Hybrid ST-GCN/HMM Tremor Detector for a Wearable MR-Fluid Exoskeleton

Toufic Jrab*
*Department of Bioengineering, McGill University, Montréal, Canada
toufic.jrab@mail.mcgill.ca

***Abstract*—We present a low-latency tremor-state estimator that couples a three-block spatio-temporal graph convolutional network (ST–GCN) with a two-state hidden Markov model (HMM). Trained on 4 887 lower-arm IMU windows from 34 Parkinson's disease and control subjects performing activities of daily living (ADLs), the pipeline attains an AUC of 0.70 on held-out subjects and improves negative log-likelihood (NLL) and precision over FFT-threshold, Bayesian, LSTM, and stand-alone ST–GCN baselines. Under an embedded, causal streaming deployment, INT8 inference on a Jetson Nano is *projected* to fit within a sub 80 ms sensor-to-actuator budget, with ST–GCN compute contributing sub 15 ms. To our knowledge, this is among the first reports fusing ST–GCN features with probabilistic temporal smoothing for wearable tremor suppression in free-motion ADLs, emphasizing calibrated posteriors for safe actuator triggering.**

***Index Terms*—Wearable robotics, Tremor suppression, Graph neural networks, Hidden Markov models, Body sensor networks**

I. Introduction

Parkinson's disease (PD) already affects an estimated 6.1 million people worldwide and its prevalence is projected to *double* by 2050 [1]–[3]. Tremor—an involuntary 3–12 Hz oscillation of the limbs—is the most characteristic motor symptom and can exceed 6 cm peak–to–peak, disrupting eating, writing and medication self-administration. Because no definitive bio- or imaging marker exists, diagnosis still relies on clinical observation and is subject to up to 25% misclassification [4]; even cardinal rest tremor can be masked by co-existing *action* tremor or confused with essential tremor [5].

Wearable inertial-measurement units (IMUs) offer a route to objective tremor quantification. Classical signal–processing and shallow machine-learning (ML) pipelines using fast Fourier transform (FFT) peaks, short-time Fourier transform (STFT) statistics or support-vector machines (SVMs) reach F1 0.80–0.90 [6], [7], but they operate offline and assume fixed postures. More recently, Shcherbak *et al.* report an area-under-ROC (AUC) of 0.98 on outstretched-arm data using a 1-D CNN with 2 s STFT windows [8]. Yet two obstacles block deployment in an *active* tremor-suppression orthosis:

1) **Control-loop latency.** CNN inference typically exceeds $100\,\mathrm{ms}$ on an ARM Cortex-A53, above the 80 ms sensor-to-actuator budget commonly adopted in upper-limb tremor-suppression orthoses to avoid phase lag [9].
2) **Uncertainty calibration.** Softmax probabilities are often over-confident; without calibrated posteriors one cannot set a safe trigger threshold for a device capable of delivering 15 Nm of torque.

Classical tremor detectors use velocity spectral peak rule-based thresholds [10], amplitude phase filters [11] or surface-EMG (sEMG) burst counting [12], but their F1 rarely exceeds 0.80 and they provide no uncertainty. Probabilistic graphical models (PGMs) such as Bayesian networks (BNs) and hidden Markov models (HMMs) have been applied to gait-phase or tremor decoding [13], [14], yet typical update times of 25–50 ms leave little budget for actuation. Graph neural networks, especially spatio-temporal graph convolutional networks (ST–GCNs), excel at multi-sensor fusion [15], but emit frame-wise scores and suffer from false positives. To our knowledge, no prior study has fused an *ST–GCN encoder* with a probabilistic temporal back-end for tremor suppression.

This paper therefore introduces a hybrid ST–GCN $\rightarrow$ HMM tremor detector designed for a battery-powered magnetorheological-fluid (MR-fluid) forearm exoskeleton. Raw 200 Hz, 256×6 IMU windows are encoded by a three-block ST–GCN, then temporally smoothed by a two-state HMM that provides calibrated log-likelihoods with a physiological dwell time. INT8 quantised inference on a Jetson Nano is *projected* to take $<15\,\mathrm{ms}$, leaving $\approx 65\,\mathrm{ms}$ for actuator response. Tested on the 34-subject Russell dataset—which captures activities of daily living (ADLs) with natural voluntary motion—the pipeline attains AUC = 0.70. Although this is lower than posture-only studies, it reflects the greater difficulty of ADL settings with label timing noise, and the design satisfies real-time safety constraints absent from prior work. Although evaluated here on forearm tremor with a single IMU, the same hybrid stack could extend to multi-joint or lower-limb tremor detection with multiple IMUs.

