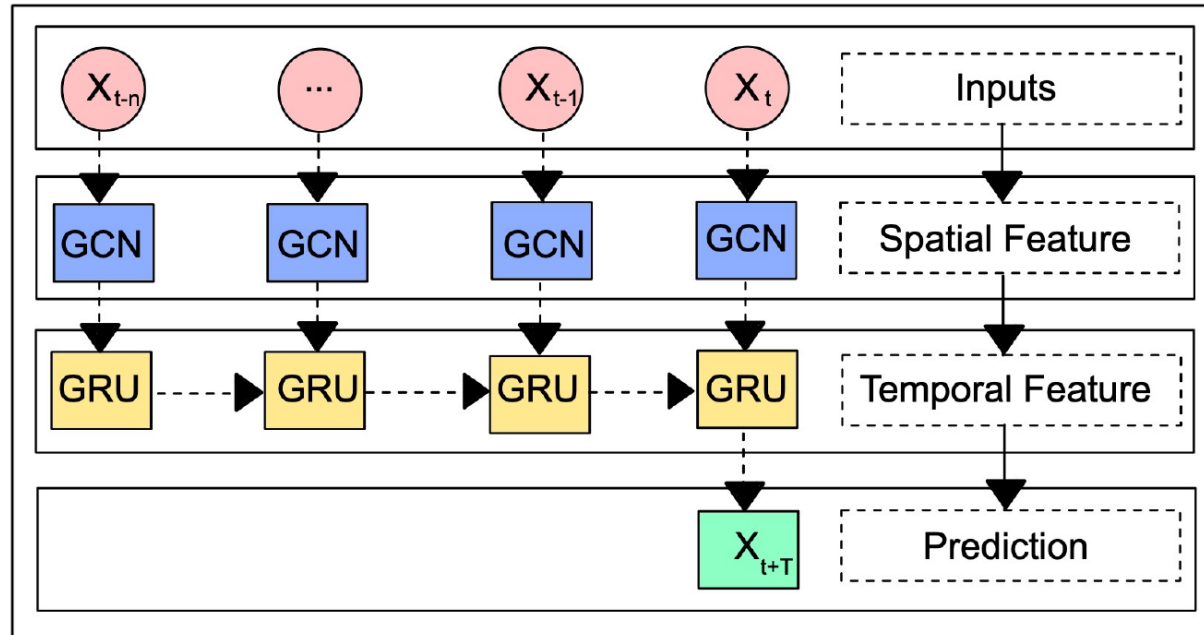
Temporal correlation

Keep the edge only if the Pearson correlation of the two series is $\geq \tau_c$.



14 IMU nodes over the thorax; colour = predicted severity.

Space with a GCN, time with a GRU



TGCN: graph convolution per time step, then recurrent aggregation, then a label per node.

- GCN per time step: each node mixes with its neighbours
- GRU carries breathing rhythm across time steps
- Final node state → 4-class severity label
- Node-level output = detection, location and severity

It detects and localize

~98%

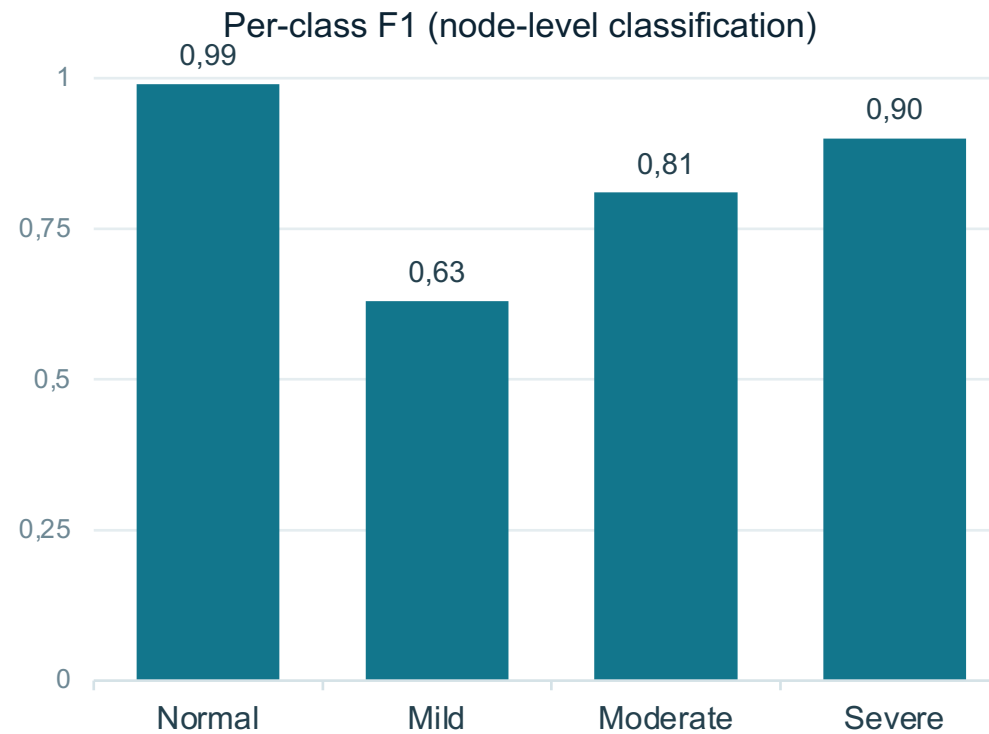
overall node-level accuracy on held-out simulated cases

0.90

F1 for the severe class (the class you must not miss)

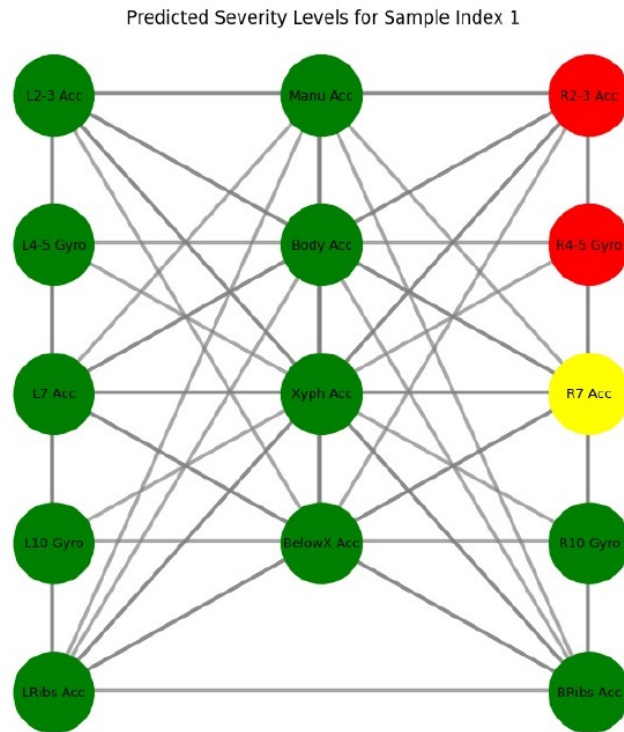
0.63

F1 for the mild class (confused with moderate)

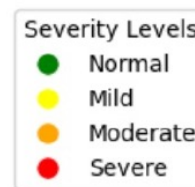


- Normal regions are near-perfectly identified
- Severe: precision 0.86, recall 0.94 (rarely missed)
- Moderate: precision 0.77, recall 0.86
- Mild: recall 0.82 but precision 0.51
- Mild is over-called as moderate: the boundary is genuinely blurry

The model points at the right part of the chest

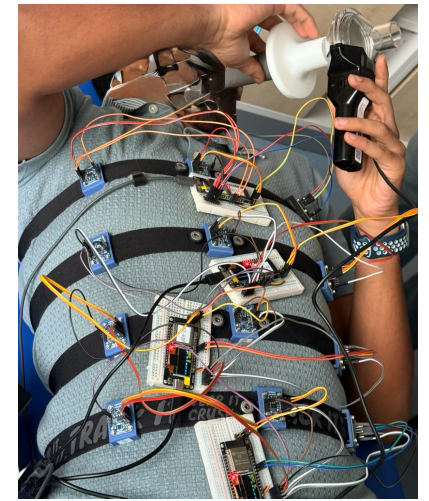


Ground Truth Severity Levels



- Predicted node severities recover the ground-truth region
- Abnormality placed on the upper right side: sensors R2-3 Acc and R4-5 Gyro
- The neighbouring node R7 Acc is picked up as mild, the influence factor made visible
- Node-level prediction

What it proved, and what it did not



What worked

- Node-level prediction gives localization
- Severity grading works where it matters most
- Proximity + correlation edges
- The influence factor is visible in the output

What it did not prove

- Synthetic data: the coupling was one we wrote ourselves
- The anatomy was assumed, maybe less sensors could also predict similar nature
- Mild vs moderate remains unresolved
- No explanation beyond the class label itself

Next step: a real dataset and an explanation that is separate from the prediction.

