mmTremor: Practical Tremor Monitoring for Parkinson's Disease and Essential Tremor in Daily Life

Qingyong Hu[1]*, Yuxuan Zhou[1]*, Jinjian Wang[1], Zirui Huang[2], Guihua Li[3], Qianhui Xu[4], Qian Zhang[1]

[1]The Hong Kong University of Science and Technology, [2]The Hong Kong University of Science and Technology (Guangzhou), [3]Jinan University Affiliated Guangdong Second Provincial General Hospital, [4]Shenzhen People's Hospital (Jinan University)

Email: [1]{qhuag, yzhoudo, jwangjx, qianzh}@cse.ust.hk, [2]zhuang285@connect.hkust-gz.edu.cn, [3]guihuali19790302@sina.com, [4]qianhuixu.jnu.edu@hotmail.com

ABSTRACT

Tremor, a prevalent symptom in various neurological disorders, significantly impacts patients' quality of life. Regular and precise tremor monitoring is essential for optimizing treatment effectiveness. Existing at-home solutions have limitations in practical daily life monitoring due to low adherence, privacy concerns, or inevitable interference from confounding body components and overlapped motions. In this paper, we propose mmTremor, the first privacy-preserving contactless system to achieve practical tremor detection during activities of daily living (ADL) in real-world settings. We design a mmWave-depth fusion tracking algorithm to handle the high interference from confounding body components. To combat the impact of activities, we design a multimodal spatiotemporal deep learning pipeline to effectively exploit latent tremor information. Additionally, a spatial contrastive unsupervised adaptation is proposed for better discriminability and adaptation capability to unseen users and environments. To fully assess mmTremor in real-world scenarios, we collect a diverse dataset of 28 patients and 9 healthy subject simulations in more than 20 distinct environments, including offices, hospitals, and homes. Extensive evaluations demonstrate that mmTremor achieves a high macro-F1 of 0.877 in tremor detection, showcasing the potential as a transformative contactless solution for tremor monitoring in daily life. The dataset will be open-source to facilitate future research.

*Both authors contributed equally to this research.

Mobicom'25, Nov 2025, Hong Kong, China

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

KEYWORDS

Smart Healthcare, Tremor Monitoring, Multi-Modality Fusion, Contrastive Adaptation

ACM Reference Format:

Qingyong Hu[1]*, Yuxuan Zhou[1]*, Jinjian Wang[1], Zirui Huang[2], Guihua Li[3], Qianhui Xu[4], Qian Zhang[1]. 2025. mmTremor: Practical Tremor Monitoring for Parkinson's Disease and Essential Tremor in Daily Life. In *Proceedings of The 31st Annual International Conference on Mobile Computing and Networking (Mobicom'25).* ACM, Hong Kong, China, 15 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Tremor is a common symptom of several significant geriatric diseases, particularly Parkinson's Disease (PD) and Essential Tremor (ET). Approximately 24.91 million people are affected by ET [51] and over 8.5 million by PD globally [13]. Tremors significantly degrade patients' quality of life (QoL), as they directly influence daily activities [34]. Since neither ET nor PD can be fully cured, effective disease management is crucial to improving quality of life by suppressing symptoms and adjusting the efficacy of the treatment. Thus, as a key solution for symptom monitoring, tremor detection shows promise for assessing and enhancing treatment efficacy [12].

A higher frequency of symptom assessment has been associated with longer survival times, fewer days of inpatient care, and lower healthcare costs [16]. Current clinical methods employ well-designed scale systems to measure the symptom severity (*e.g.*, the Hoehn and Yahr scale [25] and the Unified Parkinson's Disease Rating Scale (UPDRS) [41]), which require the patients to perform a series of predefined motions to evaluate tremor severity. Frequent hospital visits can be burdensome for patients, and a shortage of clinicians adds to the inconvenience. Additionally, the white-coat effect and predefined motion assessments fail to fully capture the

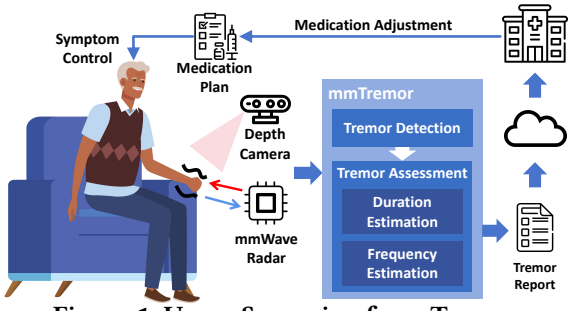


Figure 1: Usage Scenario of mmTremor.

impact of symptoms in daily life [23]. Since tremors are intermittent and unpredictable, their influence on daily activities is crucial for tailored therapy. Thus, continuous and objective monitoring of symptoms in real-life settings is essential.

To develop a practical tremor monitoring system in daily life, we must consider the following requirements: (1) non-intrusive: continuous monitoring requires comfort for effective system deployment. (2) privacy-preserving: privacy issues occur more urgently in daily life than in clinical settings [1]. (3) robust and accurate: as a severity biomarker, the system must detect tremors and assess their frequency and duration despite significant interference from daily activities, including eating and drinking, *etc.*. For example, during eating, significant arm movement can overshadow subtle tremors (millimeter to centimeter level), and other limbs may occlude or distract the monitoring system. Additionally, variations in movement across different subjects complicate the extraction of tremor patterns.

Recent research efforts have focused on monitoring tremor, the major motor symptom through various solutions, including wearable devices [3, 11, 30], RGB and depth camera-based solutions [4, 45, 61] and radio-frequency (RF)-based solutions [7, 20, 63]. However, none of these options fully meet practical monitoring requirements. Wearables require close contact and usually multiple devices, which can cause discomfort and reduce patient adherence [23]. RGB camera solutions [4, 33, 45] raise privacy concerns in daily life and are sensitive to lighting conditions while depth cameras struggle to detect small-scale tremors [17]. Several RF-based solutions have been proposed for privacy-preserving and continuous tremor monitoring [7, 20, 63]. These solutions assume that RF signals are solely changed by tremors, neglecting the variations in tremor source locations and the interference from movements of other limbs. Consequently, they may not function effectively in situations where tremor patterns vary or are overshadowed by significant movements during activities of daily living (ADL). Besides, the analysis of data derived from simulations of healthy subjects overlooks the unique characteristics of patients.

To bridge the gap towards a practical monitoring system, we recognize that depth cameras are effective for tracking ADL [42]. Nonetheless, they may overlook intricate tremor patterns due to noise stemming from distances and minor movements [44]. In contrast, sensitive mmWave signals can detect tiny movements. Thus, we propose integrating depth cameras with mmWave radar for comprehensive tremor detection at both micro and macro levels. Despite this integration, three significant challenges remain.

Firstly, accurately localizing the tremor source is difficult. mmWave signals have a coarse angular resolution, which hampers tremor source detection. While depth cameras can track body movements, they struggle with rapid actions (*e.g.*, swinging an arm) and subtle movements (*e.g.*, tremors), leading to bias in isolating the target tremor signal.

Secondly, tremor signals are heavily influenced by ADLs. Tremor (3-12 Hz) are inevitably overwhelmed by harmonic components and energy leakage from ADLs [35]. The composition of tremors and other limb movements exhibits nonlinear characteristics in both amplitude and phase [10], and the variety of ADLs introduces different motion interferences.

Lastly, tremors can manifest unpredictably due to factors like medications, activities, and emotions, distinguishing them from continuous vital signs such as heartbeats. This heterogeneity in tremors makes it challenging to create a robust system that can adapt quickly to varying conditions and accommodate different patients affordably.

In this paper, we propose mmTremor, a practical tremor monitoring system for daily life. We first develop a tremor source localization algorithm using mmWave-depth fusion tracking embedded with tremor-related prior knowledge. To address the nonlinear impact of ADLs, we propose a dynamic weighted mmWave-depth fusion deep learning pipeline, where the mmWave modality focuses on micro-movements and the depth modality extracts ADL information. To tackle the signal diversity in real-world scenarios, we propose an unsupervised contrastive adaptation scheme with customized data augmentation. We minimize the feature divergence between the labeled source data and the unlabeled real-world data through adversarial training. Then, based on the intrinsic signal diversity across different samples, we enhance model discriminability from the sample-wise label while preserving the label information. This adaptation scheme significantly improves mmTremor's robustness without incurring additional deployment costs.

Unlike prior works that primarily rely on simulated tremors from healthy subjects in laboratory settings, which lack the diversity of tremors found in real patients, we build mmTremor with 28 patients with PD or ET and 9 healthy subject simulations across over 20 real-world environments during ADLs, including offices, hospitals, and patients' homes. The dataset consists of around 40,000 samples, each with a duration of 3 seconds, labeled with IMU data [23]. It includes subjects of varying ages, genders, and disease severities assessed by medical professionals. Extensive evaluations

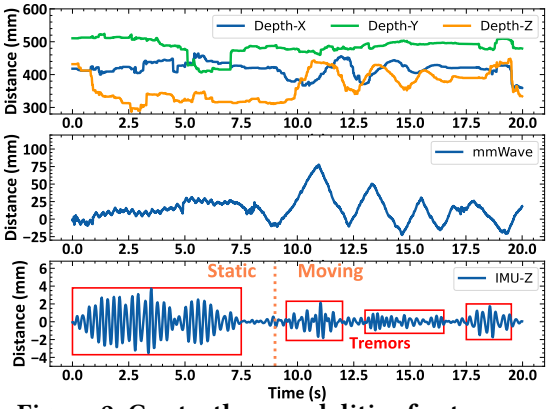


Figure 2: Contactless modalities for tremor.

demonstrate that mmTremor achieves a high and robust accuracy of 0.877 macro-F1 with unseen users. mmTremor can serve as a foundation model to support tremor analysis: (1) The tremor duration detected by mmTremor has high correlations of 0.84 with the assessment made by medical professionals and 0.97 with the IMU contact data. (2) It achieves a low frequency estimation error of 0.43 Hz, reducing the error by 46.9% compared to the state-of-the-art (SOTA) baseline [20].

