Manuscript title: Ultra-Minimal Strain-Based Tactile HCI Device for Force/Position Interaction in Severe Upper-Limb Impairments

Replies to the Comments of Editors and Reviewers

Dear Editors and Reviewers:

We are very glad to receive the encouraging comments from editors and anonymous reviewers. At first, we would like to thank all the reviewers and editors for their time and effort in moving this manuscript to the public. Those comments are very valuable and helpful for improving our paper, as well as the important guiding significance to our research work. Based on these valuable comments and suggestions, we have done a comprehensive revision on the manuscript. Moreover, we have carefully checked the whole manuscript to ensure the high-level writing quality. The detailed responses about these comments (point-to-point) are enclosed below, indicating exactly how we addressed each concern or problem and describing the changes we have made. Great thanks to you for the time and efforts expand on this paper.

With best wishes to the editors and reviewers.

Yours sincerely,
Best wishes,

All authors

* *

Associate Editor

Comments to the Author:

Four reviewers have evaluated this manuscript. While the system demonstrates practical potential, the work remains incomplete. The mian issues include insufficient user validation for the target population, inconsistent reporting of key performance metrics, inadequate technical description and validation of the proposed PFG-CMNet model, and absence of comprehensive benchmarking in the calibration method. The language is confusing and English grammar also needs further polishing. On the basis of the reviewers' ratings and comments, the AE cannot recommend the acceptance of this manuscript in this top-tier

Response: We really appreciate your time and effort for the publication of this manuscript in TII. The affirmation and encouragement of our manuscript from associate editor, editors and reviewers inspiredus greatly. Based on these valuable comments and suggestions, we have made a comprehensive and deep revision in the manuscript. The revisions include, but not limited to: (1) conducting more comprehensive user studies with the target population; (2) standardizing key performance metrics and providing clear definitions; (3) offering a more detailed technical description of PFG-CMNet with thorough validation; (4) presenting the theoretical basis of the calibration method and performing comprehensive benchmarking; and finally, further refining the manuscript’ s grammar and style to meet the requirements of TII. The detailed replies to reviewers (point to point) have been shown below. Thanks again for your affirmation.

Reply to reviewer 1

Reviewer #1: Comments to the Author

This manuscript presents the design, development, and validation of an Ultra-Minimal Strain-based Tactile System for human-computer interaction, specifically targeting individuals with severe upper-limb impairments. The core achievement is the successful balancing of three often conflicting goals: high functionality, structural simplicity, and low cost. The authors employ a topology-optimized sensor layout, a novel hierarchical CNN-MLP network trained primarily on a large Finite Element Method-generated dataset, and a Calibration Compensation Matrix to bridge the simulation-to-reality gap. Experimental results demonstrate high accuracy, real-time performance, and effectiveness in practical HCI applications like game control.

Response:

Thank you very much for your affirmation of our research work. According to reviewer’s comments, we have carefully analyzed all concerns and problems from reviewer and made a major revision on the manuscript. These comments are insightful and constructive, which are very helpful for improving the quality of this manuscript. Moreover, the rigorous and scientific attitude expressed in these comments deeply influences our revision and benefits us a lot. We believe that these valuable suggestions will further motivate us to do better in the future research. The detailed replies to the reviewer have been shown as follows. Sincerely, thanks again for the reviewer’s support and encouragement of our research work!

Comment 1: A significant limitation is that all experiments and application demonstrations were conducted by the research team. The core premise is to assist users with "severe upper-limb impairments," whose motor capabilities are not represented in the current validation. It is strongly recommended to include a preliminary user study with even a small number of individuals from the target population, or to simulate impairment conditions, to provide initial data on usability and robustness.

Response:

We are genuinely grateful for this warm and constructive recommendation, and we apologize for the limitations of our initial validation. We fully agree that evaluating the target population is essential to establish UMSTD’s practical value. We have already initiated the institutional review process in collaboration with our clinical partner hospital; because the study involves human participants with severe upper-limb impairment, the approvals are necessarily extensive and are currently in progress.

