Poster #RWD144

Social Determinants of Health, Mortality, and Extended Hospital Stay in Atrial Fibrillation: Leveraging Open-Source Large Language Models

Vishnu B Suresh[1], Won Chan Lee[1]

[1]Axtria Inc., 300 Connell Drive, Berkeley Heights, NJ, US

AXTRIA
INGENIOUS INSIGHTS

INTRODUCTION

- Social determinants of health (SDoH), including employment, social support, and relationship factors, are critical drivers of patient outcomes, yet structured EHR fields and claims data capture them only partially; The richest SDoH context often resides in unstructured clinical notes.[1]
- The inability to systematically extract and quantify SDoH at scale represents a significant gap in clinical research, quality improvement, and care coordination for high-risk populations.
- Large language models (LLMs) offer a transformative opportunity to mine SDoH from clinical narratives, but most prior studies rely on large proprietary models that are costly, inaccessible, and raise data-privacy concerns.
- Smaller open-source LLMs (<7B parameters) can be self-hosted on commodity hardware, eliminating per-token API fees and keeping clinical text fully within a secured local environment, offering a practical, cost-effective, and privacy-preserving alternative, but evidence of their real-world utility in clinical prediction remains limited.
- Atrial fibrillation (AF) is one of the most common cardiac arrhythmias, associated with significant morbidity, mortality, and healthcare resource utilization. Social risk factors may compound clinical risk yet are rarely integrated into AF outcome models.

OBJECTIVE

- Evaluate the feasibility of small open-source LLMs to extract SDoH from clinical discharge summaries in a real-world hospital database (MIMIC-IV).
- Assess whether LLM-extracted SDoH features provide incremental predictive value (beyond clinical variables) for 30-day mortality and extended LOS (>7 days) in AF patients.
- Rank SDoH attributes by feature importance to identify the strongest social predictors of adverse AF outcomes.

METHODS: DATASET AND LLM EVALUATION

- The data source used was MIMIC-IV (Medical Information Mart for Intensive Care), a large, freely-available EHR database spanning 2008-2022.[2,3,4]
- The study population consisted of 1,184 AF admissions with 1,000 discharge summaries.
- The LLM used for extraction of SDoH features was selected after comparison of multiple open-source models with <7B parameters.
- 3 different strategies were also evaluated for each model: zero-shot (no prompting), one-shot (prompted with one example) and few-shot (prompted with multiple examples), and the best strategy for each model was comparatively evaluated.

METHODS: SDoH EXTRACTION AND OUTCOME MODELING

- **13 SDoH attributes were extracted per discharge summary including**: employment status (retired, employed, unemployed, disabled), social support level (limited, adequate), relationship/marital status (married, single, divorced, widowed), living situation and housing stability.
- Notes were de-identified and section-parsed prior to extraction. Of 1,184 sampled AF admissions, 34.5% (n=409) had ≥1 SDoH attribute detected; remaining cases were treated as "not documented" rather than imputed.
- A high-confidence validation threshold was applied to minimize false-positive extractions and outputs additionally underwent manual review for quality assurance.
- **Outcome variables**: (1) 30-day all-cause; (2) Extended Length Of Stay, defined as LOS >7 days.
- **3 supervised machine learning (ML) algorithms evaluated**: Lasso logistic regression, Random Forest, and XGBoost; with binary outcomes (yes/no) for both outcome variables ;
- **Two modeling scenarios compared head-to-head for each algorithm**: [1] Scenario A (Baseline): Clinical features only; [2] Scenario B (Augmented): Clinical features + LLM-extracted SDoH features
- **Primary performance metric for evaluation**: Area Under the Receiver Operating Characteristic Curve (AUC-ROC); secondary metrics include sensitivity and specificity at clinical decision thresholds.
- Feature importance rankings generated to identify the most influential SDoH predictors within each model.

