EMORY UNIVERSITY SCHOOL OF MEDICINE
Department of Biomedical Informatics

An NLP Pipeline for Structuring Tremor Severity Ratings from Semi-Structured Clinical Notes

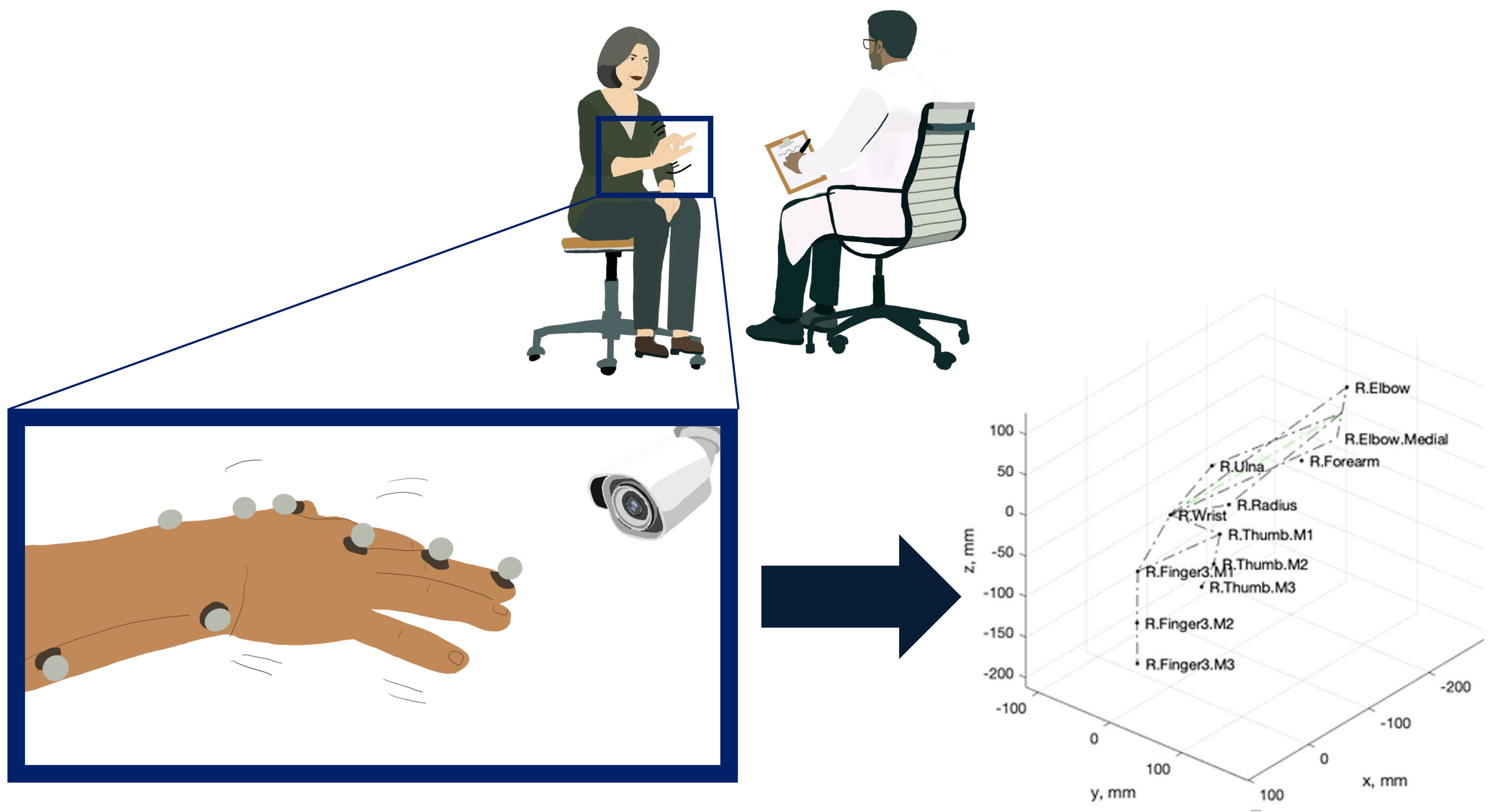
Jeanne M. Powell, PhD[1], Christine Esper, MD[2], Richa Tripathi, MD[2], Douglas Bernhard, MS, Abeed Sarker[1], J. Lucas Mckay, PhD, MSCR[1]
[1] Department of Biomedical Informatics, Emory School of Medicine, Atlanta, GA
[2] Department of Neurology, Emory University School of Medicine, Atlanta, GA

Abstract

Background: Tremor assessment in PD and ET depends on clinician notes from standardized tasks, limiting scalable ML development.
Objective: Evaluate LLaMA3-70B for automated extraction of binary tremor labels across seven body parts.
Methods: Structured clinician tables (2,980 labels) were paired with free-text notes. Multiple prompts and few-shot settings (k = 1–3) were tested.
Results: Best prompt: macro F1 = 0.98 (limbs > face/trunk). Few-shot (k = 3) outperformed zero-shot (0.89 vs 0.56).
Conclusion: LLaMA3 accurately converts clinician text into binary tremor labels suitable for ML training.