Our contributions can be summarized as:

- We introduce the first practical contactless tremor detection system, designed to assess and improve medication efficacy for patients with PD and ET.
- We propose a series of techniques, including tremor source separation, deep tremor retrieval and unsupervised contrastive adaptation for robust tremor detection.
- We evaluate mmTremor in 28 patients with PD and ET, along with 9 healthy subject simulations in more than 20 environments. Our results demonstrate that mmTremor achieves an accurate and robust performance of 0.877 macro-F1 in diverse conditions. We also showcase that mmTremor holds great potential to serve as a foundation tool for further tremor analysis.
- The data will be open-sourced [37] to facilitate further in-home care research. To the best of our knowledge, this is the first multi-modal benchmark (mmWave, depth images, and IMU data) related to PD and ET patients.

2 PRELIMINARIES

In this section, we will introduce the prior knowledge about tremor and contactless sensing for tremor detection.

2.1 Tremor Priors

Tremor is defined as an involuntary, rhythmic, and oscillatory movement of a body part, typically caused by the contractions of antagonistic muscles. It is the most common movement disorder and can manifest in various forms, such as rest tremor, postural tremor, and kinetic tremor. In Parkinson's Disease (PD), tremors often manifest as resting tremors, most noticeable when the hands are at rest, resembling a pill-rolling motion. In contrast, Essential Tremor (ET), is characterized by tremors in both hands and arms during action or when resisting gravity, without accompanying neurological signs. Tremor may also affect the head, voice, or lower limbs. Although ET shares similar tremor characteristics with PD [6], it has distinct treatments that necessitate specific tremor monitoring [24], prompting us to include ET in our investigation.

2.1.1 Time-Frequency Analysis. From a time-frequency analysis perspective, tremors manifest as a dominant main frequency accompanied by a series of harmonic frequencies [28]:

$$x(t) = \sum_k A_k sin(2\pi k f t + \phi_k), \quad k = 1, 2..., \tag{1}$$

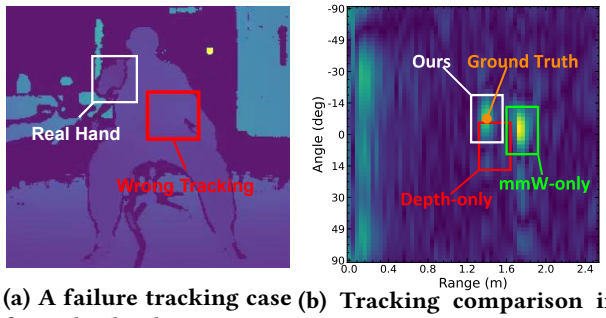
where $x(t)$ represents the tremor movement, f denotes the main frequency (ranging from 3 to 12 Hz), k means the k-order harmonic frequencies, A_k and ϕ_k denote the amplitude and initial phase of the k-order harmonic components, respectively. Tremor frequencies vary among individuals and causes of diseases, making it an important biomarker for customized treatment.

2.2 Contactless Sensing for Tremor

Given that only 30% of users are willing to wear contact sensors for a lasting period and that tremors require monitoring in daily life [23], it is desirable to develop a non-intrusive and privacy-preserving contactless tremor detection system. We will explore two privacy-preserving modalities: depth images and mmWave signals.

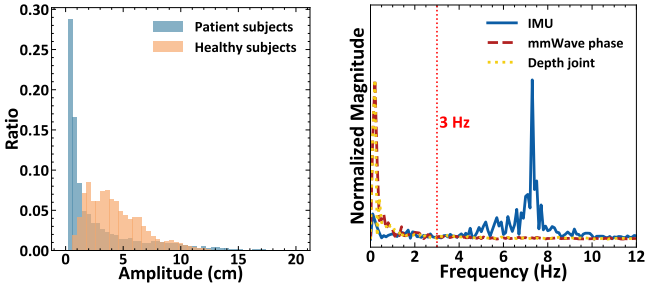
2.2.1 Depth Sensing for Tremor. Depth cameras are increasingly popular for human activity analysis, with many researchers utilizing skeleton tracking coordinates to measure or detect tremors [8]. However, these systems typically operate in the near-field (under 0.5 m) due to hardware limitations. For instance, the Azure Kinect camera suffers from random error and a systematic error of 0.1% of distance [36], which can affect the detection of common mm- and cm-level tremors. Furthermore, recent studies indicate that depth cameras exhibit greater inaccuracies in dynamic environments [44], complicating the detection of minor and rapid tremor movements in daily life.

2.2.2 mmWave Sensing for Tremor. Relying on their short wavelength, mmWave signals have proven effective for sensing tiny vibrations [21]. making them attractive for tremor monitoring [20]. However, their high sensitivity also leads to significant interference from irrelevant motions, which can be difficult to filter out, particularly in daily monitoring scenarios where movements can be large and irregular.



(a) A failure tracking case from the depth image. (b) Tracking comparison in radar Range-Angle profile.

Figure 3: Illustration of different tracking methods.



(a) Tremor Amp. distribution. (b) FFT comparison.

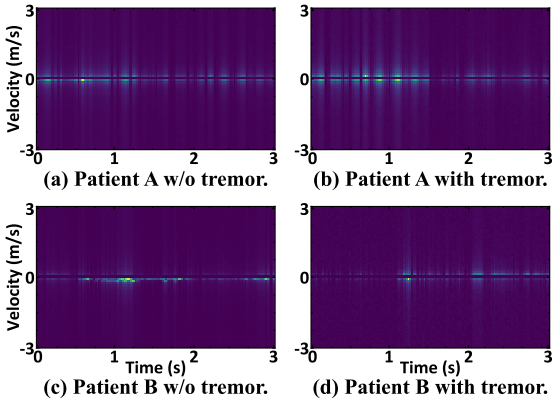
Figure 4: Illustration of different tremor characteristics between patients and healthy subjects, and the comparison of tremor signals of different modalities.

2.2.3 Comparison Between Both Modalities for Tremor. Figure 2 presents an ET patient who first holds his arm up for about 9 seconds and then draws circles for another 11 seconds, with an IMU sensor attached to the tremor-affected hand for signal collection. The amplitude for IMU is obtained by integrating the signal after applying a bandpass filter in the range of 3-12 Hz. The depth camera effectively tracks macro motion but introduces random uncertainties during the initial 9 seconds. Meanwhile, mmWave signals can detect tremors but are significantly affected by these macro movements. This highlights the necessity of fusing depth and mmWave data: by using the depth camera to cancel out macro movement interference, we can better extract the subtle tremor vibrations.

3 TREMOR DETECTION CHALLENGES

As a significant symptom in patients with PD and ET, tremor is a clinically proven biomarker that helps assess and manage drug efficacy [46]. However, several challenges exist for implementing a practical system in daily life.

(1) Dynamic Tremor Source. Tremors can significantly impact daily activities like eating and drinking, resulting in dynamic tremor sources. While depth cameras offer reasonable tracking accuracy [53], they can malfunction during rapid movements or when limbs are obscured (*e.g.*, by lifting the arm) as illustrated in Figure 3(a). This leads to considerable bias in accurately locating the tremor source. Similarly, mmWave signals face challenges in tracking a moving tremor source due to their coarse angular resolution ($\approx 15^\circ$), making it only suitable in static scenarios with the largest-energy-localization [20]. However, as shown in Figure 3(b), this approach can yield incorrect detection results, as it may capture larger reflections from the chest instead of the tremor-affected hand.



(a) Patient A w/o tremor. (b) Patient A with tremor. (c) Patient B w/o tremor. (d) Patient B with tremor.

Figure 5: Signal variations across subjects (drinking).

(2) Substantial ADL Coupling Interference. Tremor in patients typically exhibit amplitudes at the mm- and cm-level [41]. The distribution of tremor amplitudes from both patients and healthy subjects is shown in Figure 4(a). The tremor amplitudes in patients are concentrated within a quite smaller range than those of healthy simulations.

In addition to micro movements, the limited reflection area of the tremor source (*e.g.* hand) reduces the variations in the captured signals. Consequently, the frequency spectrum from common daily activities, which spans up to 12 Hz, causes significant interference with the tremor distribution. Figure 4(b) shows the FFT magnitudes of the IMU, depth, and mmWave signals from a kinetic tremor case. The tremor is simulated by a healthy subject seated with an IMU attached to his hand. The mmWave radar and depth camera are set 1.5 meters before the subject. The joint estimation from the depth camera and the phase of mmWave signals reflected from the hand are used to calculate the frequency spectrum. Compared to IMU signals, the frequency magnitudes from contactless modalities (depth and mmWave) are significantly overshadowed by the spectrum of ADLs. The ratio of tremor spectrum energy (3-12 Hz) to the interference spectrum (0-3 Hz) from contactless modalities is around 10,000 × smaller than that of IMU signals (0.003 vs. 26.01).

(3) Micro Tremor Variation with Large Domain Diversity. The primary rationale for detecting tremors in daily activities is to analyze variations induced by tremors within the composed signals. However, these variations may differ across various domains, such as in patients. To investigate

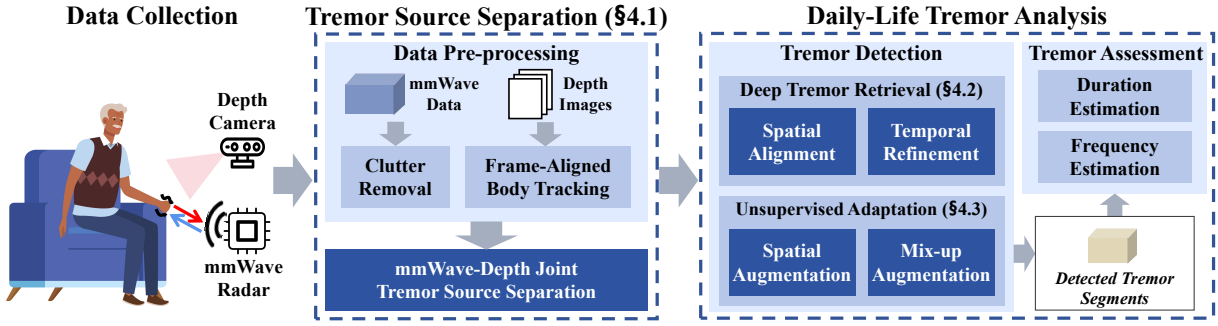


Figure 6: Overview of mmTremor.

this, we collect the motion-related radar Doppler data from two PD patients performing the same activity, *i.e.* drinking.