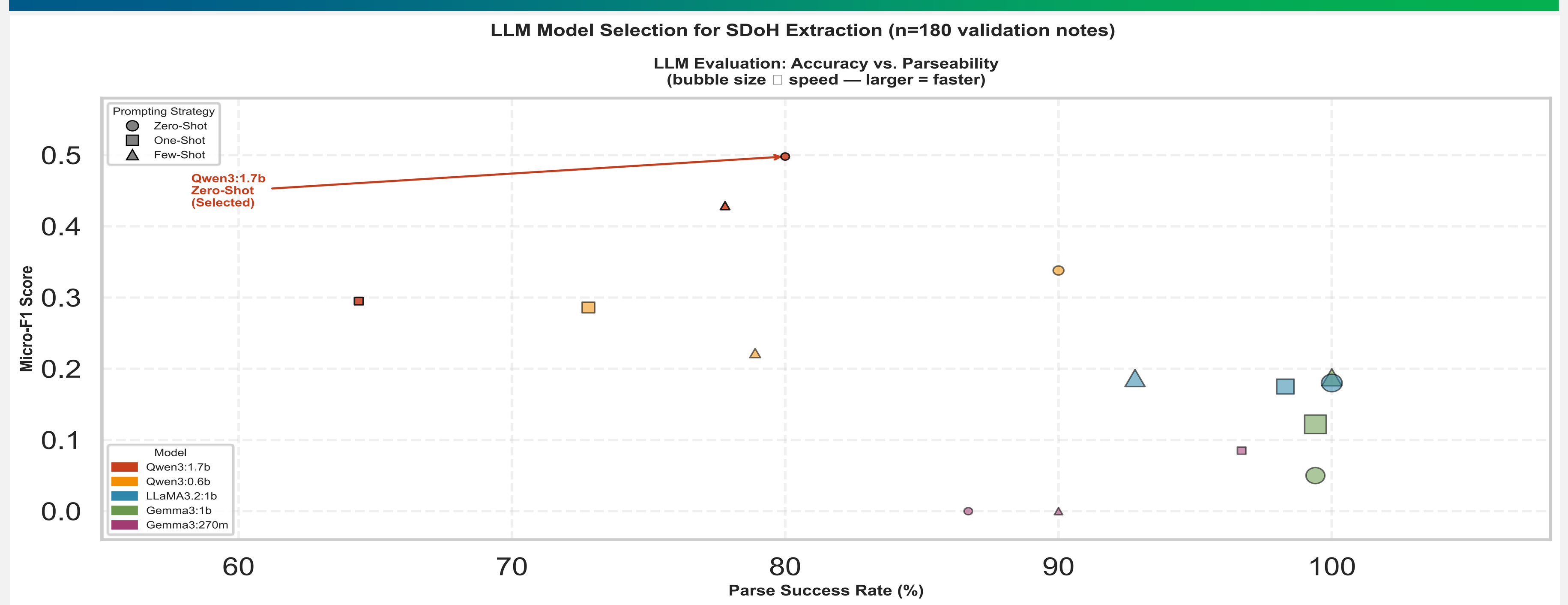
RESULTS: LLM EXTRACTION PERFORMANCE

- SDoH detection rate was 32.9% , reflecting documentation gaps rather than extraction failure. Even sparsely captured, the signal produced measurable LOS AUC gains.
- The best-performing model was Qwen3:1.7B, selected based on extraction quality, consistency, and output stability.

TABLE 1: COMPARISION OF PERFORMANCE ACROSS LLMs

LLM	Best Strategy	Micro-F1	Macro-F1	Parse Rate
Qwen3:1.7b	Zero-shot	0.498	0.445	80%
Qwen3:0.6b	Zero-shot	0.338	0.258	90%
Gemma3:1b	Few-shot	0.188	0.091	100%
LLaMA3:2.1b	Few-shot	0.187	0.112	93%
Gemma3:270m	One-shot	0.085	0.071	97%

FIGURE 1: LLM EVALUATION - ACCURACY VS PARSEABILITY



RESULTS: MODEL PERFORMANCE FOR OUTCOMES

Extended Length of Stay (LOS > 7 Days):

- Baseline prevalence: 37.1% (n=440) of AF admissions exceeded 7 days.
- Incorporating LLM-extracted SDoH features produced consistent AUC improvements across models:
- Key SDoH predictors for extended LOS: (1) Retired employment status emerged as the strongest social predictor of prolonged hospitalization; (2) Relationship/marital status was the secondary social predictor

30-Day Mortality:

- Baseline prevalence: 9.1% (n=108) of AF admissions resulted in 30-day mortality
- Best clinical-only model (XGBoost): AUC = 0.830, indicating strong discrimination using clinical features alone.
- Best SDoH-augmented model: AUC = 0.794; clinical features outperformed the SDoH-augmented models for this endpoint.
- Notable SDoH associations with mortality: limited social support showed a moderate association; however, effect sizes were smaller than those observed for LOS prediction, likely due to the low event rate and limited documentation of social factors