To make immediate progress, we conducted a focused literature review and a hospital site assessment (including a myasthenia gravis case), from which we distilled a five-category set of generic single-finger HCI actions for this population (Table II, Page. 8). We then adopted a simulated-impairment protocol by bandaging operators’ fingers to enforce single-finger use (Fig. 12a) and designed two application classes reflecting the surveyed needs. First, entertainment-based rehabilitation comprises three stage-progressive tasks aligned with increasing motor difficulty: TikTok browsing to elicit low-rate primitives (Fig. 12b, Supplementary Movie Experiment 2.1), Contra to train high-frequency basic actions (Fig. 12c, Supplementary Video Experiment 2.2), and The King of Fighters’97 to practice high-frequency composite actions (Fig. 12d, Supplementary Movie Experiment 2.3). Second, for everyday device control, we implemented single-finger, gesture-driven mode switching and used a serial smart light as a controlled proxy to emulate common home controls: CW/CCW mode selection, tap/press for power, color, and buzzer duration, and swipe for continuous brightness adjustment; an alert action (long-press) triggers a buzzer for emergency assistance (Figs. 12e-f, Supplementary Movie Experiment 3). Related usability and robustness demonstrations are provided in the Supplementary Movie. We believe these additions directly address the reviewer' s concern while we complete the formal user study with the target population.

Comment 2: A critical discrepancy exists between the abstract and main text regarding force measurement error. The abstract claims a force error of < 0.14 N, while Section V-A reports maximum absolute errors of 0.3 N for Fz and 0.5 N for Fx/Fy. This direct contradiction in a key performance metric undermines the manuscript's credibility. The authors are strongly encouraged to reconcile these values and ensure consistent reporting throughout the paper.

Response:

We thank the reviewer for the kind suggestion. We are so sorry what our description caused your misunderstanding. In the abstract, the "force error < 0.14 N" refers to the average (mean absolute) force error, whereas the 0.3 N and 0.5 N values in the main text denote the maximum absolute force errors. We have revised the manuscript accordingly to unify terminology and eliminate confusion.

Comment 3: The contribution of the proposed PFG-CMNet architecture is not sufficiently validated due to the absence of comparisons with baseline models.

Response:

We are deeply grateful for the reviewer's thoughtful and generous suggestions, which have been invaluable to improving our work. We truly appreciate your careful reading and constructive guidance. Accordingly, we have: (i) added ablation studies to isolate the contribution of each component and to clarify performance gains relative to baseline models (see Table I, Page. 7); (ii) expanded the technical description of PFG-CMNet to enhance reproducibility (see Section III-B, Page. 6); and (iii) compared our system's capabilities with representative prior work to quantitatively assess performance (see Table III, Page. 9). These changes are clearly marked in the revised manuscript. Thank you again for the insightful feedback.

Comment 4: The manuscript requires careful proofreading to address grammatical errors and improve clarity. For instance, on Page 6 in the "Deep Learning-Based Decoupling" section, the sentence "The train is conducted..." incorrectly uses the verb "train" as a noun and should be

revised to "Training is conducted...". Furthermore, on Page 3 in the "Operating Principle" section, the sentence fragment "Refers to the bridge excitation voltage..." lacks a subject and should be rephrased as "The symbol V_cc refers to the bridge excitation voltage..." to ensure grammatical correctness and maintain consistency with the terminology used in Equation (1).

Response:

We really appreciate reviewer's kind suggestion. We conducted a thorough proofreading of the manuscript. Numerous grammar and clarity improvements have been made. For example, the opening line of the conclusion (Section VI, Page 9) was rewritten into a single sentence for coherence ("To address... we propose..."). In Section II-B (Page 3), we fixed the incomplete phrase ("Refers to the bridge excitation...") by stating "Vcc (bridge excitation voltage) is set to 2.8 V, allowing..." and similarly corrected "In the Eq.(1)" to "In Eq. (1)". We also ensured that variables are clearly defined around Eq. (3) (see Section III-A, Page 4: we now explain S_{i-x}, S_{i-y} as sensor coordinates). These are just a few examples; overall, sentence structures have been polished throughout the paper for clarity and correctness. These are just a few examples; overall, sentence structures have been polished throughout the paper for clarity and correctness.

****

Reply to reviewer 2

Reviewer #2: Comments to the Author

This paper proposes an Ultra-Miniature Strain-based Tactile System (UMSTD) for human-computer interaction, aiming to address the challenge of "balancing structural simplicity with functional richness" in HCI for individuals with severe upper limb dysfunction. The system achieves high-precision 3D force/position estimation and gesture recognition with low channel counts through topology-optimized sensor layout, the construction of a PFG-CMNet hierarchical decoupling network, and the design of a CCM calibration compensation matrix. Its performance is validated through benchmark tests, trajectory tracking, and game interaction experiments. The research methodology is innovative, with relatively sufficient experimental data and convincing results.