As shown in Figure 5, we observe the following: **(1)** Tremor introduces little changes in the radar doppler signal (Figure 5(a) - Figure 5(b)). **(2)** Tremors vary significantly across different subjects (Figure 5(b) - Figure 5(d)). **(3)** Different patients exhibit considerable variability even without tremor impacts (Figure 5(a) - Figure 5(c)). We compare the Structural Similarity (SSIM) between the cases of Patient A and Patient B. The (MEAN±STD) similarity of tremor and no-tremor cases from patient A is 0.572 ± 0.073. In contrast, the similarity of tremor cases from the two patients is 0.482 ± 0.184, which is even lower than that with no-tremor cases from the same subject. This illustrates the micro tremor discrepancy is highly influenced by the huge domain diversity across different subjects.

4 SYSTEM DESIGN

To solve the above challenges, we propose mmTremor, a practical contactless tremor detection system leveraging complementary synergy from depth and mmWave modalities. Figure 6 illustrates our system's architecture with three core modules: Tremor Source Separation (Section 4.1), Deep Tremor Retrieval (*DeepTremor* in Section 4.2), and Contrastive Adaptation (ConAda in Section 4.3). Finally, the detected segments will be used to assess tremor duration and frequency.

4.1 Tremor Source Separation

To tackle the challenge of dynamic tremor source tracking, we observe that both modalities can collaborate to mitigate each other's shortcomings. Therefore, we propose the Tremor Source Separation module, which filters error-prone locations based on the joint decisions of the two modalities and prior knowledge of tremors as illustrated in Figure 7.

4.1.1 Data Pre-processing. The IF mmWave data first undergoes Range-Doppler-Angle-FFT modulation to separate reflections from different distances, velocities, and angles.

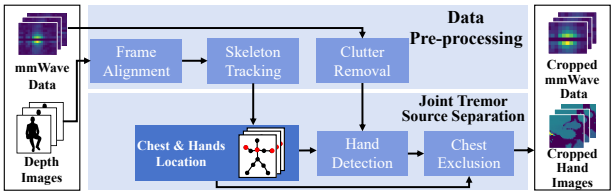
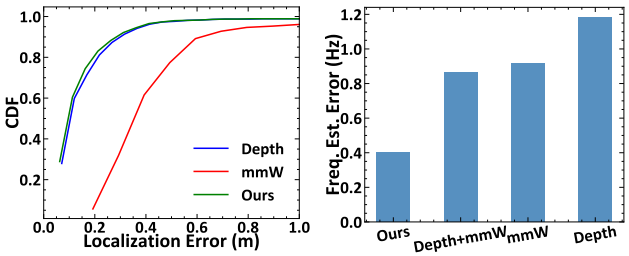


Figure 7: Illustration of Tremor Source Separation

Next, we perform clutter removal to mitigate the static multipath impacts. For the depth image, to combat the issues including the missed frames and mismatched sampling rate between the depth camera (~30 Hz) and the radar (~50 Hz), we interpolate the depth frames with the nearest timestamp sampling. Then we perform skeleton tracking [53] to extract key joints such as the chest and both hands in the spatial dimension for each depth frame.

4.1.2 mmWave-Depth Joint Tremor Source Separation. After preprocessing, depth images provide a coarse estimation for mmWave data, while mmWave data can help refine depth tracking due to its high sensitivity. Specifically, using the pre-processed body skeleton from depth images, we map the absolute spatial coordinates to the mmWave Range-Angle profile as rough localization anchors. Instead of directly selecting the energy peaks around the rough anchors which may stem from large body reflections or random movements, we design spatial filters based on the tremor priors that PD and ET tremor happens more frequently in limbs instead of the chests [41]. To mitigate such interference, we additionally leverage chest locations from depth-based tracking to exclude torso reflections. With a range resolution of around 4 cm, the hand and chest can be separated in the mmWave profile. If they are so close that only one peak appears, we still use this reflection to detect tremors. Finally, we crop the mmWave and depth data with the extracted spatial coordinates that preserve the tremor source information. Specifically, we select the radar signals with 8 range bins (around a 30 cm scope) around the tremor source considering the potential tracking error, and then perform beamforming to enhance the received signals. Then we preserve the sequence



(a) Tracking error. (b) Frequency estimation.

Figure 8: Comparison of different tracking methods.

of radar Range-Doppler FFT (RD-FFT) signals as it reflects the motion distribution. The depth images are cropped to fit the size of the tremor source combined with the upper body joints for macro motion analysis.

To validate the effectiveness of our approach, we compare our results using data from three patients with advanced RGB-based skeleton tracking [29], as shown in Figure 8(a). We observe that both ours and pure-depth results greatly surpass the mmWave-based localization results in the previous RF work [20]. Figure 8(b) displays the tremor frequency extracted from the two modalities, including the mmWave signals based on our tracking results, the mmWave purely based on the depth joints (Depth+mmW), the mmWave-only tracking results (mmW), as well as the depth-only joint estimation results (Depth). The ground truth is calculated from the peak frequency of IMU signals. A lower frequency error indicates better tremor feature extraction. Despite similar tracking errors, our design preserves the most tremor variations due to the mmWave refinement design compared to the simpler depth+mmWave scheme.

4.2 Deep Tremor Retrieval

After tremor source localization, we obtain the signal containing the most tremor information. To retrieve this information from nonlinear ADL influences, we propose *DeepTremor*in Figure 9, based on the observation that mmWave captures minute tremor displacements with a coarse angular resolution, while depth cameras track macro motions with a coarse displacement discrimination.

4.2.1 Unimodal Encoders. Since the radar beamformed RD-FFT profile and the depth inputs (cropped images and body joints) are heterogeneous modalities, we first design unimodal encoders to extract features from each modality and then project them into the same latent space. We leverage the convolution(2+1)D blocks [55] with a residual format (ResConv), which has proven effective in action recognition with a reduced risk of overfitting. We employ five ResConv blocks to aggregate the mmWave spatial-velocity information. For depth features, we use two blocks combined with learned joint embeddings from two linear layers to preserve

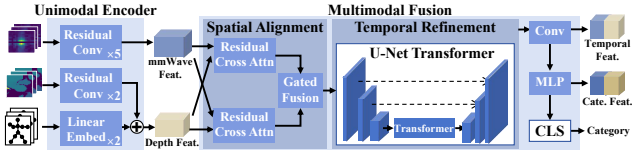


Figure 9: Model structure of *DeepTremor*.

the macro motion features. Both mmWave and depth features are further transformed to the same latent space.

4.2.2 Spatial Alignment. Since mmWave data focuses on the micro motions around the tremor source, while depth data captures the macro motions, we design a spatial alignment module to dynamically fuse the two modalities. We first leverage Cross-Attention [9] on the two modalities separately, aligning the query modality with the learned attention map from the other modality. Then we design a residual mechanism over the spatial dimension to dynamically preserve the original modality features.

Considering that different modalities pay different attention to the tremor conditions with varying amplitudes, we further design a gated fusion module to adaptively coordinate the inter-modality relationships:

$$\begin{aligned} \alpha_{mmW} &= \sigma_{mmW}(W_{mmW} \cdot [O_{mmW}; O_{depth}]), \\ \alpha_{depth} &= \sigma_{depth}(W_{depth} \cdot [O_{mmW}; O_{depth}]), \\ O_{fused} &= \alpha_{mmW} * O_{mmW} + \alpha_{depth} * O_{depth}, \end{aligned} \tag{2}$$

where σ_{mmW} and σ_{depth} are the sigmoid functions. W_{mmW} and W_{depth} are the learned linear projections. O_{mmW} and O_{depth} are the outputs from the residual cross-attention modules while O_{fused} is the output from the gated fusion module.

4.2.3 Temporal Refinement. Since tremor manifests the time-series property and modulates both modalities simultaneously, we design a UNet-based [47] transformer module to dynamically aggregate features from different temporal resolutions. As shown in Figure 9, the left part encodes the fused features using convolution-based downsampling with a stride of 2. The final layer, with the largest receptive field, is empowered with an advanced pre-trained VideoMAE transformer [54] which transfers dynamic temporal action knowledge to the relatively limited tremor datasets. The right side of the UNet concatenates the shortcut features and performs upsampling with transposed convs to generate the Temporal Features O_{Temp}. Furthermore, we aggregate multi-channels within the Temporal Features O_{Temp} with convolutions and utilize the linear transformations to generate the Category Features O_{Cate}. Finally, A multi-layer perception (MLP) is utilized as the classifier (CLS) for the category decision.

4.3 Unsupervised Tremor Adaptation

To combat the challenge of large domain diversity suppressing micro tremor patterns, we need to eliminate the domain

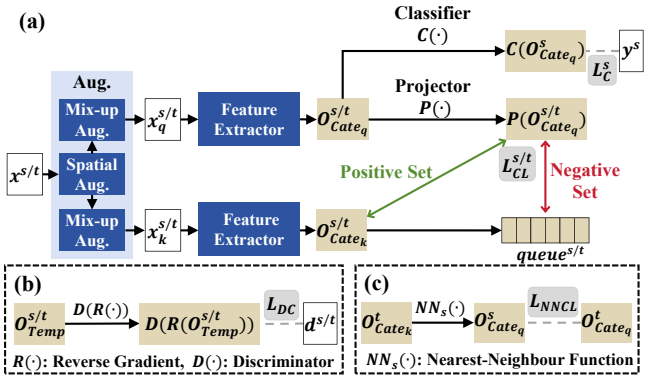


Figure 10: Architecture of ConAda.

shift while preserving the tremor information inside the samples. Additionally, considering the mobility difficulties faced by PD/ET patients, we aim to minimize labeling efforts as much as possible. In this section, we propose an unsupervised domain adaptation (UDA) scheme to transfer the model trained from the collected source data to the target user. To this end, we propose a contrastive learning-based adaptation scheme (ConAda) as shown in Figure 10 to fulfill the two requirements: (1) eliminate the domain shift. (2) enlarge the discrimination across samples.