Study Context: Motion Analysis Lab

Clinical observation of tremor occurs in tandem with 3D kinematic recordings.



📝 **Semi-Structured Notes (2022–present, n > 600 patient encounters)**
Format: *Task: Observation*
Example: *Sit-point-right: -l hand, +r mild hand tremor*

📊 **Structured Grid Added (2024–2025, n > 150 patient encounters)**
Table: task × body part → tremor (0/1)

Goal: Automate conversion of clinician text into structured tremor labels to generate training data for ML models that predict tremor from sensor signals.

Study Workflow

Clinical Visit
Individuals with PD, ET, and other movement disorders perform standardized motor tasks (*sit rest*, *point right*, *spiral draw*).

Clinical Observation
Tremor noted in real time, either in free text or structured table.

Build Supervised Dataset
Extract task–observation pairs from clinical notes and align them with structured tables.

LLM Inference
Prompt LLaMA3 to extract binary tremor labels (0/1) for 7 body regions.

Assess Model Performance
Compare model predictions to ground truth dataset.

Structured Dataset
Structured outputs ready to train predictive models linking motion to tremor.

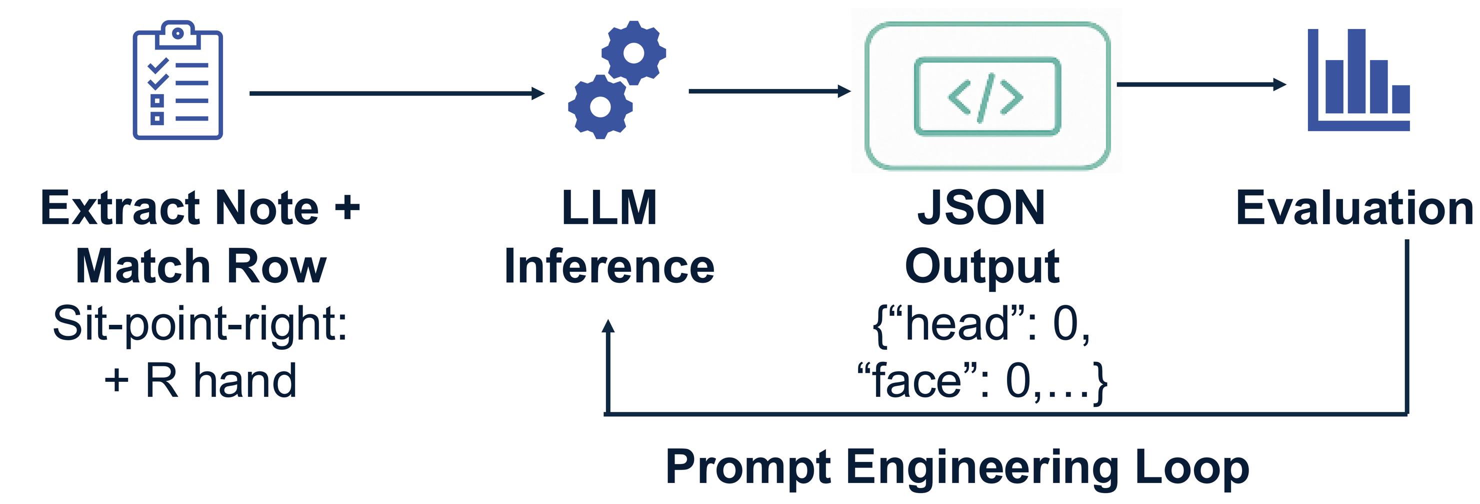
Rule-Based Data Extraction

Goal: Link clinician notes to structured tremor tables.