Response:

Thank you very much for your affirmation of our previous revision. Reviewer's suggestions are very professional and constructive, which significantly help us improve the writing quality of this manuscript. The professional skills and rigorous attitude expressed in these comments carved deep into our revision and benefited us a lot. Based on reviewer’s comments, we have made a deep revision in the manuscript. The detailed responses are shown as follows.

Comment 1:

In the introduction, the review of existing HCI modalities (FRAs, LAs) and strain-based interfaces only lists some representative studies, lacking references to and analysis of the latest research (2023–2024) in related fields.

Response:

We sincerely appreciate this insightful observation, and we agree it is crucial for the Introduction to reflect the most recent advances so that our contributions are positioned fairly and rigorously against the current state of the art. Accordingly, Section I (Page. 2) now expands the discussion of recent HCI and strain-sensing work, adding 2023 to 2024 references [16], [17], [22], [23] and summarizing their contributions, for example a 2024 multi-axis sensor review [16], a state-of-the-art MEMS tactile sensor for surgical use [17], and a 2024 wearable hydrogel strain sensor for HCI [22], as well as emerging minimalistic interfaces such as body-coupled and two-electrode HCI. These updates modernize the literature review and situate our work within the latest developments. We thank the reviewer for prompting this important improvement.

Comment 2:

Abbreviations such as "PFG-CMNet" and "FESFI-14K" are mentioned in the abstract without providing their full forms upon first appearance. It is recommended to supplement these.

Response: We sincerely thank the reviewer for this warm and helpful suggestion, and we apologize for our oversight. We have revised the abstract to spell out all abbreviations at first mention, including PFG-CMNet and FESFI-14K, and ensured consistent usage throughout the manuscript. We appreciate the reviewer’s careful attention to clarity.

Comment 3:

Derivations of some key formulas (e.g., the transformation formula for the CCM calibration matrix and the loss function calculation for PFG-CMNet) are missing, with only the final results provided.

Response: Thank you for the constructive comments. The revised Section IV-C now provides the full derivation pipeline rather than only final results. In particular, we use Rodrigues’ rotation to construct the SCIC correction for each triad, eliminating installation-induced orientation misalignment at the sensor level (principle diagram in Fig. 5). The derivation details the axis–angle construction, the per-sensor rotation $R_i \in SO(3)$, and the resulting block-diagonal composition that yields the CCM transformation; the complete steps and equations are given in Section IV-C.

Comment 4: The description of the PFG-CMNet structure is relatively vague. It is recommended to supplement with diagrams or more detailed explanations of the network architecture.

Response:

Thank you for the constructive comments. We agree and have rewritten Section IV-B and updated Fig. 5 to provide a precise, reproducible specification of PFG-CMNet. The revision details the CNN front end and residual stages with kernel size, stride, padding, activations and normalization; the 12-class localization head; the 12 expert MLPs with explicit layer widths and dropout; and the CNN–MLP coupling and routing strategy. We also include initialization, optimizer, and scaler statistics, and we release trained weights and an inference script so that the architecture can be reconstructed one-to-one from the text and figure.

For completeness, we also summarize the training data, pipeline, and reproducibility details as follows. We train on the FESFI-14K dataset with 14,000 concentrated-load samples; each sample records nine channels from three UMSTD sensors, using a 90/10 train–test split and feature scaling to [−1, 1]. PFG-CMNet uses a compact 1-D ResNet to assign each sample to one of 12 angular sectors, with 3-wide convolutions along the channel axis, global average pooling, and a 12-way linear head that outputs sector posteriors. Gated regression then routes the same 9-D input to one expert MLP per sector for fine estimation of X, Y, Fx, Fy, Fz; training employs temperature-sharpened top-2 soft routing for boundary continuity, while inference uses argmax with a top-2 fallback at low confidence. Each expert is FC 9–128–128–128–5 with ReLU and dropout 0.1. This split exploits near-linear per-sector mechanics and the statically determinate 9-channel layout, outperforming single-branch CNN/MLP baselines: force MAE drops from 0.38 ± 0.053 N to 0.18 ± 0.03 N and position MAE from 3.13 ± 0.16 mm to 2.42 ± 0.12 mm (Table I). The plain CNN baseline is faster (41 ms vs 52 ms) but does not yield valid regression. For gestures, eight classes (SHL, SHR, CW, SC, CCW, UC, SVU, SVD) with 100 samples each achieve 93.5% accuracy under the same protocol. Implementation uses Python 3.12 and PyTorch 2.6.0 on a single RTX 4070 (CUDA 12); Adam with MSE and a learning-rate sweep selects 1×10^{-4}; models converge within 300 epochs, with test MSE 0.014, MAE 0.085, and R^2 0.972.

