

Finnish Electronic Frailty Index (eFI): 2024 Progress (Mar 18 - Dec 31)

TODO

2024-12-31

Table of contents

1	Executive Summary	1
2	Methods mapping to repository	2
3	Milestones (2024)	2
4	Dataset Overview	2
5	NLP Pipeline (condensed)	2
6	Experiments Log (excerpt)	3
7	Reproducibility	3
8	Limitations	4
9	References	4

1 Executive Summary

We advanced an AI-enhanced eFI by extracting frailty-related deficits from Finnish EHR free text (NER) and integrating them with structured data (ICD-10, labs). NER identified more cases of falls and incontinence than ICD-10 alone (F1 0.81-0.87) and showed stronger mortality associations for these deficits. Mobility limitations were predictive. Loneliness was not significant.

Note: Some filenames include 2025 in their names. Confirm creation dates before labeling outputs as 2024. [TODO]

2 Methods mapping to repository

Method item	Repository artifact
Overlap/profiling	notebooks/W6.Vennd_Overlap.ipynb
HFRS pipeline	notebooks/W7.Combine_data.ipynb, notebooks/HFRS.2.Calculate_HFRS_Points.ipynb, notebooks/HFRS.3..., notebooks/HFRS.4...
CCI-HFRS correlations	notebooks/W19..., notebooks/W20..., notebooks/W21...
BMI extraction	notebooks/W22...-W27... (covariate prep)

[TODO] Verify notebook paths and commit hashes.

3 Milestones (2024)

Date	Milestone	Artifact / Script	Status
2024-04-01 - 2024-06-30	Mobility ADL scoping	[TODO]	Done
2024-05-29	“Summary Toothdata” talk	slides [TODO]	Done
2024-06-19	Labeling presentation	slides [TODO]	Done
2024-09-01 - 2024-10-31	Bathing & dressing scoping	[TODO]	Done
2024-11-18, 2024-11-20	LBW/OW seminars	slides [TODO]	Done

4 Dataset Overview

Cohort / Source	Years	Patients	Notes	Key fields	Access / Ethics
HYVÄKS (Central Finland)	2010-2022	166,147	~10.6M free-text notes	ICD-10, free text, labs	[TODO: approval IDs]

5 NLP Pipeline (condensed)

Stage	Input -> Output	Checks
Labeling / QA	Free text -> Labeled spans	Double label + adjudication
Modeling	Labeled -> NER model	[TODO: model details]
Eval / Linkage	Test -> P/R/F1, HRs	Target F1 > 0.80; HRs outperform ICD-10 baselines for falls and incontinence

6 Experiments Log (excerpt)

ID	Objective	Model	Metrics	Outcome
E-01	Falls NER vs ICD-10	[TODO]	F1 0.87, [TODO CI]	NER detects earlier onsets; HR 1.31
E-02	Incontinence NER vs ICD-10	[TODO]	F1 0.81, [TODO CI]	NER>ICD-10; HR 1.99
E-03	Mobility NER	[TODO]	F1 0.85, [TODO CI]	Positive mortality association

7 Reproducibility

- **Environment**
 - env/requirements.txt [TODO]
 - env/environment.yml (conda) [TODO]
- **Minimal run**
 1. Create conda env or pip venv.
 2. Run notebooks in notebooks/ in order: W6 -> W7 -> HFRS.* -> W19-W21 -> W22-W27. [TODO order validation]
 3. Export derived tables to tables/ and figures to figures/.
- **Synthetic data note** Provide synthetic or anonymized samples for unit tests and CI. [TODO]

8 Limitations

Under-coding in ICD-10, Finnish phrasing variability, missing confidence intervals and calibration checks, potential domain shift.

9 References

[TODO] Add formal citations for eFI, HFRS, CCI, and Finnish EHR NLP sources.