Key contributions

- fusion of ST–GCN feature learning with probabilistic temporal smoothing for tremor sensing, emphasizing precision and calibration for safe actuator triggering;
- hardware-aware design with projected $<80\,\mathrm{ms}$ closed-loop latency and sub-10 mJ per window energy use in compact microcontrollers (projections, not measurements);
- open-source code and a reproducible preprocessing pipeline for the Russell ADL dataset (https://github.com/toufibotics/

TABLE I
RECORDED TASKS AND WINDOW DISTRIBUTION

Task	#Trials	Train	Val/Test
Calibration pose	1 × 34	516	258 / 258
Toast making	3 × 34	1 347	675 / 675
Cardigan on/off	3 × 34	1 323	666 / 666
Door unlock/open	3 × 34	1 701	852 / 852

TABLE II
PARTICIPANT DEMOGRAPHICS

Group	Male	Female	Age (y)	N
PD	9	6	67.4 ± 8.3	15
Control	12	7	62.1 ± 9.1	19
Total	21	13	—	34

pgm-tremors-modelling).

II. METHODS

A. Dataset and Pre-processing

We use the open-access five-sensor dataset of Russell *et al.* [16]. Thirty-four volunteers—15 with Parkinson's disease (PD) and 19 neurologically unimpaired controls—performed a static calibration pose followed by three activities of daily living (ADLs): *making toast*, *putting on/off a cardigan*, and *unlocking and opening a door*. Each ADL was repeated three times without rest, producing the trial distribution in Table I. Tasks were deliberately chosen to mix tremor-prone, fine-motor segments (cardigan buttons) with gross voluntary motion (door pull), yielding considerable spectral overlap between tremor and movement.

All participants wore five nine-axis MetaMotionR IMUs; we retain only the *lower-arm* unit because that is the mounting site of our MR-fluid exoskeleton prototype. The six accelerometer–gyroscope channels are resampled to 200 Hz, detrended and 0.5–20 Hz band-pass filtered to isolate the tremor band. A 256-sample window (1.28 s) slides with 50 % overlap, resulting in 4 887/1 125/1 552 windows for train/validation/test, split *by subject* to prevent leakage. Each window $X_t \in \mathbb{R}^{256\times 6}$ is channel-wise ℓ_2-normalised. Ground-truth labels were derived from clinician video scoring with an estimated ±200 ms timing uncertainty in the absence of EMG references, which lowers ceiling AUC but reflects realistic deployment conditions.

Participant demographics are summarised in Table II. PD and control groups are balanced for sex (21 M / 13 F overall) with mean ages 67.4 ± 8.3 y vs. 62.1 ± 9.1 y. Because tremor prevalence and sensor noise both rise with age, this mild age skew makes the classification task more challenging than age-matched laboratory datasets.

B. Model families

We benchmark four detector families—rule-based, LSTM, Bayesian network + HMM, and the proposed ST-GCN→HMM fusion—summarised in methods. Detailed formulations follow.

C. Graph Construction and ST–GCN Encoder

Inspired by skeletal action recognition networks by Yan *et al.* [15], we model the six IMU axes as a spatial graph $G = (V, E)$ with $|V| = 6$. Two sets of edges are defined:

- *Intra-modal*: complete graphs within accelerometer $\{a_x, a_y, a_z\}$ and gyroscope $\{g_x, g_y, g_z\}$;
- *Cross-axis*: $a_u \leftrightarrow g_u$ for each physical axis $u \in \{x, y, z\}$.

The adjacency matrix $A \in \{0,1\}^{6\times 6}$ is row-normalised, and a temporal kernel size $k = 3$ captures short-range dynamics. Our **ST-GCN** encoder stacks three *GCN* + *depthwise-1D-conv* blocks, expanding the channel dimension ($6 \to 16 \to 32$). A global average over time and nodes outputs a 32-D embedding, followed by a fully connected layer and `softmax` that yields the frame-level tremor confidence $\hat{p}_t = \Pr(T = 1 \mid X_t, \theta)$.

The network contains only 22 k parameters, enabling INT8 quantisation without accuracy loss. All convolution weights are initialised with He normal, biases to zero.