Comment 5: The advantages and disadvantages of FRAs, LAs, and traditional strain-based

HCI systems are compared only through textual descriptions, without quantitative performance comparisons with recently improved strain-based HCI systems.

Response:

We thank the reviewer for comments. We have revised the **Introduction** to add quantitative comparisons and consolidated them **in a new summary table. Table III** now reports, for representative FRAs (ET/BCI/VBT), LAs-IMU, recent dense strain-array systems, and our UMSTD, the key metrics needed for fair comparison: channel count, spatial resolution, force MAE, update rate, decoupling approach, and calibration burden, together with notes on practical constraints. In addition, While protocols differ across publications, the table uses author-reported numbers under consistent definitions, allowing a quantitative view that situates UMSTD as offering competitive accuracy and rate with an order-of-magnitude lower channel count and calibration effort.

Comment 6: The application validation section only includes game interaction scenarios and does not involve real-life scenarios for the target population (e.g., daily appliance control, text input). The application scenarios are relatively limited.

Response:

We appreciate the comment and have expanded the application validation beyond game scenarios. In Section V-C and the Supplementary Movie, we analyzed the action taxonomy of single-finger HCI operations, as summarized in Table II, which comprises five categories. We introduce a simulated-impairment protocol (bandaged single-finger use) and evaluate two classes of applications. For rehabilitation, we design stage-progressive entertainment tasks that embed training to improve function or slow deterioration: Stage I TikTok browsing with taps and directional swipes; Stage II Contra for high-frequency basic actions; Stage III The King of Fighters ’97 for high-frequency composite actions.

For real-life scenarios, we add a daily appliance control task using a smart light. Single-finger gestures execute 2 mode switching mapped to device groups, swipe adjusts brightness continuously, and tap/press controls power, color, and buzzer duration. We report task completion and command recognition, together with positioning accuracy and real-time operation under multi-directional loading. Related usability and robustness demonstrations are provided in the Supplementary Movie.

We greatly appreciate your insightful comments and suggestions again, which have substantially improved our manuscript. We hope our revised manuscript and responses can address all of your concerns and comments. Thank you.

****

Reply to reviewer 3

Reviewer #3: Comments to the Author

This paper presents an ultra-minimal strain-based tactile HCI system (UMSTD) that achieves accurate 3D force/position estimation and gesture recognition with a minimal structural design. The authors address a highly relevant and challenging problem: enabling complex, stable, and affordable interaction for individuals with severe upper-limb impairments using low-DOF, low-channel strain sensing. The proposed framework integrates a topology-optimized sensing layout, a hierarchical CNN–MLP decoupling model (PFG-CMNet), and a calibration compensation matrix (CCM) trained on the FESFI-14K FEM-based dataset. The results are promising.

Response:

Thank you very much for your constructive and positive comments. To address your comments and concerns in full, we have significantly improved our manuscript and added a substantial amount of new data, analyses, discussions, and clarifications. In the following paragraphs, we will address your comments point-by-point.

Comment 1:

The PFG-CMNet is potentially novel, but its description is too brief. Key details such as number of layers, kernel sizes, activations, normalization, feature dimensions, and CNN–MLP connections are missing. Training details alone are insufficient for reproducibility, and most structural details are only shown in figure 5, which is not adequate for readers to fully understand or reconstruct the model.

Response:

Thank you for this important point. We have substantially expanded Section IV-B and updated Fig. 5 to include all architecture specifications needed for faithful reconstruction, beyond training details. Concretely, we now report: (i) exact CNN stack and dimensions: four Conv(9→32) layers with kernel 3/stride 1/padding 1, each followed by BatchNorm and ReLU; three residual stages with two units per stage at widths 32, 64, and 128 with 1×1 projection on width change; global average pooling; linear head 128 → 12 with softmax; (ii) CNN → MLP coupling: sector posteriors gate regression via a mixture-of-experts with temperature-sharpened top-2 soft routing during training and argmax with top-2 fallback at low confidence during inference; (iii) expert MLP specification per sector: FC 9→128→128→128→5 with ReLU and dropout 0.1, outputting (X, Y, Fx, Fy, Fz); L2 weight decay 1e-4; (iv) feature handling: nine-channel input, featurewise standardization to [−1, 1] before the CNN and reused by the experts; (v) initialization and optimization: He initialization, Adam, MSE regression plus cross-entropy for sector classification with label smoothing and entropy regularization; learning-rate selection procedure; batch size and epoch count. We also provide the scaler statistics, fixed seeds, trained weights for the CNN and all 12 experts, and an inference script to ensure one-to-one reproducibility.