CASE STUDY 2 · CBMS 2026

Explainable finger kinematics decoding

Real recordings, real anatomy: 34 adults, an exoskeleton glove and an explanation expressed in finger joints.

Syed M. H. Zaidi, Lisa Klemm, Benjamin Noack, Myra Spiliopoulou

IEEE CBMS 2026



Hand function mirrors motor health

- Early detection: Stroke, Parkinson's disease and ALS
- Clinical assessment today requires a lot of time
- Exoskeleton gloves capture joint kinematics at 100 Hz, everyday tasks included
- Vision: a high-fidelity digital twin of the hand
- Track disease progression and rehabilitation over months, not visits

If we can decode what the hand is doing, and say which joints told us, the same machinery can track how that changes.



SenseGlove DK1.3 exoskeleton data glove.

Twenty channels, and the information is in the coupling

Why it is hard

- 20 joint-angle channels: high-dimensional and redundant
- Timing and coordination differ per person and per age group
- “Flat” models treat channels as an unordered vector
- Anatomical relations are discarded → weaker cross-subject transfer
- Clinicians need reasons at joint level, not the accuracy

RESEARCH QUESTIONS

RQ1

Does an anatomical prior help decode hand movements across unseen subjects?

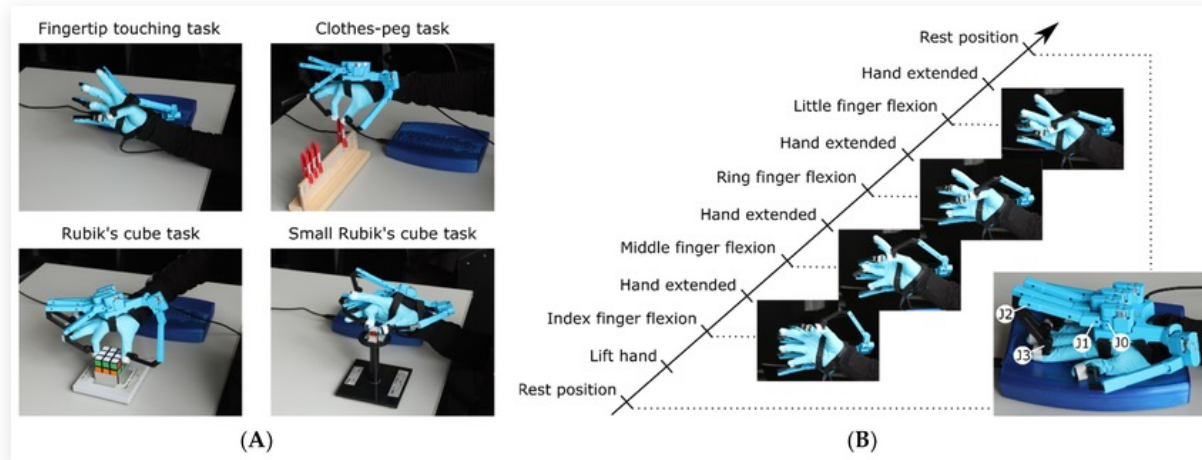
RQ2

Can we localize the evidence to individual finger joints?

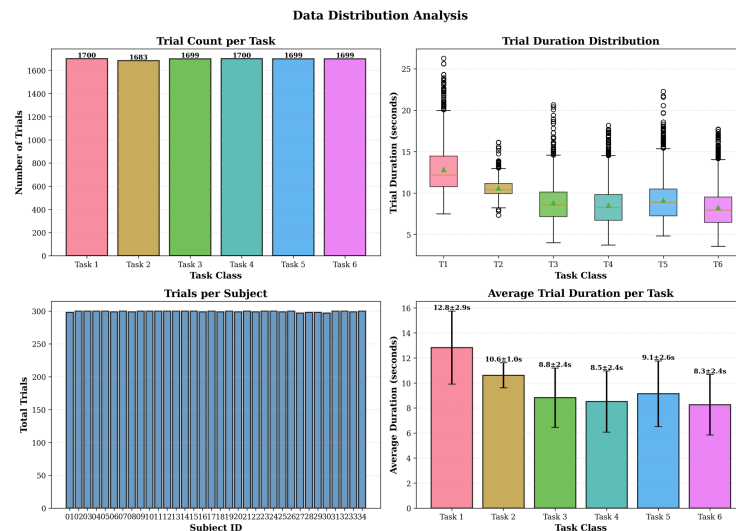
RQ3

Do coordination signatures differ between younger and older adults?

SenseGlove DK1.3: six everyday manipulation tasks



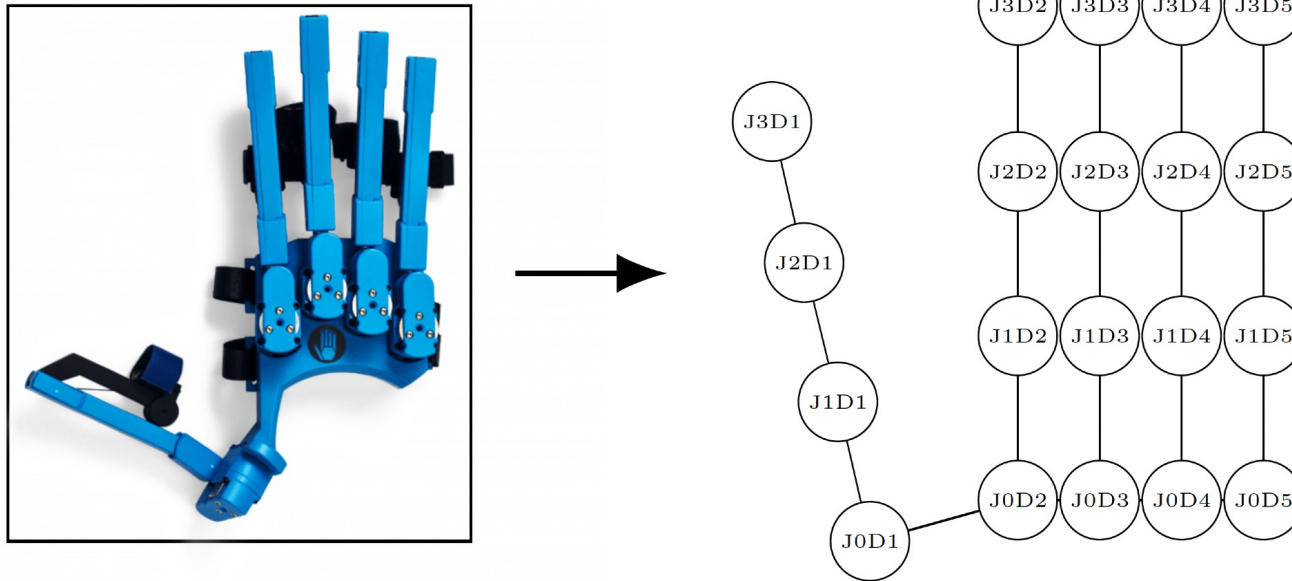
The recorded tasks (left) and the instructed movement sequence per trial (right).