FIGURE 2: SDOH FEATURE IMPORTANCE

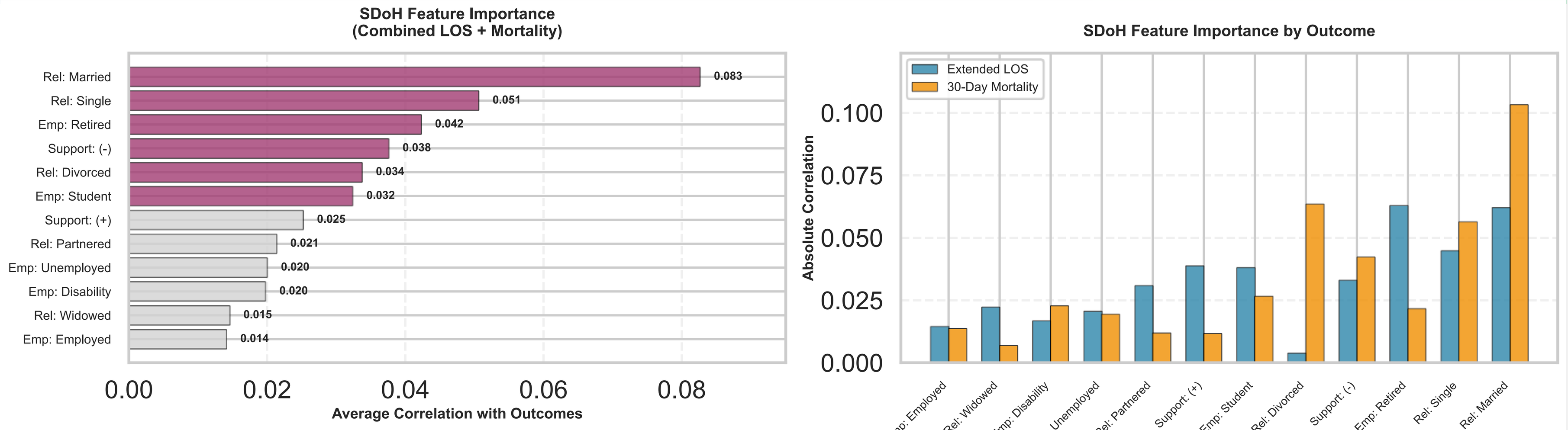
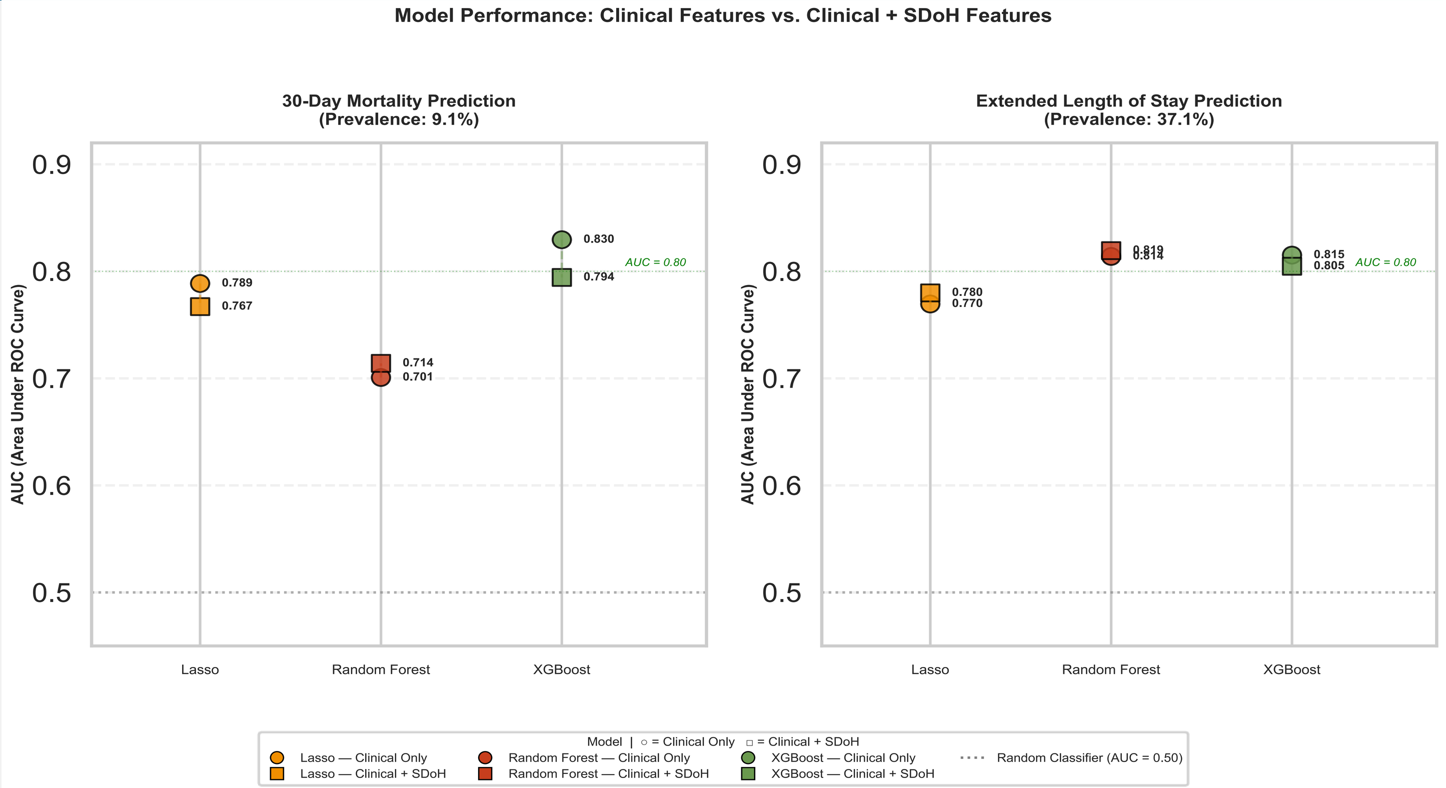