For completeness we summarize the training/data pipeline now documented in the paper: the model is trained on FESFI-14K (14,000 concentrated-load samples) with a 90/10 train – test split; each sample records nine channels from three UMSTD sensors. The standardized 9-D vector feeds the CNN for 12-sector coarse localization, after which the same 9-D input is routed to the selected expert for fine regression. This split leverages near-linear per-sector mechanics in a statically determinate 9-channel layout and outperforms single-branch CNN/MLP baselines (Table I); while a plain CNN attains lower latency, it does not yield valid regression under our setting. Implementation details (Python 3.12, PyTorch 2.6.0, RTX 4070, CUDA 12) and final metrics are reported to complete reproducibility.

Comment 2:

The described CCM method indeed focuses on practical calibration primarily based on linear correction and empirical compensation, There is no fundamental theoretical contribution or new calibration model; the matrices STC, SCIC are derived empirically from measurements.

Response:

Thank you for the thoughtful critique. We clarify that our calibration is not purely empirical. The SCIC component is a geometric correction derived from principles using Rodrigues’ rotation to remove installation-induced orientation misalignment at the triad level; its derivation and implementation are given in Section IV-C with the principle diagram in Fig. 5. The STC matrix is obtained from the sensor transfer characteristics under small-strain linear elasticity and Wheatstone readout theory, yielding a closed-form linear map that is then identified from calibration loads. The overall CCM is the block-diagonal SCIC composed with the theory-based STC, providing a physically grounded transformation rather than an ad-hoc fit. For transparency, the specific material and load parameters used for STC identification are summarized in the new table and demonstrated in the Supplementary Video, and Section IV-C reports all intermediate steps needed to reproduce the matrices.

Comment 3:

The experimental evaluation is limited to loading condition with fixed force directions, focusing only on absolute force error, response speed and position error. Consequently, the manuscript does not provide a comprehensive assessment of the system’s robustness and force-decoupling performance under more complex loading conditions, such as multi-directional forces. It is recommended that the authors include such analyses or discuss their potential impact to strengthen the validation and generality of the proposed approach.

Response:

Thank you very much for your comments. We have added multi-directional loading experiments to assess robustness and force decoupling beyond fixed-direction loads. In Section V and Fig. 10 we sweep the tilt angle from 0° to 80° and vary the resultant magnitude |F| from 1 to 5 N, reporting resultant-force RMSE and directional error across directions. The RMSE increases smoothly from ≈0.14 N near normal loading to ≈0.30 N at large tilt, while the direction error remains bounded and decreases with |F|, with the worst case ≈20° at |F| = 1 N and ≈4° at |F| = 5 N. We also retain absolute 3-axis accuracy results under 0–5 N (max abs. error 0.3 N for Fz, 0.5 N for Fx/Fy) and show stable predictions when extrapolated to 5–10 N. Serial monitoring under 0.9, 2.5, and 4.8 N shows 40–50 Hz update rates, meeting real-time requirements.

Together, these additions provide a quantitative, direction-swept evaluation of robustness and decoupling performance, complementing fixed-direction tests and demonstrating stable behavior under multi-directional forces. We also discuss the impact and remaining limitations of the current training range and outline future extensions to broaden force magnitudes and directions.

Comment 4: While Table I provides a qualitative comparison of various HCI systems, the evaluation is largely subjective and lacks quantitative benchmarking (such as resolution, range, accuracy, hysteresis and repeatability) against existing strain-based or other HCI approaches.

Response:

We have added quantitative benchmarking and ablations. First, the Introduction now cites recent strain-based and alternative HCI systems and we consolidate comparable metrics in Table III, including channel count, spatial resolution, force error (MAE), update rate, decoupling approach, and calibration burden; these values are taken from the respective papers under consistent definitions. Where range, hysteresis, or repeatability are reported by prior work, they are included in the notes; Second, Table I provides controlled ablations on our dataset: CNN-only, MLP-only, MLP+CCM, routing-only PFG-CMNet, and the full PFG-CMNet+CCM. The proposed architecture reduces error from 3.13±0.16 mm and 0.38±0.053 N (MLP) to 2.42±0.12 mm and 0.18±0.03 N, and with CCM achieves 1.83±0.09 mm and 0.14±0.02 N at 52 ms, thereby validating both the architectural choice and the contribution beyond qualitative discussion. Table III summarizes HCI approaches and contrasts our work with prior FRA and LA systems. The consolidated metrics show that UMSTD offers competitive accuracy and update rate with an order-of-magnitude fewer channels and markedly lower calibration effort.