- 34 healthy adults: 17 younger, 17 older
- Right hand only, five digits D1–D5
- 20 joint-angle channels at 100 Hz
- ~10,180 trials in total
- Windows of 150 samples, stride 75 (1.5 s, 0.75 s overlap)
- 5-fold subject-wise Cross validation

Tasks: Clothes-Peg · Fingertip-Touch · Rubik CW / CCW · Small Rubik CW / CCW

The Physio-Digital Temporal Graph

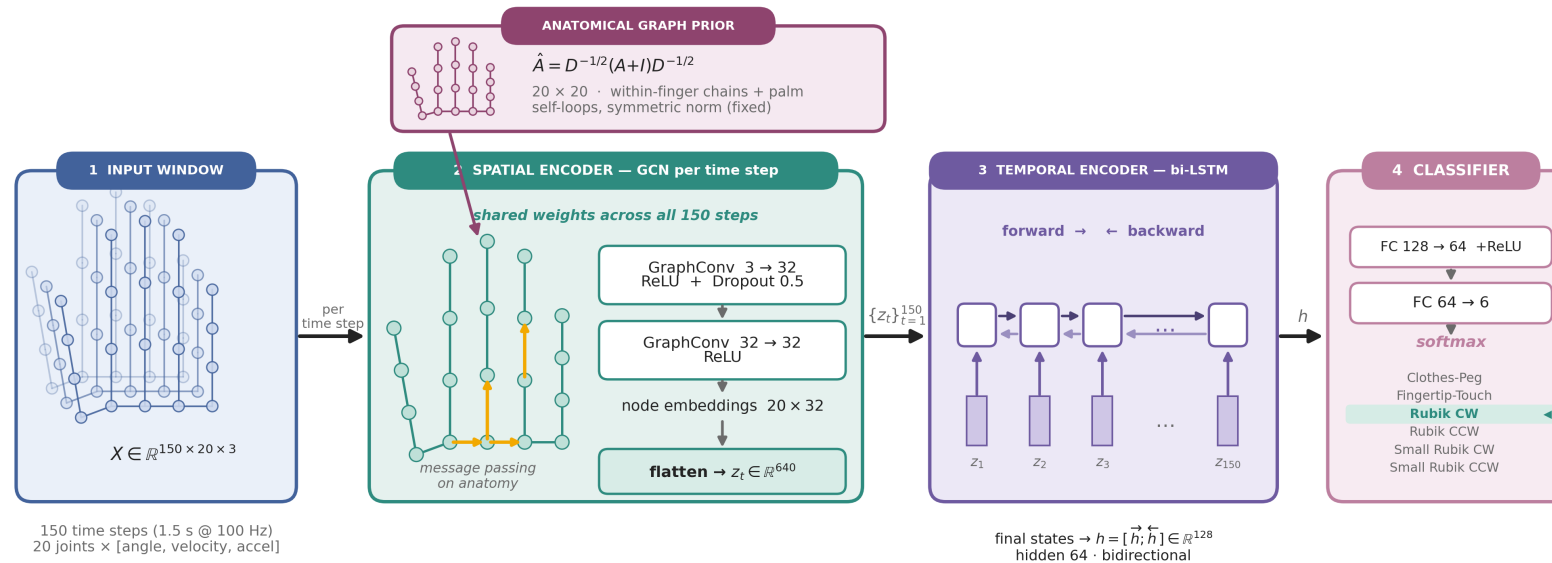


- Nodes: the 20 joint channels, $J\{0..3\}D\{1..5\}$
- Edges: joints within a finger form a chain
- Base joints are linked across the palm
- Per-node features: angle + velocity + acceleration

Physical to digital: 20 sensed joints map onto a fixed anatomical hand graph $J\{0..3\}D\{1..5\}$.

A second, data-driven layer: correlation edges ($|r| \geq 0.6$) used for interpretation only, never for training.

The TGNN: graph convolutions in space, bi-LSTM in time



1 Input window

150 time steps $\times$ 20 joints $\times$ 3 features

2 Spatial encoder

GCN shared across all 150 steps, anatomy as prior

3 Temporal encoder

bi-LSTM over the sequence of graph embeddings

4 Classifier

MLP head $\rightarrow$ 6-way softmax per window

HOW EXPLANATION WORKS

1 Attribute

Integrated gradients (Captum) on each window, against a zero baseline in standardised feature space.

2 Aggregate

Sum the absolute attributions over time and over angle, velocity and acceleration → one score per joint.

3 Overlay

Add correlation edges where $|r| \geq 0.6$ to show which joints are moving together in this task.

4 Read

Which joint, and which coordination chain, drove this classification in clinical language.

92.6% macro-F1 across unseen subjects

92.6%

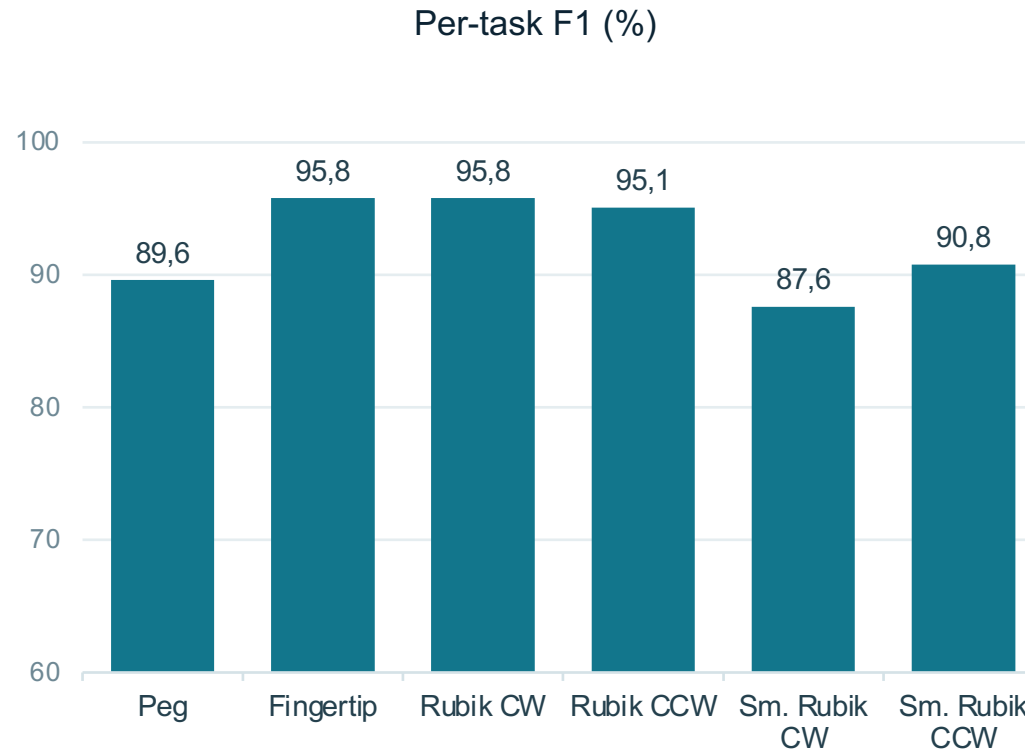
macro-F1, mean over 5 subject-wise folds (± 2.3)