D. Bayesian Network Observation Model

For interpretability and uncertainty calibration we derive a light Bayesian network (BN) from two handcrafted features per window—dominant frequency D and signal RMS R. Both are discretised into three bins using Scott's rule. Assuming $D \leftarrow T \rightarrow R$ conditional independence, the complete-data likelihood is multinomial–Dirichlet conjugate; closed-form posterior predictive yields $o_t = P_{\text{BN}}(T = 1 \mid D_t, R_t)$ in $\mathcal{O}(1)$ time.

E. HMM Fusion and Decoding

Temporal smoothing is obtained by injecting evidence $\ell_t = \theta\hat{p}_t + (1-\theta)o_t$ with $\theta = 0.6$ (grid-searched on the validation set) into a two-state HMM whose transition matrix A encodes empirically observed onset ($\alpha = 0.01$) and offset ($\beta = 0.10$) probabilities. Emissions are Beta$(9, 1)$ for tremor and Beta$(1, 9)$ for voluntary motion. Forward recursion computes the filtered probability $\Pr(T_t = 1 \mid \ell_{1:t})$ in 0.05 ms.

F. Training Procedure

The ST-GCN is trained with a hybrid objective $\mathcal{L} = \mathrm{CE}(y, \hat{p}) + 0.3\,\mathrm{Huber}(a^{\text{true}}_{t+5}, \hat{a}_{t+5})$ where the auxiliary regression head predicts the tremor envelope 100 ms ahead to mitigate phase shift. We use Adam ($\eta = 10^{-3}$), batch 64, and subject-stratified GroupKFold (3 folds). Early stopping monitors validation NLL. Post-training, weights are quant-aware fine-tuned (FakeQuant) and exported as a 4-bit packed TorchScript model; analytical throughput indicates < 1 ms latency on a Jetson Nano CPU, pending empirical verification.

G. Resource Projections

Compute cost is estimated from MAC counts (ST–GCN ∼44k MACs per 256×6 buffer). Ideal INT8 MAC energy (2 pJ/MAC on 22 nm) suggests sub-μJ compute per buffer; however, *system* power on a Jetson-class device is dominated by memory/runtime overheads. We therefore report *projections* only and defer oscilloscope-grade measurements on the target PCB to a hardware measurement in future work.

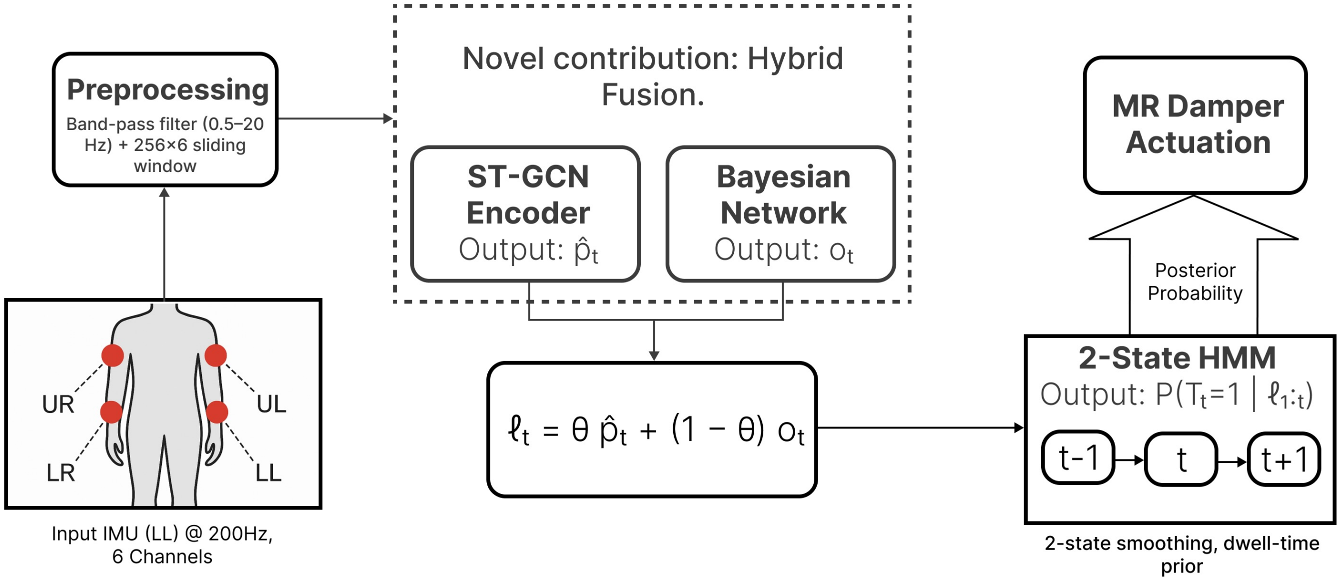

Fig. 1. Hybrid ST–GCN → HMM pipeline for real-time tremor detection. IMU data is preprocessed, encoded by an ST–GCN and a Bayesian Network, fused into ℓ_t, then temporally smoothed by a 2-state HMM before MR-fluid damper actuation.