Comment 5: No ablation studies are conducted to isolate the effects of CNN, MLP, and CCM, leaving the contribution of each component unclear.

Response:

Thank you for raising this point. We have added controlled ablations that isolate the effect of each component; results are reported in Table I (internal baselines) and summarized against external modalities in Table III. Under identical data, signals, and protocol: (i) MLP-only yields 3.13 ± 0.16 mm and 0.38 ± 0.053 N at 43 ms; (ii) MLP+CCM improves to 2.37 ± 0.12 mm and 0.29 ± 0.04 N, quantifying CCM’ s contribution; (iii) routing-only PFG-CMNet (no CCM) achieves 2.42 ± 0.12 mm and 0.18 ± 0.03 N at 52 ms, isolating the gain from sector-wise experts and the CNN – MLP coupling; and (iv) the full PFG-CMNet+CCM reaches 1.83 ± 0.09 mm and 0.14 ± 0.02 N at 52 ms, outperforming all internal baselines while preserving the nine-channel hardware budget. For completeness, a CNN-only baseline is also included and, while lower-latency (41 ms), it does not yield valid force – position regression under our setting. These ablations clarify the incremental benefit of CCM and the mixture-of-experts routing, directly addressing the reviewer’ s concern about component-wise contributions.

We greatly appreciate your insightful comments and suggestions again, which have substantially improved our manuscript. We hope our revised manuscript and responses can address all of your concerns and comments. Thank you.

****

Reply to reviewer 4

Reviewer #4: <u>Main issues</u>, Comment 1:

The authors have made commendable efforts in developing the UMSTD system, particularly in structural simplification, decoupling, and calibration. However, the manuscript would benefit from a more explicit discussion that contrasts these advancements with the current state-of-the-art in strain-based HCI. To highlight the structural innovation, it is recommended

to elaborate on the rationale behind the 9-channel design, clarifying why this specific configuration is advantageous compared to existing approaches in minimizing sensor number and optimizing spatial arrangement.

Response:

Thank you for the constructive and positive comments. We have made the rationale explicit and contrasted UMSTD with recent strain-based HCI systems. Section II formalizes the topology-optimized layout as a coverage maximization under a radius constraint, showing that three non-collinear sensing nodes form the minimal statically determinate configuration in 2D. Placing the three tri-axial modules at 120° provides full-plane coverage (A1 ∪ A2 ∪ A3), a rigid geometric reference, and just-sufficient observability to recover the contact location and 3D forces from the three node responses and their relative coordinates. This yields the nine-channel design that minimizes sensor count and wiring while limiting boundary overlap and cross-axis coupling.

We also clarify how this geometry interfaces with the network. Section IV-B explains that the 120° layout induces near-linear strain–force behavior within local sectors, enabling PFG-CMNet’s coarse-to-fine routing: a compact CNN localizes to one of 12 sectors and a per-sector MLP performs fine force/position regression with minimal parameters. Compared with dense strain arrays and visuotactile systems, our 9-channel design reduces channels and calibration burden while preserving accuracy and real-time rate; the quantitative contrasts are summarized in the revised comparison Table III and supported by ablations in Table I.

Comment 2:

The description of the PFG-CMNet architecture could be strengthened by providing a deeper rationale for its design. Specifically, the manuscript currently notes the functions of the CNN and MLP but would be improved by clarifying: 1) the process through which the CNN extracts spatial features from the 9-channel input, and 2) the mechanism for fusing the CNN and MLP features. Additionally, including a comparison with relevant state-of-the-art networks would help to firmly establish the performance advantages and efficiency of the proposed method.

Thank you for this valuable suggestion. We have strengthened Section IV-B and Fig. 5 to provide both the design rationale and the missing technical specifics. First, we explain how the CNN operates on the 9-channel input: the standardized 9D strain vector from three triaxial modules is processed by a compact 1D ResNet whose 3-wide kernels slide along the channel axis to learn cross-sensor couplings; stacked residual stages expand the receptive field from local pairs to triads, and global average pooling with a 12-way head outputs sector posteriors for coarse localization.