Eliminate the Domain Shift. To mitigate the data shift from the target user with our collected source distribution, we leverage the domain adversarial neural networks (DANN) [18] that trains a domain discriminator $D(\cdot)$ with reverse gradients to force the tremor detection model to learn domain-invariant features. To capture the temporal domain information, we feed Temporal Features O^{s}_{Temp} from the source data and O^{t}_{Temp} from the target data to the domain discriminator $D(\cdot)$ shown in Figure 10(b). Then $D(\cdot)$ is trained to separate different domains while the detection model *DeepTremor* is trained to fuse these domains simultaneously. The reverse gradient layer $R(\cdot)$ is applied to adversarially optimize both the two components with the domain classification loss:

$$L_{DC} = CE(D(R(O^{s}_{Temp})), d^{s}) + CE(D(R(O^{t}_{Temp})), d^{t}), \quad (3)$$

where CE is the cross-entropy loss. d^{s} and d^{t} are the domain labels, respectively.

Enlarge Sample Discrimination. Despite the tremor variations being suppressed, they still introduce diversity to the captured signals. Based on this intuition, we leverage contrastive learning (CL) [22] to enhance the model discriminability across different samples illustrated in Figure 10(a). Specifically, leveraging the intrinsic signal diversity across samples, CL aims to learn similar features from two augmented views of one sample (positive set) and discriminate the features from other samples (negative set), thus bringing benefits to capturing micro-tremor representations [52].

Since CL is highly influenced by the data augmentation schemes that generate two different positive samples related to the tasks [52], we propose a tremor-specific data augmentation scheme for both modalities. To preserve tremor's temporal properties, we first augment the inputs from the spatial perspective with the following schemes:

- **mmWave**: Considering the coarse angular resolution of the radar, the angle-FFT operation may cause spectrum leakage as shown in Figure 3, sharing mutual information with adjacent angle areas [26]. Thus we beamform the RD-FFT mmWave with 5 random angle bins ($\sim\pm10^{\circ}$) around the tracking results as the augmented mmWave samples.
- **Depth**: The cropped depth images contain the dynamics of tremor sources while the upper body joints preserve macro motions. Thus we randomly crop the subset of the cropped images and then resize it by interpolation, which mimics when the images are not perfectly cropped. We also randomly mask the upper body joints to simulate the joint misdetection cases.

Based on the intuition that activities composed of tremors should have tremor-like behavior we design a mix-up scheme [64] tailored for tremor cases. Specifically, we randomly form sample pairs in the same batch and mix them up with random weights. Then we set the corresponding labels as 'tremor' if there is a tremor sample in the mixed pair.

With the augmentation schemes, we can generate two views of each sample as query x_q and key x_k and leverage the momentum contrast (MoCo) framework [22] to enlarge the sample discrimination in each domain. Each x_q and x_k are processed by *DeepTremor* to generate the category emebddings O_{Cate_q} and O_{Cate_k}, where O_{Cate_k} is generated by the momentum-updated *DeepTremor* feature extractor[1]. To make the augmented features close while enlarging sample discrepancies, we add the following contrastive loss:

$$l^{s}_{ii} = P(O^{s}_{Cate_{qi}}) \cdot O^{s}_{Cate_{ki}}, \quad (4)$$

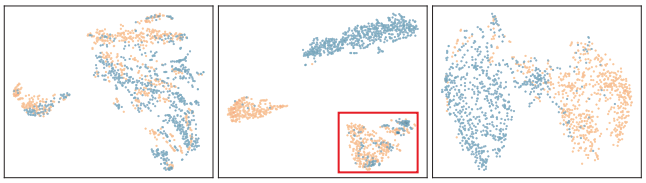
$$l^{s}_{ij} = P(O^{s}_{Cate_{qi}}) \cdot O^{s}_{Cate_{kj}}, \quad (5)$$

$$L^{s}_{CL} = -\frac{1}{N}\sum_{i=1}^{N} log\frac{\exp(l^{s}_{ii}/\tau)}{\exp(l^{s}_{ii}/\tau) + \sum_{j=1}^{J}\exp(l^{s}_{ij}/\tau)}, \quad (6)$$

where $P(\cdot)$ is the projection layer that transforms the positive features closer. $O_{Cate^{s}_{qi}}$, $O_{Cate^{s}_{ki}}$ and $O_{Cate^{s}_{kj}}$ are the i-th O_{Cate_q}, i-th O_{Cate_k} and j-th O_{Cate_k} features from the source domain, respectively. τ is the temperature scaling parameter and J is the size of the sample queue. To improve sample discrimination, we regard the augmented views from the sample as positive samples while the others as negative samples.

The contrastive loss pushes the sample away from the queue that stores other negative samples and thus enhances the model's discriminability. This loss also constraints the

[1]The feature extractor is *DeepTremor* without the last classifier.



(a) No adaptation. (b) DANN. (c) Our ConAda.

Figure 11: t-SNE visualization for the embeddings from unseen subjects with different adaptation schemes. The yellow dots are for 'tremor' cases while the blue ones denote 'no-tremor' cases.

model trained on the unlabelled target data denoted as L^t_{CL} to enhance the target sample discrimination.

Besides unlabelled contrastive learning, we need to align the contrasted domain representations in Figure 10(c). Since the target domain has no available labels, to preserve the label information, we adopt the nearest-neighbor contrastive learning [43] to align the target representations with the nearest source representations:

$$O^s_{Cate_{qi}} = NN_s(O^t_{Cate_{ki}}), \tag{7}$$

$$L_{NNCL} = -\frac{1}{N_t}\sum_{i=1}^{N_t} log\frac{\exp(O^t_{Cate_{qi}} \cdot O^s_{Cate_{qi}}/\tau)}{\sum_{j=1}^{N_s}\exp(O^t_{Cate_{qi}} \cdot O^s_{Cate_{qj}}/\tau)}, \tag{8}$$

where $NN_s(\cdot)$ is the nearest-neighbor function with mean-squared error as the distance metric to retrieve the source embedding closest to the target domain embedding. The final contrastive loss in our ConAda is:

$$L_{CL} = \alpha \cdot (L^s_{CL} + L^t_{CL}) + \beta \cdot L_{NNCL}, \tag{9}$$

where α and β are hyperparameters to control loss weights.

Capture Tremor Information. To ensure *DeepTremor* maps the contactless modalities to the tremor detection results, we adopt the cross-entropy loss on the final embeddings with the tremor labels:

$$L^s_C = CE(C(O^s_{cate_q}), y^s), \tag{10}$$

where $C(\cdot)$ is the *DeepTremor* last classifier. $O^s_{cate_q}$ and y^s are the features and labels in the source domain, respectively.

Summarize all losses: Finally, our adaptation scheme ConAda optimizes the following loss with the domain adaptation weight γ:

$$L = L^s_C + L_{CL} + \gamma \cdot L_{DC}. \tag{11}$$

Figure 11 visualizes the Temporal Features O_{Temp} from the unseen subjects by averaging the timestamps with t-SNE [56]. The 'No adaptation' scheme fails to separate the correct labels, highlighting the necessity of domain adaptation in the feature space. Although DANN benefits from the learned features, it still mixes the 'tremor' samples with 'no-tremor' samples since it cannot preserve label information [43]. With the contrastive learning scheme, our ConAda better separate the features by enhancing sample discrimination and eliminating domain shifts.

Figure 12: Various environments in data collection.

5 IMPLEMENTATION

5.1 Data Collection

Ground Truth. Inertial measurement units (IMUs) have been widely utilized for tremor assessment and acknowledged as a gold standard [39, 57, 61]. We build a glove with 6 IMUs (MPU6050) to collect tremor data from five fingers and the back of the hand with 50Hz. A 3-12 Hz band-pass filter is applied to the data from six sensors. Then the root mean square (RMS), peak power [23] and the tremor amplitudes are calculated using a 3-second sliding window with 0.5-second overlap [49]. Segments exceeding predefined thresholds are labeled as 'tremor' cases. For example, following the previous work [15], segments with amplitudes greater than 5 mm will be regarded as the 'tremor' cases.

Patients Recruitment. We collaborate with two medical centers with our institute's IRB approval. Patients are recruited from the outpatient clinic and inpatient department under the supervision of experienced clinicians. Experiments were conducted with informed consent from the patients and under the guidance of these clinicians. Our study involves 37 volunteers including 28 patients (24 patients with PD and 4 patients with ET) and 9 healthy simulations. The experiments were conducted in 23 diverse settings (2 outpatient clinic rooms, 8 inpatient units, and 13 patients' homes) in Figure 12. Two clinicians evaluated the patients with UPDRS tremor subscores 3.15-3.18 [41]. The patients have a MEAN±STD tremor subscores of 4.71 ± 2.72 and an age distribution of 67.14 ± 8.13 years old, with the disease lasting from 2 years to more than 20 years.

Experiment Design. Subjects are seated approximately 150 cm from the device, with distances varying from 50 to 300 cm and orientations ranging from -30 to 30 degrees based on environmental constraints. Patients perform five ADLs affected by tremors: eating, drinking, drawing, resting, and using a phone, as referenced in prior research [24, 48]. Using the phone lasts 1 minute for twice, while other activities last

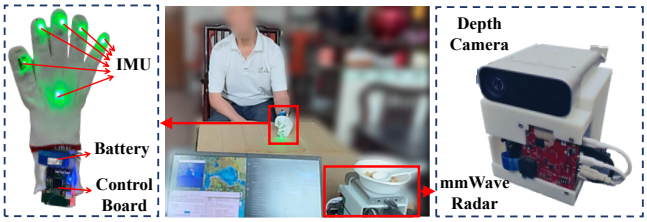


Figure 13: Implementation of the sensing platform and ground truth hardware.

about 15-20 seconds and are repeated around 10 times. After providing informed consent, patients watch a demonstration video before performing tasks based on their daily habits.

For healthy subjects, simulation experiments involve 21 combinations of tremor types, limb movements, and hand orientations to mimic patient scenarios. Subjects watch video examples and simulate tremors for 10 seconds, repeating the process 12 times for rest tremors and 3 times for kinetic tremors. In total, over 40,000 segments are collected, each lasting 3 seconds with a sliding window of 0.5 seconds.

5.2 System Implementation

Sensing Platform. Our sensing platform, depicted at the right of Figure 13, combines TI IWR1843 mmWave radar with an Azure Kinect DK depth camera. The depth camera operates at a maximum frame rate of 30 Hz in the WFOV mode. The radar operates at 77 GHz with a bandwidth of 4 GHz, equipped with 3 transmitting antennas and 4 receiving antennas. It sends frames at 50 Hz, each with 64 chirps and 256 sampling points for each chirp.