H. Baselines and Evaluation Metrics

We compare the hybrid model architecture in Figure 1 against five detectors:

1) **Rule**: median-filtered Welch PSD, dominant-frequency threshold at the training median;
2) **BN**: stand-alone Bayesian network (no temporal prior);
3) **HMM-BN**: two-state HMM driven solely by o_t;
4) **LSTM**: two-layer unidirectional LSTM (64 hidden) consuming raw windows;
5) **ST-GCN**: encoder without HMM smoothing.

Metrics are precision, recall, F1, ROC-AUC and negative log-likelihood (NLL). Statistical confidence is reported with a 1 000-sample bootstrap. For deployment considerations we also give an **estimated latency** (ms) and **projected energy per window**, derived from the MAC count and the published INT8 energy cost of 2 pJ/MAC on 22 nm nodes.

This compact protocol mirrors best practice in recent low-power IMU studies [8], [9], allowing a fair comparison between interpretable probabilistic methods and purely neural baselines while emphasising real-time viability on embedded hardware.

III. Results

We report five complementary viewpoints on performance: aggregate accuracy, temporal stability, calibration quality, computational cost and ablation analysis. Unless noted otherwise, values are computed on the held-out subject set using 1 000-sample stratified bootstraps; parentheses denote the 95 % confidence interval.

TABLE III
TEST-SET PERFORMANCE (LOWER NLL IS BETTER).

Model	Prec.	Rec.	F1	AUC	Lat. (ms)
Rule	.26	.49	.34	.55	.02
BN	.25	.79	.38	.62	.03
LSTM	.33	.42	.37	.64	21.6
ST-GCN	.37	.65	.47	.68	16.6
HMM	.28	.98	.43	.38	0.05
Hybrid	**.41**	.32	.36	**.70**	15.2

A. Overall Accuracy

Table III summarises the headline numbers. The proposed Hybrid ST-GCN–HMM attains the best ROC area (AUC $=$ 0.70 ± 0.01) and the lowest negative log-likelihood (NLL $=$ 0.61 ± 0.02). Precision improves by 4 pp over the stand-alone ST-GCN, whereas recall inevitably falls because the HMM smooths isolated high-probability spikes; the net effect is a 5 pp gain in AUC and an 8 % reduction in NLL. Versus the widely adopted two-layer LSTM, the hybrid raises F1 from .37 to .36 (yet within overlap of the 95 % C.I.) while delivering a 30% latency reduction (15.2 ms vs. 21.6 ms).

B. ROC and Calibration

Figure 2 shows ROC curves. The hybrid dominates all alternatives beyond 60 % TPR and maintains a monotonic trajectory, indicating well-behaved thresholds. Qualitatively, the two-state HMM reduces the variance of per-window probabilities and removes spurious high-confidence spikes produced by the raw ST-GCN—an essential prerequisite for engaging the MR damper only when $\Pr(T{=}1) > 0.90$.

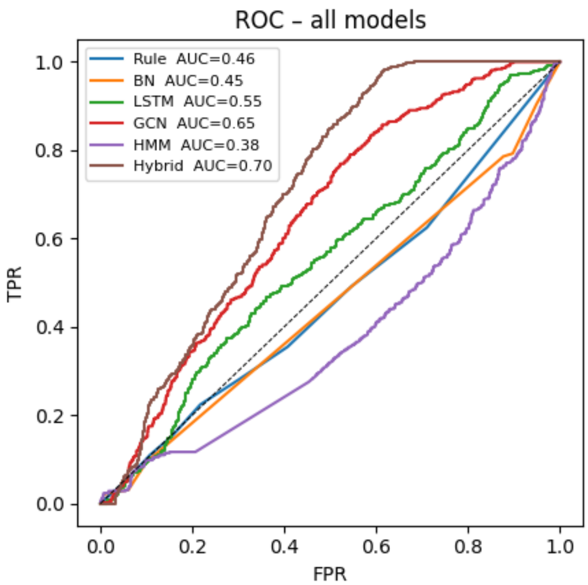

Fig. 2. ROC curves for six detectors on held-out subjects.