Second, we clarify the fusion: sector posteriors gate a mixture-of-experts in which the same 9D input is routed to one of twelve expert MLPs for fine regression of (X, Y, Fx, Fy, Fz); training uses temperature-sharpened top 2 soft routing to ensure boundary continuity, while inference uses argmax with a top 2 fallback at low confidence. We now report all implementation details needed for reconstruction, includes number of layers, kernel sizes, strides, padding, activations, normalization, exact MLP widths and dropout, initialization/optimizer/scaler statistics—and release weights plus an inference script for one-to-one reproducibility.

To ground the rationale, we note that the 120 ° three-node geometry yields a statically determinate 9-channel layout and near-linear mechanics within local sectors; the coarse-to-fine routing exploits this structure to reduce model complexity and latency while improving accuracy. We quantify advantages in two ways: (i) Table I provides ablations against CNN-only and MLP-only baselines and a routing-only variant, isolating the gain from CNN feature extraction and CNN→ MLP fusion; and (ii) Table III compares UMSTD with representative state-of-the-art networks/modalities (FRAs, LAs-IMU, dense strain arrays) on channel count, spatial resolution, force accuracy, update rate, decoupling method, and calibration burden, showing competitive accuracy/rate with an order-of-magnitude fewer channels.

Comment 3:

The reproducibility and overall value of the FESFI-14K dataset would be significantly improved by including critical methodological details. Specifically, the authors should describe the parameters for the random 3D forces, clarify the distribution of sampling points across the sectors, and provide justification for the chosen material parameters in the finite element method.

Response:

Thank you for highlighting the reproducibility requirements. We have revised the dataset section to include the requested methodological specifics. The disk surface is partitioned into 12 sectors of 30 ° ; sampling points are laid out on a fixed grid with 2 mm radial spacing and 1 mm circumferential spacing, giving 120 points per sector and 1,440 points in total (Fig. 5f). Ten load sets are generated: five purely vertical cases with Fz ∈ [0, 5] N, and five fully 3-D cases where Fx, Fy, Fz are independently sampled from Uniform[0, 5] N and then batched into Abaqus; Python scripts and the random seed are provided to reproduce the exact draws. All simulations use a static general step and output only the strain field to reduce computation; meshes are C3D10H (10-node hybrid tetrahedra). Post-processing extracts integration-point strains for the seven gauges per sensor and maps them to voltages through a Wheatstone-bridge model.

We also clarify the material model choices and their justification. Young’s moduli are set to: PICK 194,000 MPa; Epoxy 250 MPa; ADH 600 MPa; FPC 4,100 MPa; PI 50,000 MPa; Strain Gauge 90,000 MPa. These values are grounded in datasheet ranges and refined by calibration. The exact parameters, load scripts, and the point-distribution grid are summarized in the revised text and demonstrated in the Supplementary material, enabling one-to-one regeneration of FESFI-14K.

Comment 4: A key limitation of the force and position accuracy experiment is the lack of a comparative control group. The manuscript only presents the error data of the proposed UMSTD system. Without a quantitative comparison under identical conditions against the existing methods cited in Table I, the validation of the system's performance advantages remains incomplete.

Response:

Thank you for the constructive comments. We agree that fair, same-condition controls are important. Because competing approaches rely on different sensors and mechanics, reproducing them under identical hardware is not feasible. To address this, we introduce strict in-system controls that use the exact same signals, ground truth, and protocol: CNN-only, MLP-only, MLP+CCM, routing-only PFG-CMNet, and the full PFG-CMNet+CCM (Table I). These controls establish a same-condition baseline and show that the proposed architecture reduces error from 3.13±0.16 mm and 0.38±0.053 N (MLP) to 2.42±0.12 mm and 0.18±0.03 N with routing, and to 1.83±0.09 mm and 0.14 ± 0.02 N with CCM at the same 52 ms latency. To contextualize beyond our hardware, Table III compiles quantitative metrics reported by recent strain-based and alternative HCI systems, including channel count, spatial resolution, force error, update rate, decoupling method, and calibration burden. Taken together, these same-signal controls and cross-study metrics show that UMSTD with PFG-CMNet attains competitive accuracy and real-time rate while using only nine channels and low calibration effort, thereby substantiating the performance advantages despite modality differences across prior work.