92.4%

accuracy (± 2.5); precision 93.0, recall 92.5

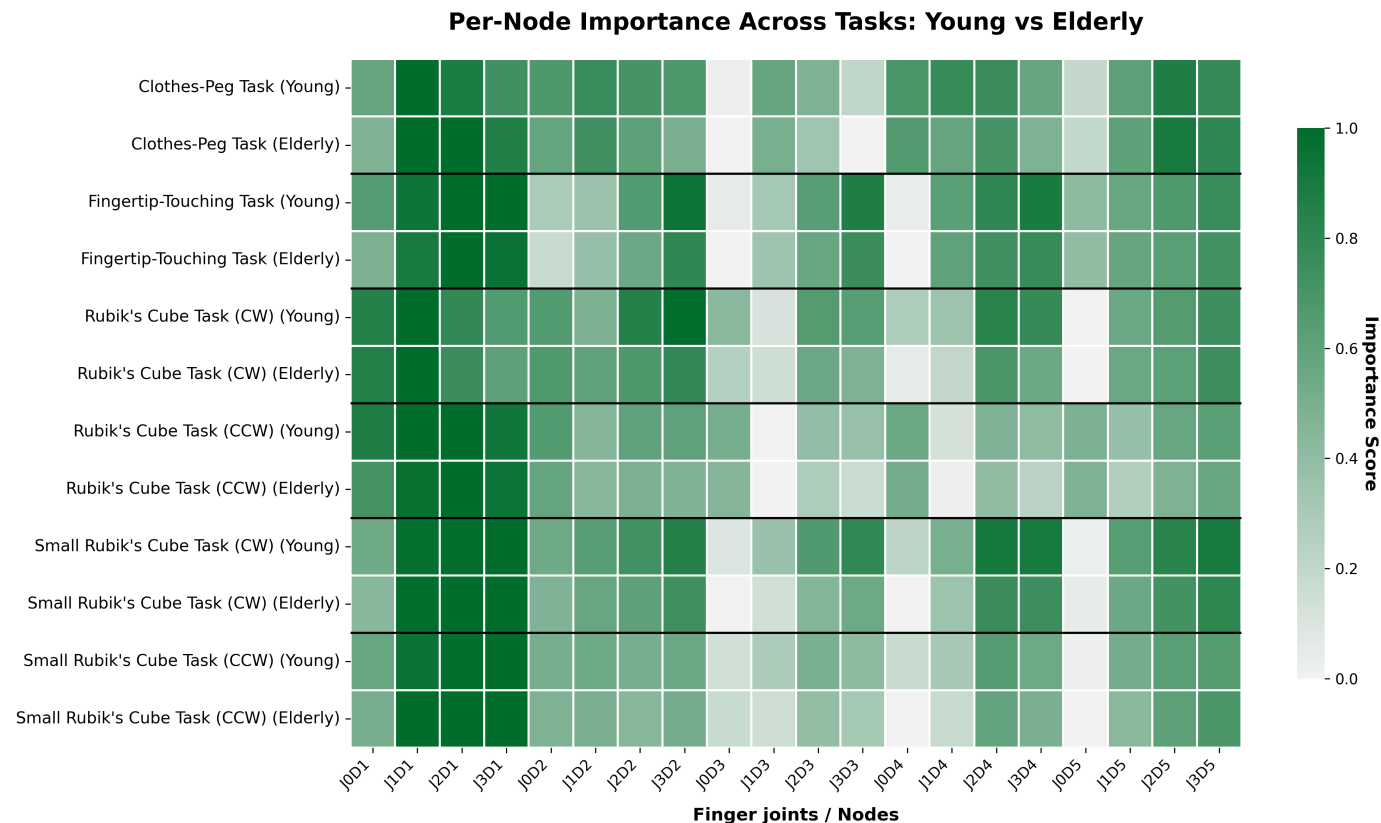
< 2.8

percentage points std on every reported metric



- Fingertip-Touch and the large cube are easiest ($> 95\%$)
- Small Rubik CW is hardest at 87.6%
- Confusions follow manipulation similarity
- Peg $\leftrightarrow$ Small Rubik: both are sustained pinch-and-hold
- Rubik CW $\leftrightarrow$ CCW: same object, mirrored direction

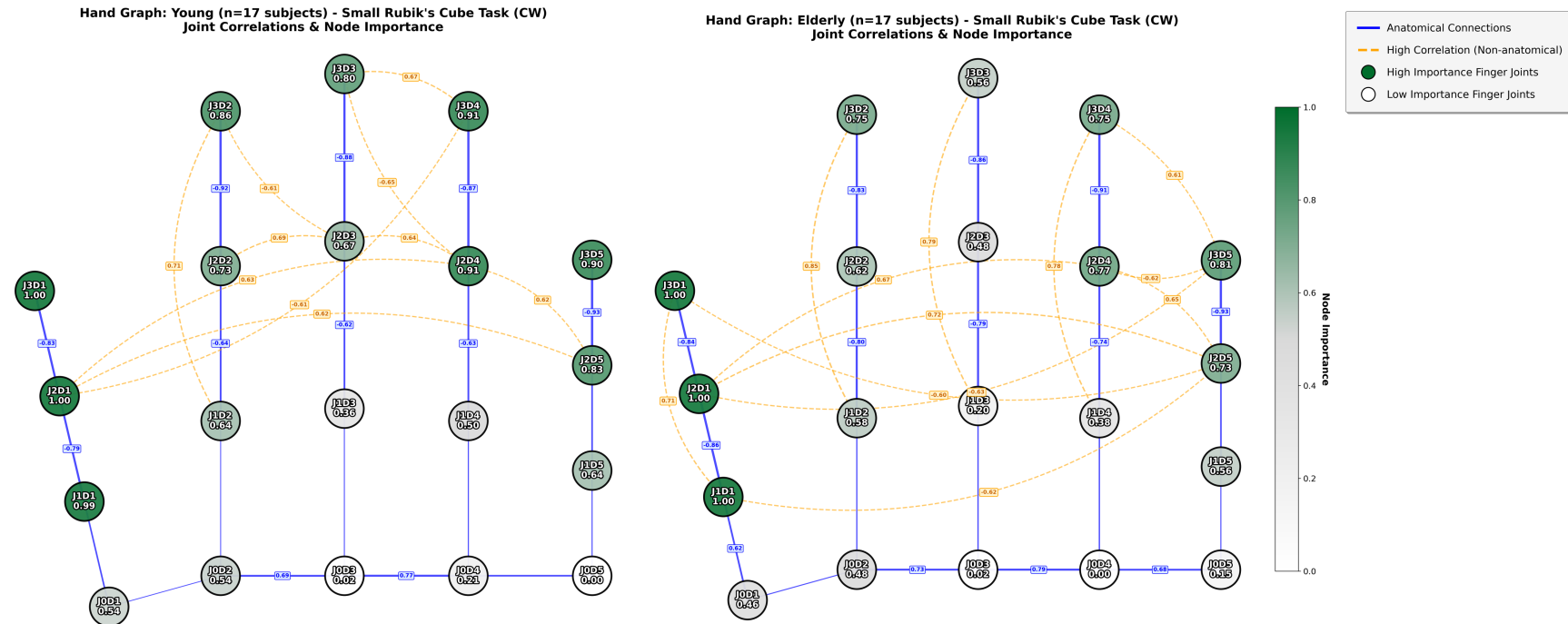
Which joints carry the evidence?



Per-node integrated-gradients importance, one row per task and age group.

- The thumb chain (D1) dominates in every task and both groups
- Consistent with its central role in precision grip
- Beyond the thumb, the profiles diverge by age
- Younger adults: importance spread across more joints
- Older adults: importance concentrated on fewer joints

Two age groups, two coordination strategies



Task 5 (Small Rubik CW). Left: younger adults. Right: older adults. Solid edges = anatomy; dashed = $|r| \geq 0.6$; node colour = joint importance.

Younger adults

Broad inter-finger coordination, strong non-anatomical thumb-to-ring couplings (J0D1–J2D4, J0D1–J3D4).

Older adults

Concentrated within-finger coupling, centred on D4/D5 (J1D2–J3D2, J1D3–J3D3, J2D4–J2D5).

From localization to explanation

What worked

- Anatomy + time → 92.6% macro-F1 across unseen subjects
- Explanations expressed in joints, not channel indices
- Group-level comparison of coordination becomes possible
- The PDTG transfers: same recipe as the chest, new anatomy

What still need

- Healthy volunteers only - no clinical population yet
- The graph is fixed: coordination cannot evolve over time
- Attribution ranks joints; it never says what would change the answer
- An importance score is not contrastive, and clinicians think contrastively

“Which joint mattered?” is answerable. “What would this hand have had to do for the model to call it something else?” is not.

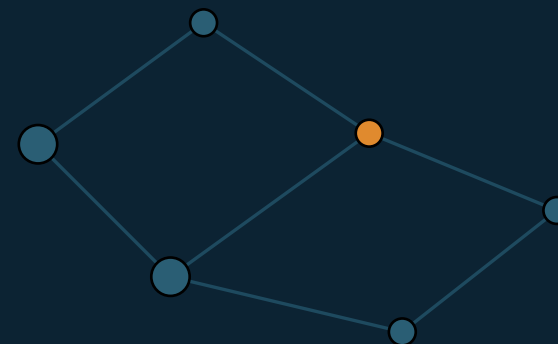
CASE STUDY 3 · XKDD @ ECML PKDD 2026

Counterfactual transition graphs

Stop asking what flips one window. Ask how the classifier connects its own classes to each other.

