Syed Taswar M.

Team Romberger — CS 410 Long Project

MODEL ACCURACY
73.9%

ML Pipeline & Model Training

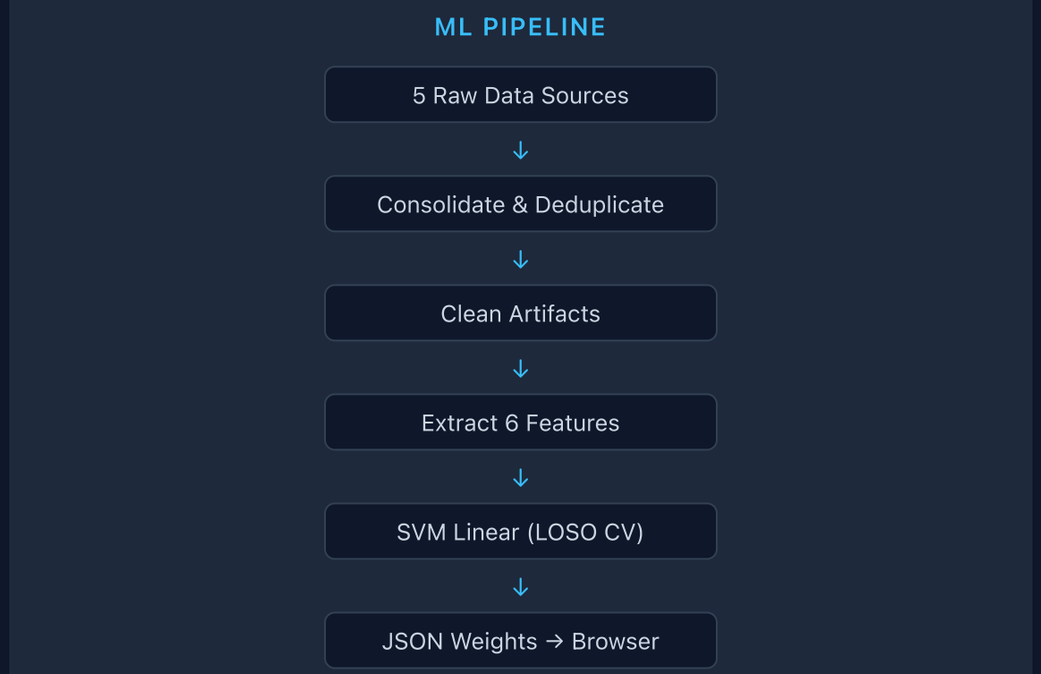
Built the full data pipeline: consolidation of 5 data sources, artifact cleaning, 30-second chunking, and feature extraction (6 summary statistics). Trained SVM Linear classifier with GroupKFold LOSO cross-validation across 9 subjects (142 samples). Identified and removed corrupted data from subject_08.

Backend Integration (Supabase)

Integrated Supabase database to store every prediction with sample_id, session_id, label, extracted features, and prediction output. Added prediction history UI, toast notifications, and Row Level Security policies. Dataset grows with each upload.

Frontend & UX

Built the main website with dark mode, demo data generator, stick figure animation, responsive design, and accessibility (ARIA labels, keyboard nav, reduced-motion). Added CSV drag-and-drop upload with real-time validation.



GitHub: A2AppRom/Manik_Data_For_Romberg

Live Site: a2approm.github.io/Manik_Data_For_Romberg

Artifacts: /courses/cs410/s26/hdeblois/GROUP2/longproj01/t2-WR/taswar01-artifacts/

T-Shaped Skills

Deep expertise in ML, broad contributions across web design and database

DEEP — PRIMARY

Machine Learning

- Designed and built the full 7-stage data pipeline (Python, pandas, scikit-learn)
- Trained and compared 3 classifiers: SVM RBF, SVM Linear, Logistic Regression
- Implemented GroupKFold leave-one-subject-out cross-validation (9 folds)
- Engineered 6 features: mean, median, std, skewness, kurtosis, path length
- Discovered and removed corrupted data (subject_08 — identical open/closed files)
- Exported model weights to JSON for client-side inference in the browser
- Wrote automated test scripts for pipeline, model, and website validation

BROAD — SECONDARY

Web Design

- Built the main index.html (2,100+ lines) with responsive layout
- Dark/light mode toggle with localStorage persistence
- Stick figure animation, canvas particle effects, scroll-reveal
- Accessibility: ARIA labels, keyboard navigation, reduced-motion support

BROAD — SECONDARY

Database

- Integrated Supabase (PostgreSQL) for persistent storage of predictions
- Configured Row Level Security policies (INSERT + SELECT)
- Built prediction history UI with auto-refresh and manual reload
- Stored raw CSVs in Supabase Storage bucket for dataset growth

Initial Comments: /home/taswar01/cs410/4454.txt

Additional Comments & Plans

What I Learned

- End-to-end ML workflow: raw sensor data to browser-based inference
- Data quality matters more than model complexity — removing 20 corrupted samples improved accuracy significantly
- LOSO cross-validation is essential for small datasets to avoid overfitting
- Serverless backends (Supabase) can replace a full server for CRUD-style apps

Artifacts

- Task_Breakdown.txt — overview of all 13 tasks and Thea's 7 requirements
- 15 task reports documenting each pipeline stage
- model_comparison.csv — 3-model benchmark results
- cv_fold_results.csv — per-fold LOSO accuracy breakdown
- features_dataset.csv — 142-sample feature matrix
- romberg_model_weights.json — trained SVM weights
- 8 website screenshots (hero, predict, performance, etc.)

Next Steps

- Coordinate acceptance testing with Thea (client)
- Collect more data per subject to reduce class imbalance
- Add individual baseline calibration for personalized predictions
- Server-side CSV validation via Supabase Edge Functions

Contributions Timeline

Mar 25 — Initial website + ML dataset pipeline

Apr 1 — Dark mode, demo data, PDF export, accessibility

Apr 1 — ML classifier integration with frontend

Apr 22 — Full ML pipeline: 9 subjects, SVM model, tests

Apr 26 — Backend: prediction history, toasts, NaN filtering

Research Paper: /home/taswar01/cs410/rpaper.pdf

GitHub: A2AppRom/Manik_Data_For_Romberg

Live Site: a2approm.github.io/Manik_Data_For_Romberg

Artifacts: /courses/cs410/s26/hdeblois/GROUP2/longproj01/t2-WR/taswar01-artifacts/