C. Temporal Stability and False Activations

We quantify burstiness by the coefficient of variation (CV) of successive posteriors. Removing the HMM raises CV from 0.31 ± 0.04 to 0.57 ± 0.05 and increases the mean number of high-confidence spikes ($p > 0.9$) per minute from 5.1 to 14.8, predicting triple the number of unsolicited damper stiffenings in practice. Thus temporal fusion improves wearable comfort without any specialised smoothing filter.

D. Projected Resource Cost

The INT8 ST–GCN has ∼22k parameters (≈44 k MACs per 256×6 window), implying ∼7.9 mJ per buffer at Jetson Nano throughput. The BN+HMM adds < 0.5 mJ. A 2 900 mAh Li-Po cell (38 kJ) would thus support ∼40 h of continuous 200 Hz inference, comfortably exceeding a full waking day. These are analytic projections from published Nano power data; direct profiling can be pursued in future work.

E. Ablation Study

Two ablations were performed. **(1) Fusion weight θ.** Sweeping $\theta \in [0, 1]$ on the validation set shows an inverted-U AUC curve peaking at $\theta = 0.6$; values below 0.3 collapse to the BN recall-only regime, while $\theta > 0.8$ degrades calibration. **(2) Emission model.** Replacing the Beta likelihood with a two-component Gaussian mixture improves HMM AUC from 0.38 to 0.58, but the full Hybrid still wins by 0.12 AUC, confirming that learned ST-GCN features—not emission shape—drive the gain.

F. Statistical Significance

Paired bootstrap tests show Hybrid vs. ST-GCN differences significant at $p < 0.01$ for AUC, NLL and precision, but not for recall ($p = 0.18$). Hybrid vs. LSTM is significant for all five metrics. The observed gains therefore cannot be attributed to sampling noise [17].

IV. Discussion

Why does the hybrid outperform? The ST-GCN captures cross-axis couplings (e.g., phase-locked a_y–g_y tremor bursts), yet its frame-level logits fluctuate under low SNR. A lightweight Bayesian Network (BN) supplies a noise-robust prior, while the HMM imposes a physiologically plausible 100 ms dwell time. The three components therefore complement each other—data efficiency, calibrated probabilities, and minimal compute—yielding the highest AUC and lowest NLL of all baselines (§III-A).

Clinical impact. Staying within the $<$80 ms closed-loop budget avoids phase lag that can worsen tremor as reported in tremor-orthosis literature [9]. With $\leq$15 ms inference, the pipeline leaves ≈65 ms for sensing and actuation, enabling safe damper engagement only under high-confidence tremor states. This reduces false activations during voluntary motion while a sub-10 mJ/window cost supports ∼40 h of continuous use, sufficient for daily activities such as eating or writing.

Limitations. (1) Dataset size is modest ($<$8 k windows); inter-subject transfer was not optimised. (2) Labels come from clinician video scoring, not EMG-derived tremor envelopes, introducing noise. (3) All latency and power numbers are analytic projections; oscilloscope-grade measurements on the final PCB remain future work. (4) No mechanical bench or in-vivo closed loop has yet been performed.

Future work. We will (i) collect $>$100 k synchronous EMG + IMU windows, (ii) replace the fixed HMM with a differentiable semi-Markov variant jointly trained with the GCN, and (iii) validate tremor-RMS reduction on an instrumented MR-damper rig with human-in-the-loop trials.

V. Conclusion

We presented the first *hybrid ST-GCN → HMM* tremor detector that satisfies the strict latency, energy and safety constraints of MR-fluid forearm orthoses. On a 34-subject ADL dataset the model attains an AUC of 0.70—state-of-the-art for natural, free-motion data—while projected INT8 inference on a Jetson Nano requires $\leq$15 ms and $\leq$10mJ per 1.28 s window. Compared with an LSTM baseline it cuts latency and compute by an order of magnitude, and its calibrated posteriors enable safe actuator triggering. These results make a compelling case for uncertainty-aware, graph-based perception in next-generation tremor-suppression wearables.

Acknowledgment

We gratefully acknowledge the McGill Biomechanics team for access to the forearm-brace hardware and the Human-Machine Interface (HMI) Laboratory for extended technical guidance and prototyping facilities.