Software. We build mmTremor with one NVIDIA RTX 3090 GPU based on PyTorch 2.0.1. The body joints are exclusively processed from the depth images with the Azure Body Tracking SDK 1.4.1. The cropped mmWave RD-FFT inputs have 64 doppler bins and 8 range bins around the extracted tremor source with I/Q channels. The depth cropped images have a size of 32×32 pixels to cover the tremor source (*e.g.*, hands) while we select 14 upper body joints including the head, chest, left/right arms, and left/right hands. The details of *DeepTremor* and ConAda are as follows.

***DeepTremor* mmWave Encoder.** The input channels of the five ResConvs are (2, 16, 32, 64, 128) with the output channel of 256. The first ResConv has a spatial/temporal kernel size of (1, 1) while the others are (9, 11). The spatial strides are (1, 2) at the second ResConv while the last three ResConvs have the same (2, 2) spatial strides. Finally, we reshape the embeddings and map them into a 128-dim latent space with a linear projection.

***DeepTremor* Depth Encoder.** The input channels of the two ResConvs are (1, 16) with 64 output channels. The spatial/temporal kernel sizes are (11, 9) with the spatial strides of (4, 4). We convert the reshaped embeddings into the 128-dim latent space with a linear projection. The joint embeddings are also projected to the 128-dim space with two liner layers.

***DeepTremor* Spatial-Temporal Feature Extractor.** The spatial alignment module takes the 128-dim embeddings from both modalities and outputs a 64-dim embedding. Then the temporal refinement module adopts a three-layer U-Net, where each layer changes the temporal resolution and increases the feature channel. The last layer contains a conv to align the temporal features with the pretrained VideoMAE encoder input of 384-dim [54]. The feature extractor outputs the Temporal Features $O_{Temp} \in \mathbf{R}^{128\times150}$, which will be processed by three convs with a linear layer for the Category Features $O_{Cate} \in \mathbf{R}^{128}$. Finally, the MLP classifier $C(\cdot)$ with two linear layers maps O_{Cate} to the target tremor category.

ConAda Domain Discriminator. The domain discriminator analyzes the domain information from the Temporal Features O_{Temp} with four Conv1D blocks with input channels of (64, 64, 64, 32) with the output channel of 128 and the stride of 2. The final domain classification layer converts the reshaped hidden latents to the domain labels.

Training Parameters. *DeepTremor* is trained with Adam optimizer with a cosine learning rate scheduler that increases from 0 to $1e^{-3}$ in 5 epochs and gradually reduces to $5e^{-5}$ in later 25 epochs. The temperature, momentum and queue size parameters of ConAda are 0.07, 0.99, and 8192, respectively. The ConAda loss weights in Equations 9 and 11 are $\alpha = 0.1, \beta = 1, \gamma = 5$ to ensure them in the same scale.

6 EVALUATION

6.1 Evaluation Methodology

Experiment Settings. We conduct the evaluation based on the 37 subjects. To evaluate the generalization of mmTremor, we use a **cross-subject** evaluation scheme with a random five-fold split. The subjects are split into 5 groups, each with 7-8 people. One subject group is selected as the test set while the remaining groups function as the training set by turns.

Metrics. We use the *F1-score*, *recall*, and *precision* for positive samples (*i.e.*, tremors) to evaluate the sensitivity and specificity of tremor detection. Since the non-tremor cases also count for the 'OFF' period for tremor duration estimation, it is important to assess our system on both tremor and non-tremor cases. Thus, we also use the *macro-F1*, and *accuracy* by averaging the metrics in both categories. In addition, *Matthews correlation coefficient (MCC)* [5] is used to evaluate the correlations of detection results with the ground truth.

Baselines. Since there are currently no available settings that extract tremors from the two modalities in daily activities, we established the baselines to understand the contributions of our proposed components. Specifically, these baselines are designed by removing or replacing specific modules of our system. We set the baselines as follows:

	ACC	M-F1	F1	R	P	MCC
BL-A	0.728	0.666	0.521	0.501	0.543	0.332
BL-B	0.773	0.713	0.581	0.521	0.656	0.434
BL-C	0.793	0.749	0.644	0.634	0.655	0.499
BL-D	0.826	0.799	0.726	0.717	0.736	0.598
BL-E	0.842	0.799	0.706	0.642	0.784	0.605
BL-F	0.848	0.808	0.720	0.654	0.801	0.623
BL-G	0.859	0.830	0.759	0.733	0.786	0.660
BL-H	0.872	0.848	0.787	0.803	0.772	0.696
Ours	**0.897**	**0.877**	**0.829**	**0.828**	**0.829**	**0.759**

Table 1: Overall performance of all subjects. 'BL': 'Baseline'. 'ACC': 'Accuracy'. 'M-F1': 'Macro-F1'. 'R': 'Recall'. 'P': 'Precision'.

- **Baseline-A:** It uses the cropped depth images as input with the unimodal depth encoder and the linear classifier without adaptation.
- **Baseline-B:** Compared with Baseline-A, it adds the pretrained video backbone VideoMAE [54] to enhance the feature extraction.
- **Baseline-C:** We replace the VideoMAE backbone with our proposed Temporal Refinement module.
- **Baseline-D:** It utilizes only mmWave signals to localize tremor sources, with mmWave RD-FFT as input to our unimodal mmWave encoder and the linear classifier.
- **Baseline-E:** It fuses mmWave RD-FFT with depth images based on an advanced multi-modal fusion backbone Multiway Transformer [60].
- **Baseline-F:** We replace the fusion Multiway Transformer of Baseline-D with our *DeepTremor* fusion.
- **Baseline-G:** Baseline-E + DANN adaptation [18].
- **Baseline-H:** We replace our tremor-specific augmentation in ConAda with the time-series augmentation scheme [43] and adopt the contrastive adaptation with *DeepTremor*.
- **Ours:** Our *DeepTremor* + tremor-specific ConAda adaptation scheme, which is the whole of mmTremor.

6.2 Overall Performance

The overall results for all subjects are presented in Table 1 while Figure 14 details the performance of both patients and healthy subjects.

Effectiveness of *DeepTremor*. The results of Baseline-A to C indicate that coarse information from a single depth camera can interfere with tremor data. While Baseline-D (mmWave-only) shows comparable performance to the fusion approach (Baseline-E) on some metrics, this is primarily because mmWave-only indirectly identifies vibrations by locating the patient's thorax. This approach, however, restricts fine-grained limb tremor analysis, like frequency estimation (see Figure 19(a)), making it inadequate for our goal of comprehensive tremor assessment. For comparison, Baseline-E fuses both modalities with an advanced Multiway Transformer. With the mmWave gains, Baseline-E achieves a high performance of 0.842 accuracy. Compared with Multiway Transformer, our dedicated *DeepTremor* (Baseline-F) considers the spatial-temporal characteristics of both modalities and tremors and achieves better performance, especially for healthy subjects who perform larger tremors.

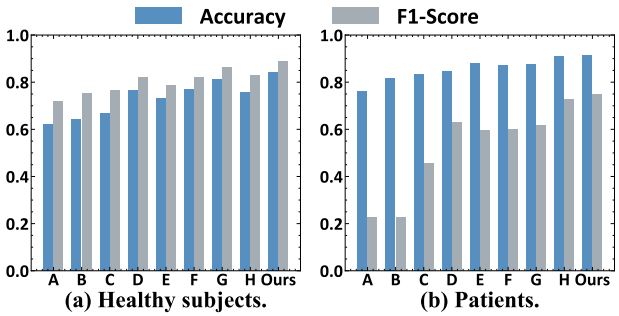


Figure 14: Overall performance on healthy subjects and patients.

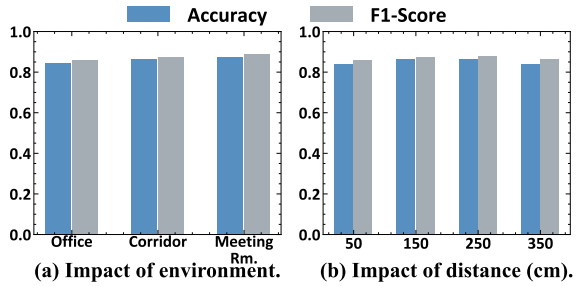
Effectiveness of ConAda. Despite the gain brought by *DeepTremor*, its F1-score is not satisfactory without adaptation as shown in the patients' performance of Figure 14 (b). Baseline-G integrates the common adaptation strategy DANN [18] and obtains minor enhancement. In contrast to DANN which omits the label information for the target domain [43], we could observe our proposed ConAda achieves the highest performance, including an accuracy of 0.897, macro-F1 of 0.877, and F1-score of 0.829. This is because ConAda adopts the tremor-specific augmentation and leverages contrastive learning to preserve the tremor features. To prove the effectiveness of the augmentation module in ConAda, we replace it with the common time-series augmentation [43]. Since patients typically exhibit minor and intermittent natural tremors (Figure 4(a)), it imposes more challenges on patients than healthy simulations. Compared with Baseline-H, our solution has a stable gain in both patients and healthy subjects, demonstrating the tremor-specific augmentation module's effectiveness.

6.3 Robustness Study

To evaluate the robustness of mmTremor, we randomly select 3 healthy subjects simulating tremors with the same ADLs as the patients under varied conditions.

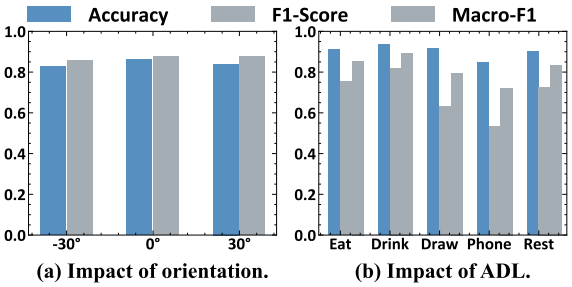
6.3.1 Impact of Environment. We first evaluate the impact of different environments, including a corridor, an office, and a meeting room. Results are shown in Figure 15(a). Our system maintains a high accuracy and a high F1-score over 0.86 across various environments.

6.3.2 Impact of Distance. Users may appear at varied distances from the system. We evaluate our system at distances



(a) Impact of environment. (b) Impact of distance (cm).

Figure 15: Impact of environment and distance.



(a) Impact of orientation. (b) Impact of ADL.