Comment 5: The theoretical foundation of this work would be significantly strengthened by providing a more comprehensive exposition of the mathematical derivations. Specifically, the correlation between rotation angles and sensor position coordinates, the solution strategy for determining matrix dimensions and compensation coefficients, and the statistical analysis methods.

Response:

Thank you for the thoughtful critique. We clarify that our calibration is not purely empirical. The SCIC component is a geometric correction derived from principles using Rodrigues’ rotation to remove installation-induced orientation misalignment at the triad level; its derivation and implementation are given in Section IV-C with the principle diagram in Fig. 5. The STC matrix is obtained from the sensor transfer characteristics under small-strain linear elasticity and Wheatstone readout theory, yielding a closed-form linear map that is then identified from calibration loads. The overall CCM is the block-diagonal SCIC composed with the theory-based STC, providing a physically grounded transformation rather than an ad-hoc fit. For transparency, the specific material and load parameters used for STC identification are summarized in the new table and demonstrated in the Supplementary Video, and Section IV-C reports all intermediate steps needed to reproduce the matrices. In future work, we will systematically examine how sensor signal amplitude and inter-element spacing jointly affect prediction accuracy.

Comment 6: The study would be greatly enhanced by a more thorough discussion that delves into the interpretation of the experimental results. Expanding the discussion to include the potential sources of error and the system's practical boundaries would provide a more balanced perspective. This would support the authors' conclusion regarding the practical utility of the UMSTD for the target patient population.

Response:

We thank the reviewer for the constructive comments. We have added a dedicated discussion of limitations and error sources, primarily in the Conclusion (page 9). We now explicitly mention:

(1) The force range limit: we note that beyond 5 N, accuracy degrades (e.g., error ~0.7 N at 10 N) and that the model would need retraining or a larger dataset to handle that (page 9).

(2) Assumption of single contact: we state that our system currently assumes one contact at a time and would not handle multi-touch without further development.

(3) We reiterate the observed cross-axis error sources and how we compensated them, as well as remaining small errors.

Furthermore, in Section V-A (page 7) we explain why errors increased in 5 – 10 N (lack of training data in that range). And in Section IV-C (page 6) we describe the two main error sources (STC and SCIC) which is part of interpreting where errors come from. Combined, these additions give a transparent view of our system’ s reliability and the boundaries of its performance. We also indicate that future work will tackle some of these limits (cross-device calibration, etc., page 9).

<u>Minor Issues</u>, Comment 1: The organization of some sections could be optimized for clarity. The description of the relationship between strain resistance change and voltage output would fit better in Section II. Similarly, Section IV.A, which pertains to experimental data preparation, should be moved to the beginning of Section V to provide a more logical progression.

Response:

We thank the reviewer for the constructive comments. We have made corrections in the revised manuscript.

Comment 2: The clarity of the manuscript would benefit from a thorough proofreading to address instances of ambiguous phrasing and undefined terms. As representative examples, the antecedent of "enabling..." in the Abstract should be specified, and all variables in Equation (3) need to be clearly defined upon first use.

Response:

We thank the reviewer for the comments. We thoroughly proofread the entire manuscript. All variables are now defined when introduced. For example, around Eq. (3) (Section III-A, page 3-4), we explicitly define $d=[\delta_x;\delta_y;\delta_z;\theta_x;\theta_y;\theta_z]^T$ as translations and rotations of the plate, and S_{i_x}, S_{i_y} as the sensor coordinates. We have checked each equation:

Eq. (1) – (2): added that R is nominal resistance, $\Delta R_{i_1}, \Delta R_{i_2}$ are the changes due to strain, etc.

Eq. (4): defined d and the size of A_i.

Eqs. (7) – (9): clarified they come from force and moment balance, and defined S_{force} for Eq. (9).

Eq. (17) – (19) (formerly 25 – 26 in original): defined M, Z, C, T clearly.

We also corrected minor notations. All acronyms have been expanded as noted, and any ambiguous references have been resolved. For instance, we replaced phrases like “the sensor” with specific references when needed. In summary, the revised manuscript has been carefully edited for clarity and completeness in terminology.

****

To all reviewers：

We sincerely appreciate all reviewers for their time and effort in moving this manuscript to the public. These comments are insightful and constructive, which are very helpful for improving the quality of this manuscript. We believe that these valuable suggestions will further motivate us to do better in the future research. Moreover, the rigorous and scientific attitude expressed in these comments deeply influences our revision and benefits us a lot. We hope that the corrections can meet with your approval. Once again, thank you for your kind comments.