Syed M. H. Zaidi, Szymon Bobek, Grzegorz J. Nalepa, Myra Spiliopoulou

Jagiellonian University Kraków · CAISR, Halmstad University



From “which joint” to “what would have to change”

Attribution says

“The thumb chain drove this decision.”

A ranking. Useful, but not actionable.

A counterfactual says

“Had the ring finger flexed 8° less, this would have been a peg.”

Contrastive, and in the units of the sensor.

- A counterfactual is the smallest edit that flips the prediction
- The edit must stay plausible: a hand-movement window must still look like real motion
- Standard practice evaluates one instance at a time, abnormal to normal and vice versa.
- Local -> global. How does the model connect its own classes to each other?
- So: what does the whole set of class-to-class edits say about the model?

A local question, asked globally

The usual question: *what minimal edit flips this single window's prediction?*

Our question: *how does the classifier connect its own classes to each other?*

RQ1

Plausibility

How can we capture the plausibility of a counterfactual, independently of whether it flipped the label?

RQ2

Class similarity

How can we link that plausibility to the similarity between the source class and the target class?

RQ3

Validation

How can we validate counterfactual plausibility in a controlled way, across explainers?

Building the counterfactual transition graph

1 Prototypes

Per class and fold, embed the training windows and keep the $m = 3$ prototypes closest to the class centroid.

2 Transition attempt

Generator core edits prototype p_a towards class B under one scan of its solution space.

3 Two measures

Validity $\in \{0,1\}$: did the label flip?
Plausibility: inverse Mahalanobis distance to the nearest target prototype independent of validity.

4 Edge weight

CP-proximity $w(A,B) = \text{mean validity over folds, prototype pairs and strategies.}$

Low Class-Prototype-Proximity = class boundary is very hard for in-distribution counterfactuals to cross

Four generator cores, one framework

EXPERIMENTAL SCALE

75

binary models: 15 class pairs × 5 folds

0.97

mean validation accuracy of those models

13,500

transition attempts across the graph

NEPO

GRADIENT

Optimises the window towards the target class with small, sparse, smooth edits guided by attributions. Borrows nothing from other examples.

Residual

GRADIENT

Splits the signal into a shared component and a class-carrying residual, then attenuates the residual.

GACP

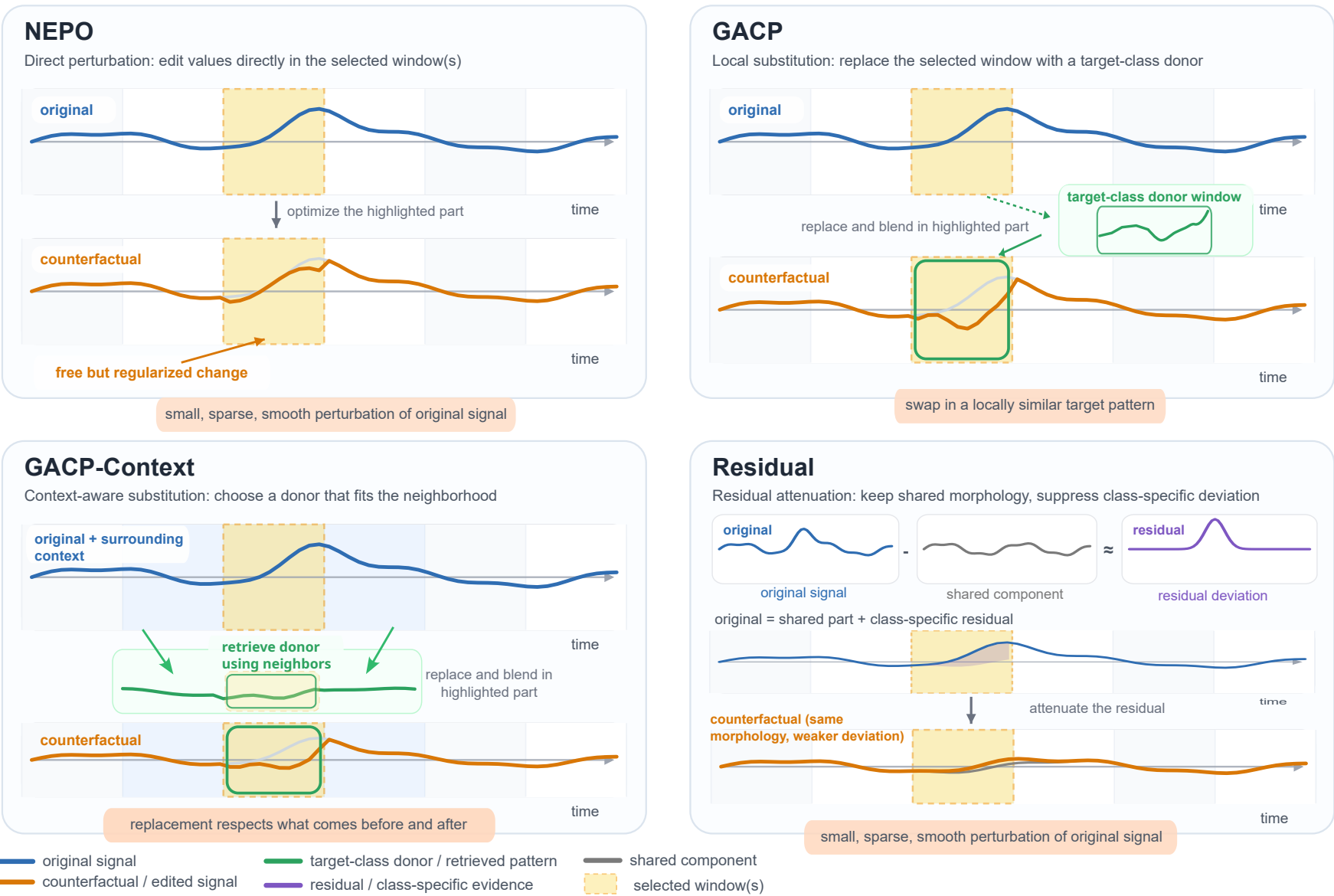
REPLACEMENT

Substitutes the window with a nearest unlike neighbour of the target class, blended back with graph-aware smoothing.

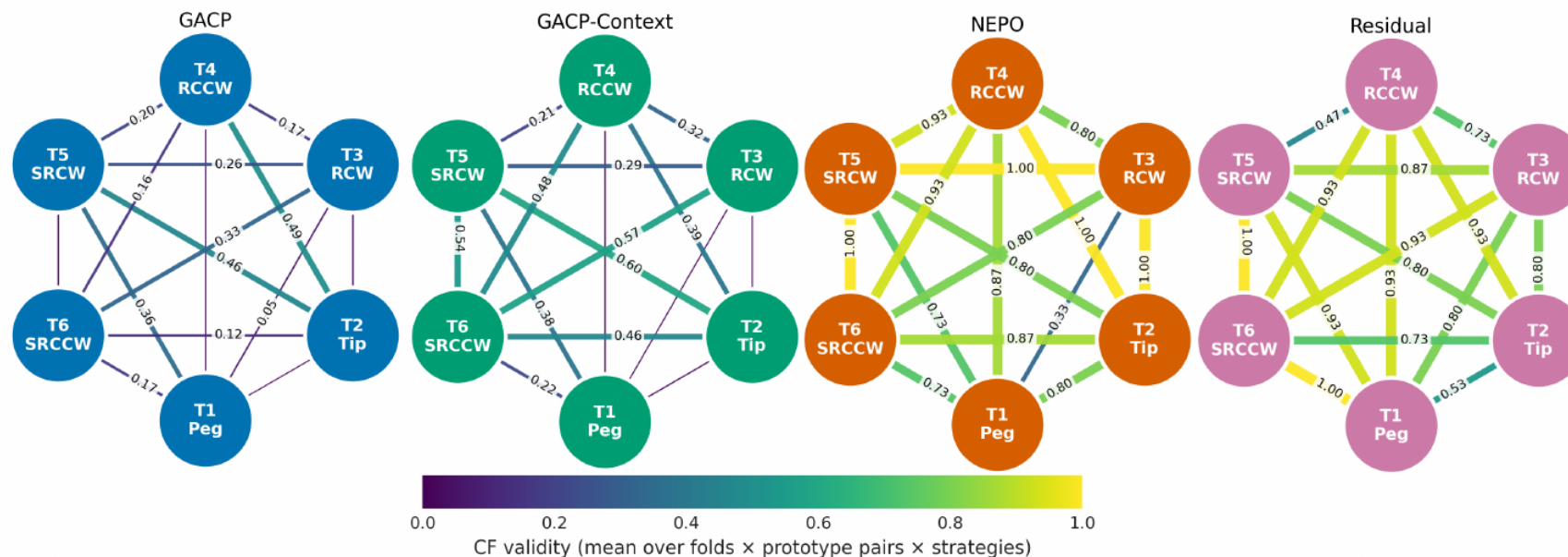
GACP-Context

REPLACEMENT

Same idea, but retrieval matches the surrounding neighbourhood rather than the window alone.