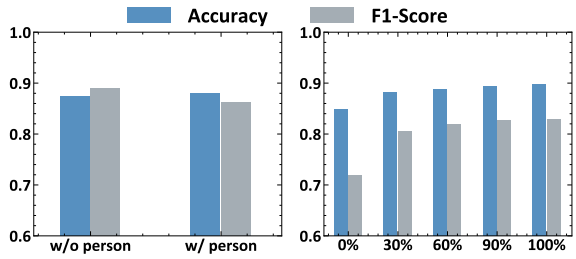
Figure 16: Impact of user orientation and movement.

of 0.5 m, 1.5 m, 2.5 m, and 3.5 m with a 0-degree orientation. Results in Figure 15(b) indicate that our system maintains good performance of above 0.84 F1-score at reasonable distances in real-world scenarios, from monitoring at the same desk (0.5 m) and across a room scale (3.5 m).

6.3.3 Impact of Orientation. In real-world settings, users may not face the devices accurately, so our system should tolerate variations in user orientation. We conduct experiments with orientations varied from -30° to 30°. The positive angle indicates the user is located at the clockwise direction towards the system. Results in Figure 16(a) demonstrate the system's good performance across a wide range of angles.

6.3.4 Impact of Interfering Person. We also simulated scenarios where other individuals walked around the subject. The subjects were seated 150 cm in front of the system at 0° orientation, while another person moved at a distance of 0.5~1 m, keeping the line of sight path clear. As shown in Figure 17(a), our system performs effectively even with nearby individuals present.

6.3.5 Impact of Movements. We analyze patient data during various activities. As shown in Figure 16(b), our system effectively mitigates the effects of large movements, maintaining high performance across different ADLs. However, the F1-scores for drawing and using a phone are lower, likely due to an imbalanced data distribution-—samples from drawing and using a phone are both only about 1/3 of those from drinking .Future work will focus on implementing imbalance learning techniques to address this issue.



(a) Impact of interfering person. (b) Impact of adaptation rate.

Figure 17: Impact of interference and adaptation rate.

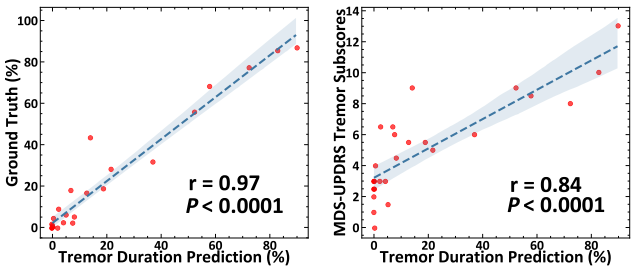
6.3.6 Impact of Adaptation Rate. To evaluate the impact of data volume on adaptation, we assessed our system using various adaptation rates (0, 30%, 60%, 90%, and 100%), which represent the proportion of data used for adaptation. As shown in Figure 17(b), even a small adaptation rate yields significant improvements, with performance gradually increasing as the rate rises. These results indicate that our system achieves strong generalization performance with only a small amount of data (approximately 5 minutes).

6.4 Case Study

6.4.1 Duration Estimation. One downstream application of mmTremor is to accurately assess the tremor duration after tremor direction, and further provide evidence for improving tremor choices [12]. We compute the tremor duration proportion to the total duration for each patient as shown in Figure 18(a). Our contactless solution has a high correlation of 0.97 compared with the contact IMU ground truth. This result presents the accurate tremor duration estimation across patients with different tremor severity, demonstrating that mmTremor is practical to provide useful results for clinicians while reducing user burdens.

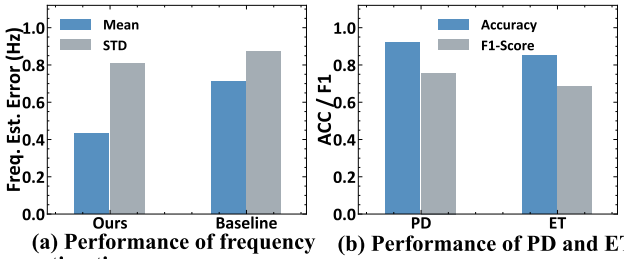
6.4.2 Correlation with Medical Assessment. As an important symptom and biomarker, tremor duration is highly related to the severity of PD. To demonstrate the effectiveness of mmTremor, we analyse the correlations between our tremor duration outputs with the average assessment from two experienced medical professionals, *i.e.* MDS-UPDRS [41] tremor subscore (3.15-3.18). The results shown in Figure 18(b) prove that our predicted tremor duration is strongly correlated with the MDS-UPDRS tremor subscore ($r = 0.84, P < 0.0001$). This result demonstrates that mmTremor holds great potential to serve as a clinically relevant assessment of tremor severity in daily scenarios without any intrusiveness.

6.4.3 Frequency Estimation. As another important downstream task of tremor analysis, frequency illustrates the tremor characteristics, especially the cause of tremors. Here we evaluate the frequency estimation supported by mmTremor. The ground truth is obtained via the peak frequency of IMU signals within 3-12 Hz. For ours, we extract the phase of



(a) Correlation with ground truth. (b) Correlation with MDS-UPDRS subscores.

Figure 18: Correlation with IMU ground truth and MDS-UPDRS subscores.



(a) Performance of frequency estimation. (b) Performance of PD and ET.

Figure 19: Performance analysis of tremor frequency estimation and patient groups.

mmWave signals based on our tremor source separation, while the baseline [20] tracks with only mmWave signals. For fairness, we compare our results under the same condition of the baseline that only analyzes the indicated tremor cases. The results in Figure 19(a) present that the tremor frequency estimation error with our tracking strategy is only 0.43 Hz, much lower than the baseline (0.81 Hz), which is often misled by confounding limbs and chest.

6.4.4 Performance under Different User Statuses. A reliable tremor detection system should track the user's status. We invite a PD patient who has undergone the Deep brain stimulation (DBS) treatment. The symptoms of PD will be significantly suppressed when the DBS machine is on. We present the duration estimation results for different statuses in Table 2. The results show that mmTremor accurately monitors the duration at different statuses and validates it is practical to track the patient conditions in daily life.

6.4.5 Performance for PD and ET. Since our dataset contains patients with PD and ET, we present the performance across different diseases in Figure 19(b). Our system performs well in both diseases, proving that mmTremor captures characteristics for both tremor types. Here the performance for ET is slightly worse than PD, probably due to the lack of tremor samples from only 4 ET patients. It is promising to significantly improve the performance of ET by extending our dataset with more reasonable ET tremor samples.

DBS Status	Predicted Duration (%)	Ground Truth (%)
ON	2.09	2.93
OFF	100	100

Table 2: Duration estimation for different DBS statuses.

6.5 Computational Cost

Though trained on a GPU RTX 3090 server, mmTremor is expected to perform inference with low computational cost. Thus mmTremor, we conduct inference measurement on the trained *DeepTremor* by averaging 1000 samples with one RTX 3090 GPU. The inference latency is 31.29 ms and the GPU memory consumption is about 2920 MB. This proves that mmTremor is affordable for most consumer-grade GPUs in real-time. It achieves such low consumption thanks to our dedicated design in the training stage optimization.

7 DISCUSSION AND FUTURE WORK

Privacy Concerns. While the depth camera largely mitigates the privacy issues by capturing only the user's silhouette, privacy concerns remain, like potential identity disclosure [40]. In our system, these images are processed into skeleton and cropped hand images locally and will not appear in final reports to clinicians, minimizing the risk of transmitting sensitive user information to external organizations. For adaptation, we think privacy protection methods like differential privacy [59] can further reduce this issue.

Further Analysis on Tremor mmTremor represents a significant advancement in practical tremor monitoring by enabling accurate tremor detection in everyday life. While mmTremor already offers clinically relevant metrics such as tremor duration and frequency, there remains a need to capture additional biomarkers of tremor. Firstly, tremor amplitude, a key criterion in the MDS-UPDRS, is critical for assessing Parkinson's disease progression. mmTremor shows promise in contactless assessment of amplitude through our extracted tremor features. Secondly, tremors typically initiate in the upper limbs before spreading to the head and other extremities. mmTremor has the potential to detect and track tremors across various body regions. Thirdly, as tremors are common in multiple neurological conditions, misdiagnosis rates are notable. By conducting in-depth frequency analysis and evaluating other motor characteristics, mmTremor holds the potential to aid in diagnosing the underlying causes of tremors, thereby enhancing treatment effectiveness.

Generalizability. We collect the first multi-modal dataset consisting of 28 patients with PD and ET across diverse environments, covering disease histories from 2 to over 20 years. However, our dataset inevitably has limitations, as most patients exhibit mild hand tremors. In future efforts, we aim to expand the dataset to include a larger, demographically

diverse population with varying tremor severities, different affected limbs, and more daily conditions that present tremors. By doing so, we envision developing a comprehensive and practical tremor assessment tool to enhance current diagnosis, therapy, and rehabilitation practices.

Broader Scenarios Our focus on detecting tremors in seated scenarios arises from the understanding that a significant portion of daily activities occur while seated, such as eating, writing, and using electronic devices. By capturing tremor data during regularly performed routine activities like eating, our system facilitates a more scientific evaluation of tremor symptoms. In contrast to prior studies that often concentrate on fixed postures and limited actions, our research offers a more comprehensive insight into how tremors impact the daily lives of individuals. Looking forward, our goal is to broaden our research scope to encompass whole-home monitoring, incorporating a wider range of daily activities. This expansion will enable the development of comprehensive contactless solutions tailored specifically for the elderly.

Extension to Other Symptoms As a major symptom, tremor significantly affects the quality of daily life. Our system focuses on tremor modeling based on the observation that tremor introduces abnormal variations compared with normal cases. Based on this observation, our system holds the potential to generalize to additional motion abnormalities like bradykinesia and rigidity, which are common biomarkers in elderly and PD patients. In our future work, we will explore additional motion symptoms based on our hardware and build a comprehensive home care system for early detection and monitoring of abnormalities.

8 RELATED WORKS

Wearable-based Tremor Monitoring. Electroencephalography (EEG) and electromyography (EMG) have been extensively utilized for tremor assessment, capturing signals from the nervous system and muscles during tremor [14, 38]. Surface EMG, in particular, serves as the gold standard for tremor diagnosis. However, its application for continuous monitoring in daily life is impractical [57]. Recently, IMUs have emerged as potential solutions for daily tremor monitoring and assessment. These include the wearable systems worn on the limbs [3, 11] and those integrated into smartwatches or smartphones [2, 32]. Despite their promise, wearable devices face challenges like discomfort with prolonged wear [23] and poor long-term adherence [50]. Additionally, frequent charging requirements further diminish their practicality for continuous monitoring.