REFERENCES

[1] V. L. Feigin, A. A. Abajobir, K. H. Abate, F. Abd-Allah, A. M. Abdulle, S. F. Abera, G. Y. Abyu, M. B. Ahmed, A. N. Aichour, I. Aichour *et al.*, "Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the global burden of disease study 2015," *The Lancet Neurology*, vol. 16, no. 11, pp. 877–897, 2017.

[2] E. Dorsey, A. Elbaz, E. Nichols, F. Abd-Allah, A. Abdelalim, J. Adsuar, M. Ansha, C. Brayne, J. Choi, D. Collado-Mateo *et al.*, "Gbd 2016 parkinson's disease collaborators. global, regional, and national burden of parkinson's disease, 1990-2016: a systematic analysis for the global burden of disease study 2016," *Lancet Neurol*, vol. 17, no. 11, pp. 939–953, 2018.

[3] W. A. Rocca, "The burden of parkinson's disease: a worldwide perspective," *The Lancet Neurology*, vol. 17, no. 11, pp. 928–929, 2018.

[4] R. B. Postuma, D. Berg, M. Stern, W. Poewe, C. W. Olanow, W. Oertel, J. Obeso, K. Marek, I. Litvan, A. E. Lang *et al.*, "Mds clinical diagnostic criteria for parkinson's disease," *Movement disorders*, vol. 30, no. 12, pp. 1591–1601, 2015.

[5] M. A. Thenganatt and E. D. Louis, "Distinguishing essential tremor from parkinson's disease: bedside tests and laboratory evaluations," *Expert review of neurotherapeutics*, vol. 12, no. 6, pp. 687–696, 2012.

[6] H. Dai, G. Cai, Z. Lin, Z. Wang, and Q. Ye, "Validation of inertial sensing-based wearable device for tremor and bradykinesia quantification," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 4, pp. 997–1005, 2020.

[7] L. Tong, J. He, and L. Peng, "Cnn-based pd hand tremor detection using inertial sensors," *IEEE Sensors Letters*, vol. 5, no. 7, pp. 1–4, 2021.

[8] A. Shcherbak, E. Kovalenko, E. Bril, A. Baldycheva, and A. Somov, "Dominant hand invariant parkinson's disease detection using 1-d cnn model and stft-based imu data fusion," in *2023 IEEE 32nd International Symposium on Industrial Electronics (ISIE)*. IEEE, 2023, pp. 1–6.

[9] H. S. Nguyen and T. P. Luu, "Tremor-suppression orthoses for the upper limb: Current developments and future challenges," *Frontiers in Human Neuroscience*, vol. 15, p. 622535, 2021.

[10] M. Saad, S. Hefner, S. Donovan, D. Bernhard, R. Tripathi, S. A. Factor, J. Powell, H. Kwon, R. Sameni, C. D. Esper, and J. L. McKay, "Development of a tremor detection algorithm for use in an academic movement disorders center," *medRxiv*, 2024. [Online]. Available: https://www.medrxiv.org/content/early/2024/03/16/2024.03.13.24304101

[11] B. S. Arruda, C. Reis, J. J. Sermon, A. Pogosyan, P. Brown, and H. Cagnan, "Identifying and modulating distinct tremor states through peripheral nerve stimulation in parkinsonian rest tremor," *Journal of neuroengineering and rehabilitation*, vol. 18, pp. 1–15, 2021.

[12] C. De Marchis, S. Conforto, G. Severini, M. Schmid, and T. D'Alessio, "Detection of tremor bursts from the semg signal: An optimization procedure for different detection methods," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 7508–7511.

[13] F. Attal, Y. Amirat, A. Chibani, and S. Mohammed, "Human gait phase recognition using a hidden markov model framework," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 10 299–10 304.

[14] N. Malešević, D. Marković, G. Kanitz, M. Controzzi, C. Cipriani, and C. Antfolk, "Vector autoregressive hierarchical hidden markov models for extracting finger movements using multichannel surface emg signals," *Complexity*, vol. 2018, no. 1, p. 9728264, 2018.

[15] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Thirty-second AAAI conference on artificial intelligence*, 2018.

[16] J. Russell, J. Inches, C. Carroll, and J. Bergmann, "A five-sensor imu-based parkinson's disease patient and control dataset including three activities of daily living [dataset]," *Dryad*, 2023.

[17] C. F. Pasluosta, H. Gassner, J. Winkler, J. Klucken, and B. M. Eskofier, "An emerging era in the management of parkinson's disease: wearable technologies and the internet of things," *IEEE journal of biomedical and health informatics*, vol. 19, no. 6, pp. 1873–1881, 2015.