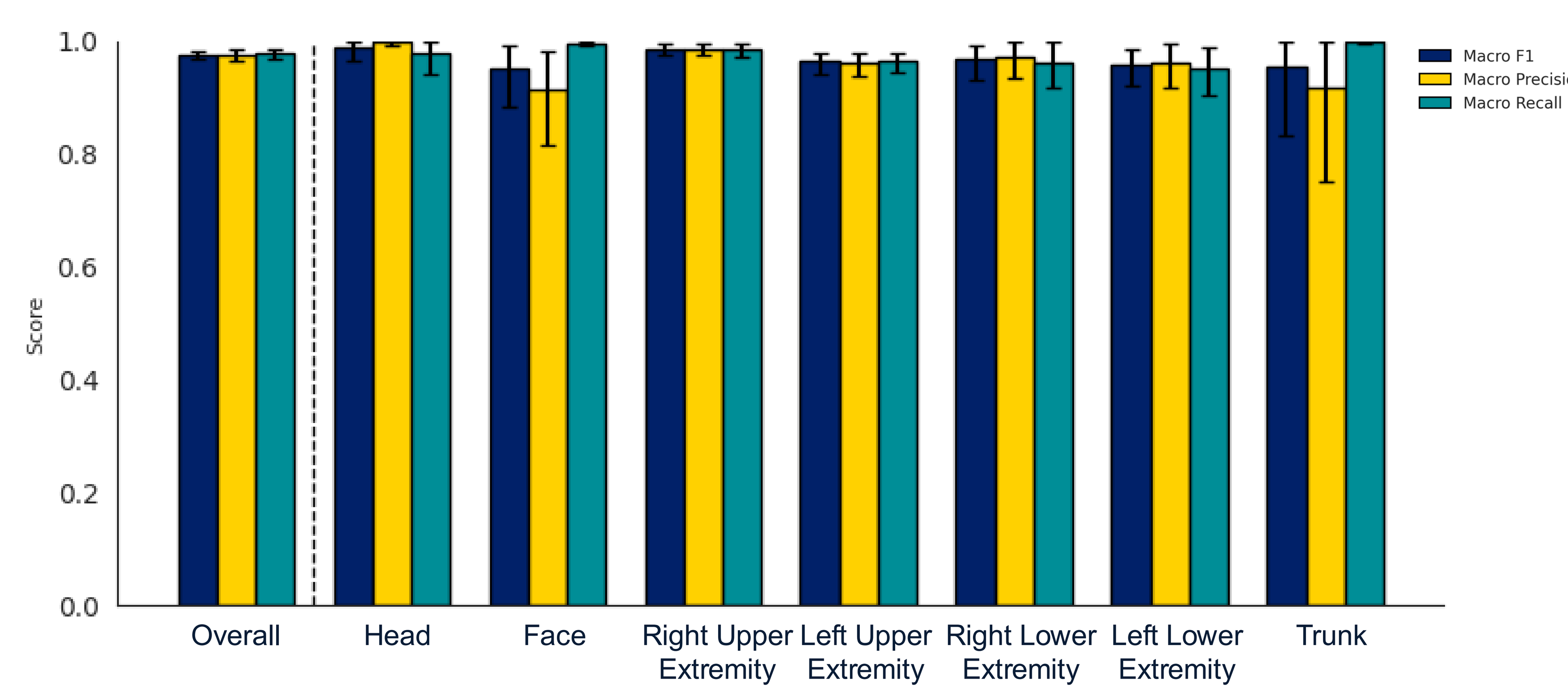
Input:
DOCX notes (*task : observation*)
Regex Parsing:
Normalize task names, Match to TRC motion files, Handle "same as above" references
Output:
Unified dataset aligning free-text and binary (0/1) tremor labels

LLM Classification

Goal: Convert extracted text into binary tremor labels (0/1) across seven body parts from 2,980 labeled instances (42 encounters, 460 tasks).



Classification performance by body region



Retrieval Augmented Prompting

Used 10 new examples, tested k = 1, 2, 3
Injected similar examples based on cosine similarity directly into prompt (few-shot)

k	Overall F1
1	0.559 (95% CI: 0.524 - 0.595)
2	0.579 (95% CI: 0.541 - 0.616)
3	**0.886 (95% CI: 0.855 - 0.915)**

Conclusion

Accurate: Aligns with expert labels (F1 = 0.98)
Few-shot > Zero-shot: Performance improves with retrieval examples.
Weak points: Tabular data overrepresent pertinent negatives compared to text
Integration: Structured outputs link with 3D kinematic features
Future work: Extend to unstructured notes; use active learning for uncertain cases

Acknowledgements

We extend our heartfelt gratitude to all the individuals who generously donated their data for this research project. We also thank members of the Emory Motion Analysis Lab, Sarker Lab, Emory Brain Health Center

EMORY UNIVERSITY

Contact Information:
Email: jmpowe6@emory.edu
LinkedIn: linkedin.com/in/jeanne-powell/