Contactless Tremor Monitoring. Numerous RGB camera-based methods [33, 45, 61] and depth camera methods [4, 8] have been explored. However, RGB camera solutions raise privacy concerns and are inadequate for non-line-of-sight scenarios. Moreover, depth cameras perform poorly in dynamic environments [44], limiting their feasibility for tremor monitoring in daily activities. RF solutions have also been explored. Blumrosen et al. [7] attempt tremor quantification with an ultra-wideband radar, despite their simulations with a metal plate that mimics tremor behaviors. Lin et al. [31] leverage a dedicated Doppler radar to quantify tremors in fixed postures. The work [19] employs an FMCW radar for tremor detection. Some studies have explored daily tremor monitoring [20, 63] but only involve simulated tremors with fixed tremor sources and lack validation in real-world patients. In contrast, mmTremor not only resolves dynamic tremors under huge ADL interference but also is validated on patients with PD and ET in home settings.

mmWave-based Tiny Biomarker Monitoring. Continuous monitoring of tiny human biomarkers is a key application for mmWave sensing. ViMo [58] explores channel impulse responses for multi-user vital sign monitoring. RF-SCG [21] reconstructs seismocardiogram waveforms, while VED [65] generates blood pulse waveforms. MoVi-Fi [10] extracts respiration and heartbeat waveforms despite motion interference. WaveBP [27] estimates the arterial blood pressure waveform. However, these methods rely on strong reflections from the chest or static areas. In contrast, mmTremor tracks dynamic hand movements and extracts tremor patterns using weak signals reflected from hands, even amidst significant movement interference.

9 CONCLUDING REMARKS

This paper presents mmTremor, the first practical contactless tremor detection system, designed to assess medication effectiveness for patients with PD and ET. To achieve this, we propose a multi-modal fusion pipeline, including the mmWave-depth tracking for tremor source separation, *DeepTremor* for extracting micro tremor features from non-linear large activity interference, and ConAda for unsupervised domain adaptation. Extensive evaluations demonstrate the capability to accurately track tremor conditions with high correlations of medical assessment and contact sensors. mmTremor shows great promise for wide deployment in real-world scenarios to facilitate in-home healthcare for target patients.

ACKNOWLEDGMENTS

The authors would like to thank all the anonymous reviewers for their insightful comments. We also thank the Turing AI Computing Cloud (TACC) [62] and HKUST iSING Lab for computation resources on their platform. This research is supported in part by RGC under Contract CERG 16206122, 16204523, AoE/E-601/22-R, R6021-20, and Contract R8015. Dr Qian Zhang, Dr Guihua Li, and Dr Qianhui Xu are the corresponding authors.

REFERENCES

[1] Karim Abouelmehdi, Abderrahim Beni-Hessane, and Hayat Khaloufi. 2018. Big healthcare data: preserving security and privacy. *Journal of big data* 5, 1 (2018), 1–18.

[2] Noreen Akram, Haoxuan Li, Aaron Ben-Joseph, Caroline Budu, David A Gallagher, Jonathan P Bestwick, Anette Schrag, Alastair J Noyce, and Cristina Simonet. 2022. Developing and assessing a new web-based tapping test for measuring distal movement in Parkinson's disease: a Distal Finger Tapping test. *Scientific reports* 12, 1 (2022), 386.

[3] Sheik Mohammed Ali, Sridhar Poosapadi Arjunan, James Peters, Laura Perju-Dumbrava, Catherine Ding, Michael Eller, Sanjay Raghav, Peter Kempster, Mohammod Abdul Motin, PJ Radcliffe, et al. 2022. Wearable sensors during drawing tasks to measure the severity of essential tremor. *Scientific reports* 12, 1 (2022), 5242.

[4] Mehmet Akif Alper, John Goudreau, and Morris Daniel. 2020. Pose and Optical Flow Fusion (POFF) for accurate tremor detection and quantification. *Biocybernetics and Biomedical Engineering* 40, 1 (2020), 468–481.

[5] Pierre Baldi, Søren Brunak, Yves Chauvin, Claus AF Andersen, and Henrik Nielsen. 2000. Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* 16, 5 (2000), 412–424.

[6] Sergi Barrantes, Antonio J Sánchez Egea, Hernán A González Rojas, Maria J Martí, Yaroslau Compta, Francesc Valldeoriola, Ester Simo Mezquita, Eduard Tolosa, and Josep Valls-Solè. 2017. Differential diagnosis between Parkinson's disease and essential tremor using the smartphone's accelerometer. *PloS one* 12, 8 (2017), e0183843.

[7] G Blumrosen, M Uziel, B Rubinsky, and D Porrat. 2010. Non-contact UWB radar technology to assess tremor. In *XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010: May 27–30, 2010 Chalkidiki, Greece*. Springer, 490–493.

[8] Abdul H Butt, E Rovini, C Dolciotti, G De Petris, P Bongioanni, MC Carboncini, and F Cavallo. 2018. Objective and automatic classification of Parkinson disease with Leap Motion controller. *Biomedical engineering online* 17 (2018), 1–21.

[9] Chun-Fu (Richard) Chen, Quanfu Fan, and Rameswar Panda. 2021. CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification. In *ICCV*. IEEE, 347–356.

[10] Zhe Chen, Tianyue Zheng, Chao Cai, and Jun Luo. 2021. MoVi-Fi: motion-robust vital signs waveform recovery via deep interpreted RF sensing. In *MobiCom*. ACM, 392–405.

[11] Houde Dai, Pengyue Zhang, and Tim C. Lueth. 2015. Quantitative Assessment of Parkinsonian Tremor Based on an Inertial Measurement Unit. *Sensors* 15, 10 (2015), 25055–25071.

[12] Tsviya Fay-Karmon, Noam Galor, Benedetta Heimler, Asaf Zilka, Ronny P Bartsch, Meir Plotnik, and Sharon Hassin-Baer. 2024. Home-based monitoring of persons with advanced Parkinson's disease using smartwatch-smartphone technology. *Scientific Reports* 14, 1 (2024), 9.

[13] Parkinson's Foundation. [n. d.]. Prevalence & Incidence of Parkinson's disease. https://www.parkinson.org/understanding-parkinsons/statistics/prevalence-incidence

[14] Luay Fraiwan, Ruba Khnouf, and Abdel Razaq Mashagbeh. 2016. Parkinson's disease hand tremor detection system for mobile application. *Journal of medical engineering & technology* 40, 3 (2016), 127–134.

[15] Maximilian U. Friedrich, Anna-Julia Roenn, Chiara Palmisano, Jane Alty, Steffen Paschen, Günther Deuschl, Chi Wang Ip, Jens Volkmann, Muthuraman Muthuraman, Robert L. Peach, and Martin M. Reich. 2024. Validation and application of computer vision algorithms for video-based tremor analysis. *npj Digit. Medicine* 7, 1 (2024).

[16] Takako Fujita, Akira Babazono, Sung-a Kim, Aziz Jamal, and Yunfei Li. 2021. Effects of physician visit frequency for Parkinson's disease treatment on mortality, hospitalization, and costs: a retrospective cohort study. *BMC geriatrics* 21, 1 (2021), 707.

[17] Brook Galna, Gillian Barry, Dan Jackson, Dadirayi Mhiripiri, Patrick Olivier, and Lynn Rochester. 2014. Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease. *Gait & posture* 39, 4 (2014), 1062–1068.

[18] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor S. Lempitsky. 2016. Domain-Adversarial Training of Neural Networks. *J. Mach. Learn. Res.* 17 (2016), 59:1–59:35.

[19] Nazia Gillani and Tughrul Arslan. 2021. Unobtrusive detection and monitoring of tremors using non-contact radar sensor. In *2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART)*. IEEE, 01–04.

[20] Nazia Gillani, Tughrul Arslan, and Gillian Mead. 2023. An Unobtrusive Method for Remote Quantification of Parkinson's and Essential Tremor using mm-Wave Sensing. *IEEE Sensors Journal* (2023).

[21] Unsoo Ha, Salah Assana, and Fadel Adib. 2020. Contactless seismocardiography via deep learning radars. In *MobiCom*. ACM, 62:1–62:14.

[22] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross B. Girshick. 2020. Momentum Contrast for Unsupervised Visual Representation Learning. In *CVPR*. Computer Vision Foundation / IEEE, 9726–9735.

[23] Margot Heijmans, Jeroen GV Habets, Christian Herff, Jos Aarts, An Stevens, Mark L Kuijf, and Pieter L Kubben. 2019. Monitoring Parkinson's disease symptoms during daily life: a feasibility study. *npj Parkinson's Disease* 5, 1 (2019), 21.

[24] Dustin A Heldman, Joseph Jankovic, David E Vaillancourt, Janey Prodoehl, Rodger J Elble, and Joseph P Giuffrida. 2011. Essential tremor quantification during activities of daily living. *Parkinsonism & related disorders* 17, 7 (2011), 537–542.

[25] Margaret M Hoehn and Melvin D Yahr. 1967. Parkinsonism: onset, progression, and mortality. *Neurology* 17, 5 (1967), 427–427.

[26] Qingyong Hu, Hua Kang, Huangxun Chen, Qianyi Huang, Qian Zhang, and Min Cheng. 2023. CSI-StripeFormer: Exploiting Stripe Features for CSI Compression in Massive MIMO System. In *INFOCOM*. IEEE, 1–10.

[27] Qingyong Hu, Qian Zhang, Hao Lu, Shun Wu, Yuxuan Zhou, Qianyi Huang, Huangxun Chen, Ying-Cong Chen, and Ni Zhao. 2024. Contactless Arterial Blood Pressure Waveform Monitoring with mmWave Radar. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 4 (2024), 178:1–178:29.

[28] Anas Ibrahim, Yue Zhou, Mary E Jenkins, Ana Luisa Trejos, and Michael D Naish. 2021. Real-time voluntary motion prediction and Parkinson's tremor reduction using deep neural networks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29 (2021), 1413–1423.