The same classifier, four different topologies

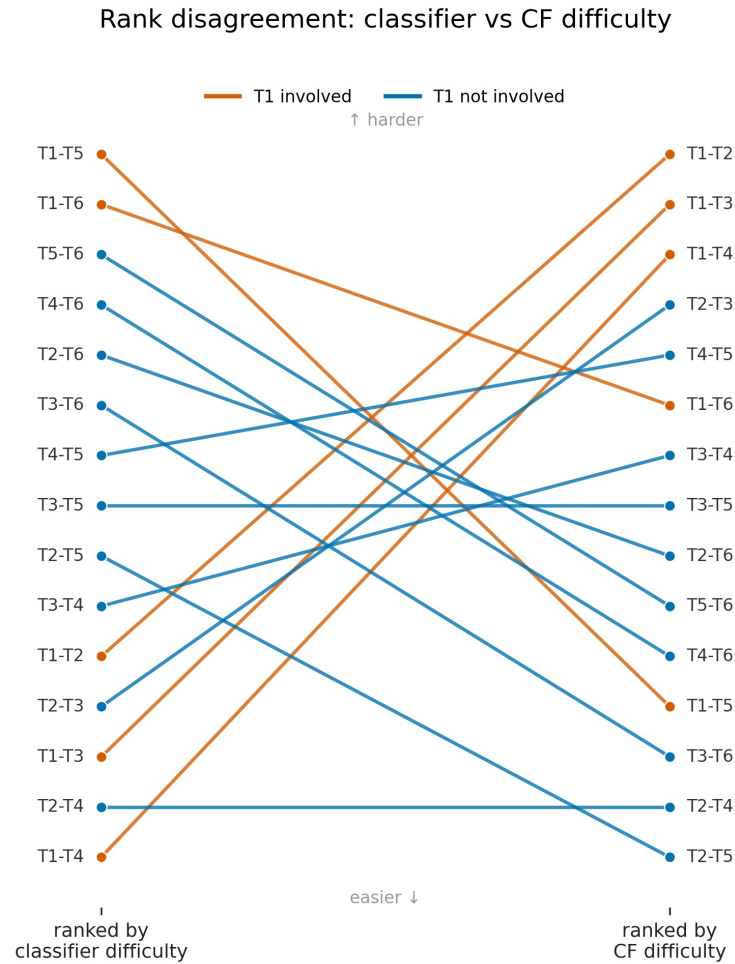


Nodes are classes; edge colour and thickness give the mean validity of that boundary.

Replacement cores (left) recover a sparse, structured topology and leave every Clothes-Peg boundary dark.

Gradient cores (right) flood almost every boundary, they reach any class by leaving the data manifold.

Reachability does not follow accuracy

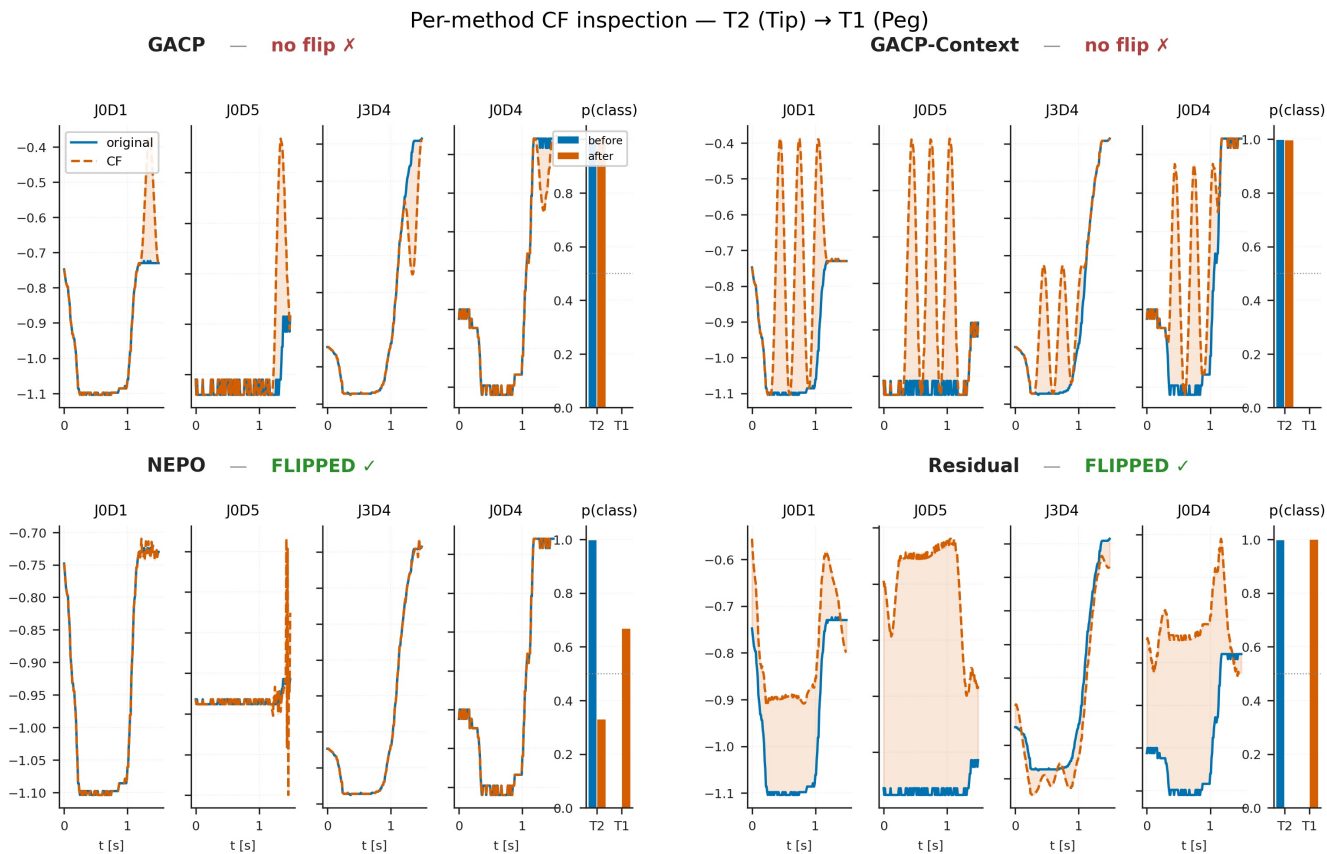


$$\rho = -0.37$$

Spearman correlation between classifier difficulty and CF difficulty

- T1–T3 and T1–T4: separated at accuracy ≥ 0.988 , yet an in-distribution edit crosses almost never.
- T1–T5 is the mirror case: hardest to separate, easy to flip

Valid is not the same as plausible



Fingertip-Touch (T2) → Clothes-Peg (T1) on four representative joints. Blue: original. Orange: counterfactual.