TABLE 2: COMPARATIVE MODEL PERFORMANCE

Model	Outcome	Clinical-only AUC	Clinical + SDoH AUC	% Change
Lasso Regression	30-Day Mortality	0.789	0.767	-2.8%
Lasso Regression	Extended LOS (>7d)	0.770	0.780	**+1.3%**
Random Forest	30-Day Mortality	0.701	0.714	**+1.9%**
Random Forest	Extended LOS (>7d)	0.814	0.819	**+0.6%**
XGBoost	30-Day Mortality	0.830	0.794	–4.3%
XGBoost	Extended LOS (>7d)	0.815	0.805	-1.2%

FIGURE 3: COMPARATIVE MODEL PERFORMANCE



KEY FINDINGS

- Small open-source LLMs (Qwen3:1.7B) can reliably extract clinically meaningful SDoH from unstructured discharge summaries with no proprietary AI infrastructure required
- LLM-extracted SDoH features add modest but consistent incremental value for extended LOS prediction in AF patients (up to +1.9% AUC improvement), suggesting they may complement clinical features for LOS risk stratification, pending external validation in multi-center cohorts.
- Retired employment status and relationship/marital status are the strongest social predictors of prolonged hospitalization, underscoring the importance of discharge planning tailored to social circumstances.

CLINICAL IMPLICATIONS

Clinical Implications:

- Findings highlight actionable opportunities for healthcare systems to leverage open-source AI tools to systematically capture SDoH without requiring expensive proprietary technology.
- Patients identified as socially vulnerable - particularly those who are retired, lack social support, or live alone - may benefit from targeted discharge planning, early social work consultation, and enhanced care coordination.
- Integration of SDoH screening into clinical workflows for AF management could improve risk stratification and reduce preventable prolonged admissions.
- The LLM-based extraction approach described here provides a scalable, cost-effective pathway for real-world implementation across healthcare systems.

REFERENCES

1. Lybarger K, Dobbins NJ, Long R, et al. Leveraging natural language processing to augment structured social determinants of health data in the electronic health record. J Am Med Inform Assoc. 2023;30(8):1389-1397. doi:10.1093/jamia/ocad073
2. Johnson, A., Bulgarelli, L., Pollard, T., Gow, B., Moody, B., Horng, S., Celi, L. A., & Mark, R. (2024). MIMIC-IV (version 3.1). PhysioNet. RRID:SCR_007345. https://doi.org/10.13026/kpb9-mt58
3. Johnson, A.E.W., Bulgarelli, L., Shen, L. et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 10, 1 (2023). https://doi.org/10.1038/s41597-022-01899-x
4. Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220. RRID:SCR_007345.

DISCLOSURES

VBS and WCL are employees of Axtria Inc., USA. All authors declare no conflicts of interest related to this research.

POSTER PRESENTED AT

ISPOR 2026, Philadelphia, PA, USA