[29] Tao Jiang, Peng Lu, Li Zhang, Ningsheng Ma, Rui Han, Chengqi Lyu, Yining Li, and Kai Chen. 2023. RTMPose: Real-Time Multi-Person Pose Estimation based on MMPose. *CoRR* abs/2303.07399 (2023).

[30] Pieter L Kubben, Mark L Kuijf, Linda PCM Ackermans, Albert FG Leentjes, and Yasin Temel. 2016. TREMOR12: An open-source mobile app for tremor quantification. *Stereotactic and functional neurosurgery* 94, 3 (2016), 182–186.

[31] Chia-Hung Lin, Jian-Xing Wu, Jin-Chyr Hsu, Pi-Yun Chen, Neng-Sheng Pai, and Hsiang-Yueh Lai. 2021. tremor class scaling for Parkinson disease patients using an array X-band microwave Doppler-based upper limb movement quantizer. *IEEE sensors journal* 21, 19 (2021), 21473–21485.

[32] Florian Lipsmeier, Kirsten I Taylor, Ronald B Postuma, Ekaterina Volkova-Volkmar, Timothy Kilchenmann, Brit Mollenhauer, Atieh Bamdadian, Werner L Popp, Wei-Yi Cheng, Yan-Ping Zhang, et al. 2022. Reliability and validity of the Roche PD Mobile Application for remote monitoring of early Parkinson's disease. *Scientific reports* 12, 1

(2022), 12081.

[33] Weiping Liu, Xiaozhen Lin, Xinghong Chen, Qing Wang, Xiumei Wang, Bin Yang, Naiqing Cai, Rong Chen, Guannan Chen, and Yu Lin. 2023. Vision-based estimation of MDS-UPDRS scores for quantifying Parkinson's disease tremor severity. *Medical Image Anal.* 85 (2023), 102754.

[34] Elan D Louis and Duarte G Machado. 2015. Tremor-related quality of life: a comparison of essential tremor vs. Parkinson's disease patients. *Parkinsonism & related disorders* 21, 7 (2015), 729–735.

[35] Kenneth A Mann, Frederick W Wernere, and Andrew K Palmer. 1989. Frequency spectrum analysis of wrist motion for activities of daily living. *Journal of Orthopaedic research* 7, 2 (1989), 304–306.

[36] Microsoft. [n. d.]. Azure Kinect DK hardware specifications. [Online]. https://learn.microsoft.com/en-us/previous-versions/azure/kinect-dk/hardware-specification.

[37] mmTremor. 2024. Dataset link. https://drive.google.com/drive/folders/1xBptp7txXTQ59h-TUwRqTq-U2WBD8oWq?usp=sharing

[38] Morteza Moazami-Goudarzi, Johannes Sarnthein, Lars Michels, Renata Moukhtieva, and Daniel Jeanmonod. 2008. Enhanced frontal low and high frequency power and synchronization in the resting EEG of parkinsonian patients. *NeuroImage* 41, 3 (2008), 985–997.

[39] Fábio Henrique Monteiro Oliveira, Daniel Fernandes da Cunha, Amanda Gomes Rabelo, Luiza Maire David Luiz, Marcus Fraga Vieira, Adriano Alves Pereira, and Adriano de Oliveira Andrade. 2021. A non-contact system for the assessment of hand motor tasks in people with Parkinson's disease. *SN Applied Sciences* 3 (2021), 1–11.

[40] Wiktor Mucha and Martin Kampel. 2022. Addressing privacy concerns in depth sensors. In *International conference on computers helping people with special needs*. Springer, 526–533.

[41] Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease. 2003. The unified Parkinson's disease rating scale (UPDRS): status and recommendations. *Movement Disorders* 18, 7 (2003), 738–750.

[42] Xiaomin Ouyang, Xian Shuai, Yang Li, Li Pan, Xifan Zhang, Heming Fu, Sitong Cheng, Xinyan Wang, Shihua Cao, Jiang Xin, Hazel Mok, Zhenyu Yan, Doris Sau-Fung Yu, Timothy Kwok, and Guoliang Xing. 2024. ADMarker: A Multi-Modal Federated Learning System for Monitoring Digital Biomarkers of Alzheimer's Disease. In *MobiCom*. ACM, 404–419.

[43] Yilmazcan Özyurt, Stefan Feuerriegel, and Ce Zhang. 2023. Contrastive Learning for Unsupervised Domain Adaptation of Time Series. In *ICLR*. OpenReview.net.

[44] Alvaro Lopez Paredes, Qiang Song, and Miguel Heredia Conde. 2023. Performance Evaluation of State-of-the-Art High-Resolution Time-of-Flight Cameras. *IEEE Sensors Journal* (2023).

[45] Silvia L. Pintea, Jian Zheng, Xilin Li, Paulina J. M. Bank, Jacobus J. van Hilten, and Jan C. van Gemert. 2018. Hand-Tremor Frequency Estimation in Videos. In *ECCV Workshops (6) (Lecture Notes in Computer Science, Vol. 11134)*. Springer, 213–228.

[46] Walter Pirker, Regina Katzenschlager, Mark Hallett, and Werner Poewe. 2023. Pharmacological treatment of tremor in Parkinson's disease revisited. *Journal of Parkinson's disease* 13, 2 (2023), 127–144.

[47] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *MICCAI (3) (Lecture Notes in Computer Science, Vol. 9351)*. Springer, 234–241.

[48] Arash Salarian, Heike Russmann, Christian Wider, Pierre R. Burkhard, François J. G. Vingerhoets, and Kamiar Aminian. 2007. Quantification of Tremor and Bradykinesia in Parkinson's Disease Using a Novel Ambulatory Monitoring System. *IEEE Trans. Biomed. Eng.* 54, 2 (2007), 313–322.

[49] Rubén San Segundo, Ada Zhang, Alexander Cebulla, Stanislav Panev, Griffin Tabor, Katelyn Stebbins, Robyn E. Massa, Andrew S. Whitford, Fernando De la Torre, and Jessica K. Hodgins. 2020. Parkinson's Disease Tremor Detection in the Wild Using Wearable Accelerometers. *Sensors* 20, 20 (2020), 5817.

[50] Ana Lígia Silva de Lima, Tim Hahn, Luc JW Evers, Nienke M De Vries, Eli Cohen, Michal Afek, Lauren Bataille, Margaret Daeschler, Kasper Claes, Babak Boroojerdi, et al. 2017. Feasibility of large-scale deployment of multiple wearable sensors in Parkinson's disease. *PloS one* 12, 12 (2017).

[51] Peige Song, Yan Zhang, Mingming Zha, Qingwen Yang, Xinxin Ye, Qian Yi, and Igor Rudan. 2021. The global prevalence of essential tremor, with emphasis on age and sex: a meta-analysis. *Journal of global health* 11 (2021).

[52] Yonglong Tian, Chen Sun, Ben Poole, Dilip Krishnan, Cordelia Schmid, and Phillip Isola. 2020. What Makes for Good Views for Contrastive Learning?. In *NeurIPS*.

[53] Michal Tölgyessy, Martin Dekan, and L'uboš Chovanec. 2021. Skeleton tracking accuracy and precision evaluation of kinect v1, kinect v2, and the azure kinect. *Applied Sciences* 11, 12 (2021), 5756.

[54] Zhan Tong, Yibing Song, Jue Wang, and Limin Wang. 2022. VideoMAE: Masked Autoencoders are Data-Efficient Learners for Self-Supervised Video Pre-Training. In *NeurIPS*.

[55] Du Tran, Heng Wang, Lorenzo Torresani, Jamie Ray, Yann LeCun, and Manohar Paluri. 2018. A Closer Look at Spatiotemporal Convolutions for Action Recognition. In *CVPR*. Computer Vision Foundation / IEEE Computer Society, 6450–6459.

[56] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008).

[57] Basilio Vescio, Andrea Quattrone, Rita Nisticò, Marianna Crasà, and Aldo Quattrone. 2021. Wearable devices for assessment of tremor. *Frontiers in Neurology* 12 (2021), 680011.

[58] Fengyu Wang, Feng Zhang, Chenshu Wu, Beibei Wang, and KJ Ray Liu. 2020. ViMo: Multiperson vital sign monitoring using commodity millimeter-wave radio. *IEEE Internet of Things Journal* 8, 3 (2020), 1294–1307.

[59] Qian Wang, Zixi Li, Qin Zou, Lingchen Zhao, and Song Wang. 2020. Deep Domain Adaptation With Differential Privacy. *IEEE Trans. Inf. Forensics Secur.* 15 (2020), 3093–3106.

[60] Wenhui Wang, Hangbo Bao, Li Dong, Johan Bjorck, Zhiliang Peng, Qiang Liu, Kriti Aggarwal, Owais Khan Mohammed, Saksham Singhal, Subhojit Som, and Furu Wei. 2023. Image as a Foreign Language: BEIT Pretraining for Vision and Vision-Language Tasks. In *CVPR*. IEEE, 19175–19186.

[61] Stefan Williams, Hui Fang, Samuel D Relton, David C Wong, Taimour Alam, and Jane E Alty. 2021. Accuracy of smartphone video for contactless measurement of hand tremor frequency. *Movement Disorders Clinical Practice* 8, 1 (2021), 69–75.

[62] Kaiqiang Xu, Decang Sun, Hao Wang, Zhenghang Ren, Xinchen Wan, Xudong Liao, Zilong Wang, Junxue Zhang, and Kai Chen. 2025. Design and Operation of Shared Machine Learning Clusters on Campus. In *ASPLOS (1)*. ACM, 295–310.

[63] Xiaodong Yang, Syed Aziz Shah, Aifeng Ren, Dou Fan, Nan Zhao, Shufeng Zheng, Wei Zhao, Weigang Wang, Ping Jack Soh, and Qammer H Abbasi. 2018. *S*-band sensing-based motion assessment framework for cerebellar dysfunction patients. *IEEE Sensors Journal* 19, 19 (2018), 8460–8467.

[64] Hongyi Zhang, Moustapha Cissé, Yann N. Dauphin, and David Lopez-Paz. 2018. mixup: Beyond Empirical Risk Minimization. In *ICLR (Poster)*. OpenReview.net.

[65] Shujie Zhang, Tianyue Zheng, Zhe Chen, and Jun Luo. 2022. Can We Obtain Fine-grained Heartbeat Waveform via Contact-free RF-sensing?. In *INFOCOM*. IEEE, 1759–1768.