- Replacement cores stay on the manifold and simply fail: 0.19 and 0.30 mean validity
- GACP-Context invents oscillations no finger joint can perform
- Gradient cores flip almost everything: 0.84 and 0.83
- But on rigid pairs they only succeed by stepping off the manifold

Results

What the graph buys us

- High classifier accuracy does not imply good counterfactuals
- Class topology is visible, and it is not the confusion matrix
- Conclusions about the classifier separate from artefacts of the explainer
- A benchmark instrument for comparing CF explainers

Limitations

- One dataset so far.
- Fixed hyperparameters: $m = 3$ prototypes, one distance metric
- 15 pairs is too few for statistical significance
- Undirected edges average two directions that may differ (see next section)

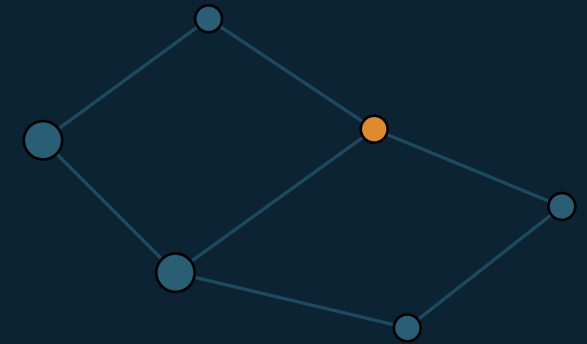
The boundaries a classifier separates most confidently can be exactly the ones no realistic edit can cross.

ONGOING WORK

Where this goes next

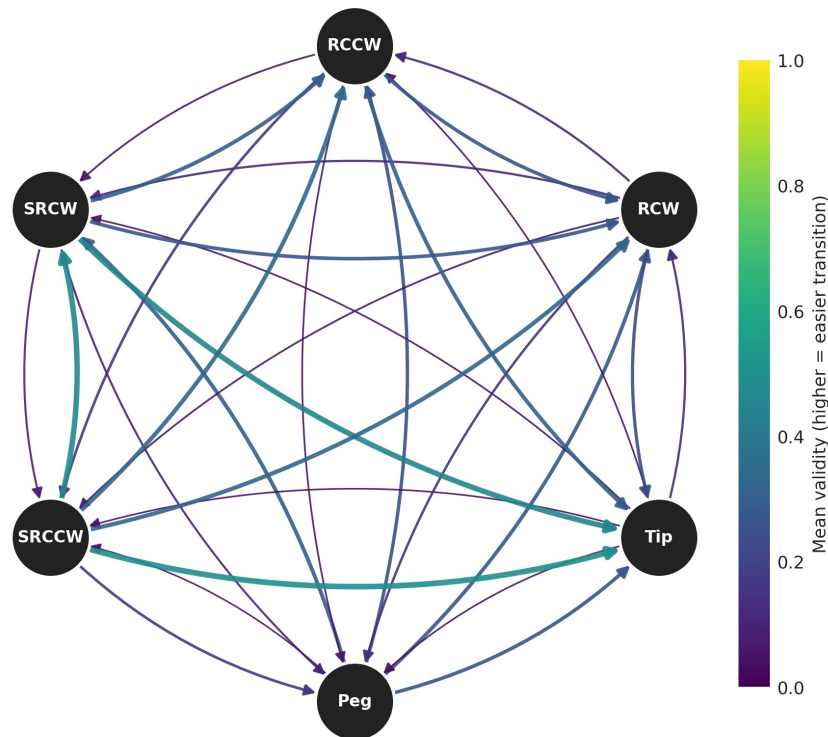
Two threads in flight: making the transition graph directed, and taking the whole approach to a real patient data.

- Directed counterfactual transition graphs (in progress)
- Graph transformers on an ALS hand-tracking (in progress)



Transitions have a direction

Directed counterfactual transition graph: all generators
(one arrow per direction; thicker = easier)

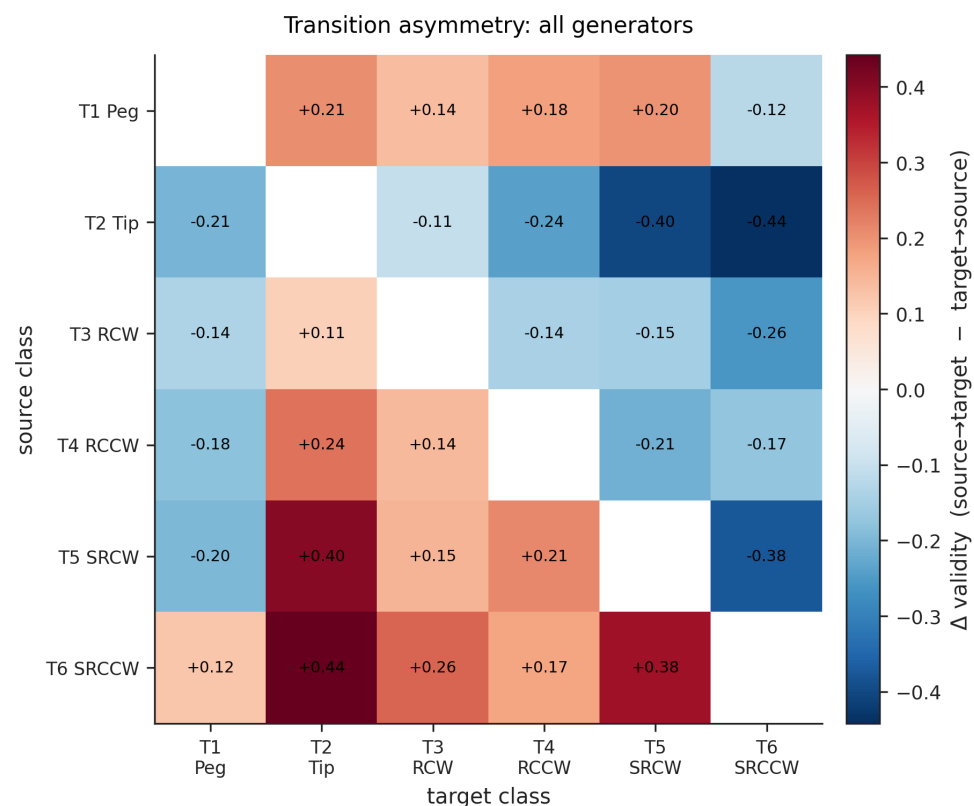


Clothes-Peg is easy to leave and hard to enter

T1 boundary	into T1	out of T1	ratio
T1 ↔ T2 Fingertip-Touch	0.050	0.257	5.1×
T1 ↔ T3 Rubik CW	0.133	0.271	2.0×
T1 ↔ T4 Rubik CCW	0.060	0.240	4.0×
T1 ↔ T5 Small Rubik CW	0.096	0.293	3.1×
T1 ↔ T6 Small Rubik CCW	0.161	0.038	0.2×

- The old pipeline measured only one direction
- Rebuilt: both directions
- The undirected reading averaged away an effect as large as the signal it reported

Classes are attractors or sources, not just “near” or “far”



Δ validity per ordered pair: red = the target pulls, blue = the target repels.

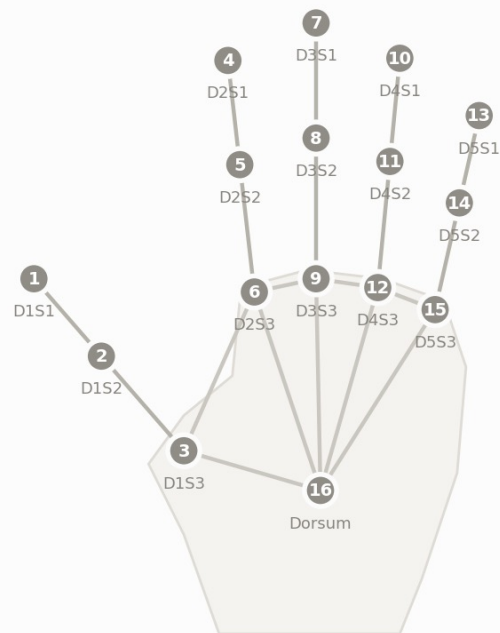
Class	out	in	attractor index
T2 Fingertip-Touch	0.065	0.344	+0.280
T3 Rubik CW	0.130	0.246	+0.116
T4 Rubik CCW	0.169	0.206	+0.037
T5 Small-Rubik CW	0.239	0.201	-0.038
T1 Clothes-Peg	0.220	0.100	-0.120
T6 Small-Rubik CCW	0.355	0.081	-0.274

Fingertip-Touch draws counterfactuals in and rarely lets them out (+0.28).
Small-Rubik CCW does the opposite (-0.27).

Second new edge weight: prototype distance. T2 $\rightarrow$ T1 and T3 $\rightarrow$ T1 are the closest pairs in prototype space (0.19, 0.18) yet have the lowest validity. Close prototypes, rigid boundary.

Hand graph: a real ALS patient data

Sensor montage (right-hand convention)



D1..D5 = thumb..little S1 = distal, S3 = proximal
node 16 = dorsum reference sensor

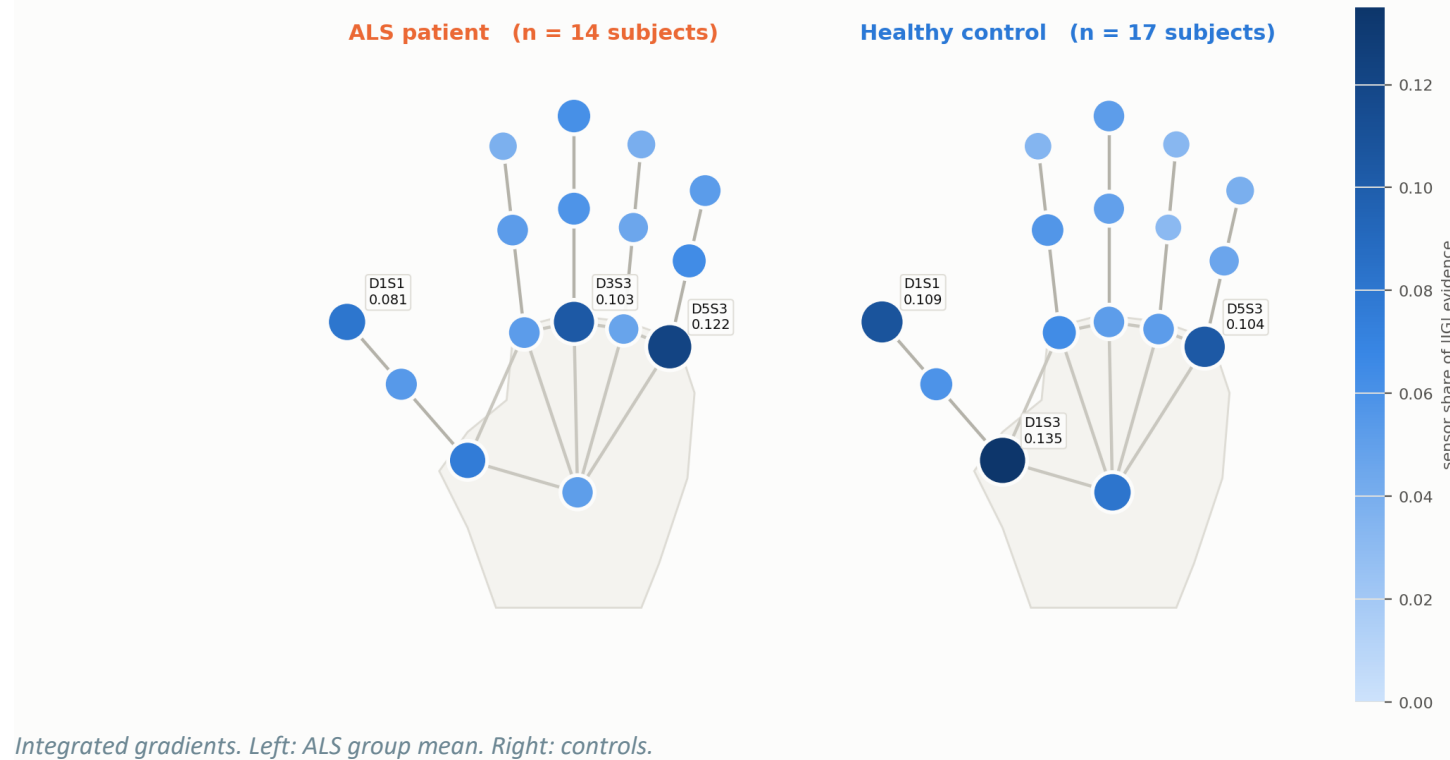
- ALS attacks fine hand motor control early
- 32 subjects: 15 ALS patients, 17 Healthy controls
- 16 sensors per hand, four instructed movements
- 16 electromagnetic sensors
- Graph transformer: anatomy enters as a learned bias on attention
- Leave-one-subject-out across all 32 folds

81.2%
subject accuracy

26 / 32
subjects correct

Where the evidence lives

Sensor contribution to the diagnosis - Finger spreading



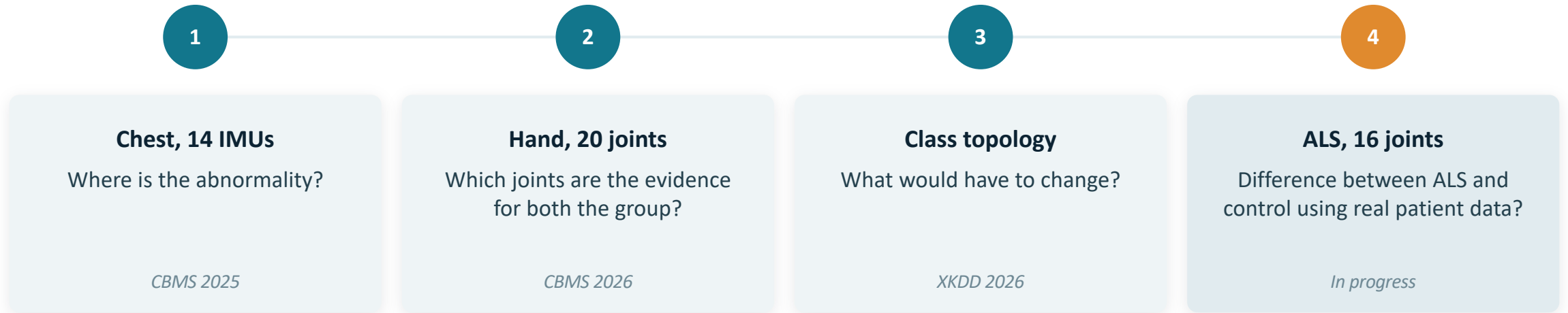
D3S3

middle-finger contributing more for ALS patients

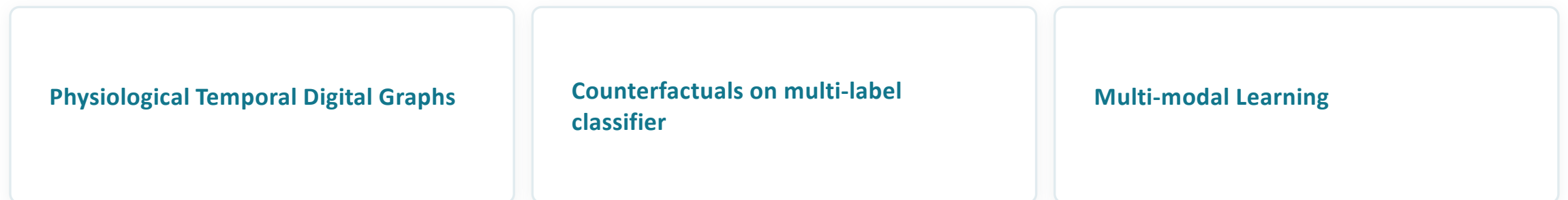
D1S1

thumb is contributing more for Healthy controls

One representation, four questions, and where I want help



I WOULD LIKE TO COLABORATE HERE



Thank you

Questions, and especially the awkward ones, are very welcome.

Collaborators

Myra Spiliopoulou, Benjamin Noack (OVGU) · Lisa Klemm, Christoph Reichert, Elena Azañón (Leibniz Institute for Neurobiology, Magdeburg) · Szymon Bobek, Grzegorz J. Nalepa (Jagiellonian University / CAISR Halmstad)

The work of the author is financed by the graduate school TACTIC: “Towards Co-Evolution in Human-Technology Interfaces”, co-funded by the Federal State of Saxony-Anhalt and the EU (ESF+), Funding Nr. ZS/2023/12/182067. Case study 3 was additionally supported by the DigiWorld Priority Research Area at Jagiellonian University and by Horizon Europe Grant Agreement 101120406.

