THE UNIVERSITY OF ALABAMA AT BIRMINGHAM

Leveraging Large Language Models to Enhance Machine Learning Interpretability and Predictive Performance: A Case Study on Emergency Department Returns for Mental Health Patients

Abdulaziz Ahmed, PhD[1], Mohammad Saleem[1], Mohammed Alzeen[1], Badari Birur[1] , Rachel E. Fargason[1] , Bradley G. Burk[1] , Hannah R. Harkins[1] , Ahmed Alhassan[1] , Mohammed Ali Al-Garadi[2] [1]University of Alabama at Birmingham, Birmingham, AL, USA; [2]Vanderbilt University Medical Center, Nashville, TN, USA

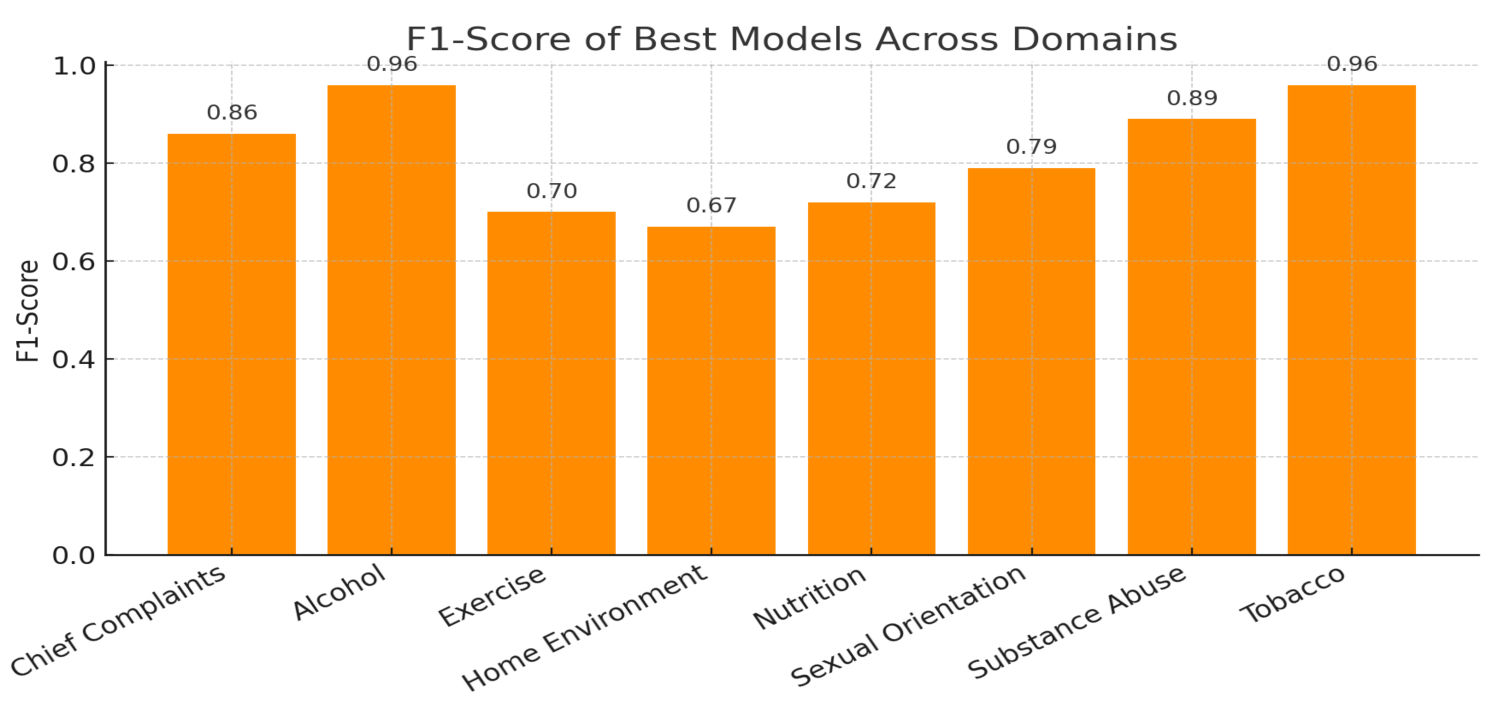
INTRODUCTION

- **Critical Problem**: Nearly one in four (25.2%) psychiatric patients return to the Emergency Department (ED) within 30 days after discharge.
- **Study Goal**: Evaluate the integration of Large Language Models (LLMs) with traditional Machine Learning (ML) to improve the prediction of 30-day ED returns for mental health patients.
- **Key Focus**: Leverage both structured and unstructured data (e.g., chief complaints) to create actionable clinical insights and enhance model interpretability

RESULTS

LLM Feature Extraction Performance

LLaMA 3 (10-shot) significantly outperformed conventional methods in classifying chief complaints.

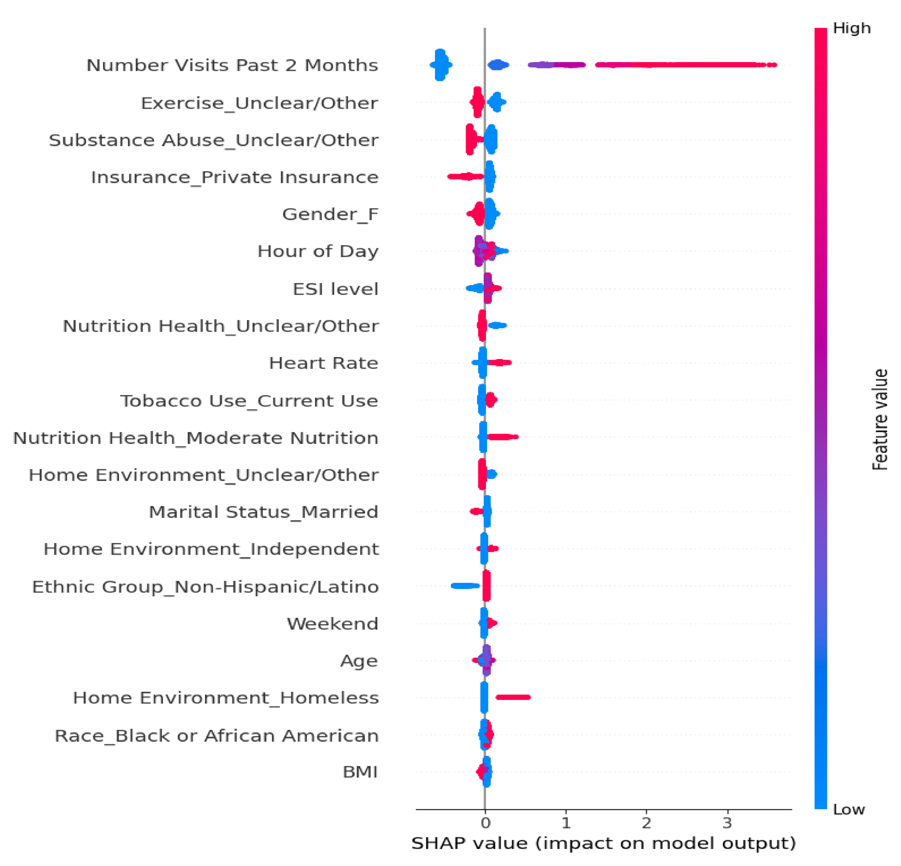


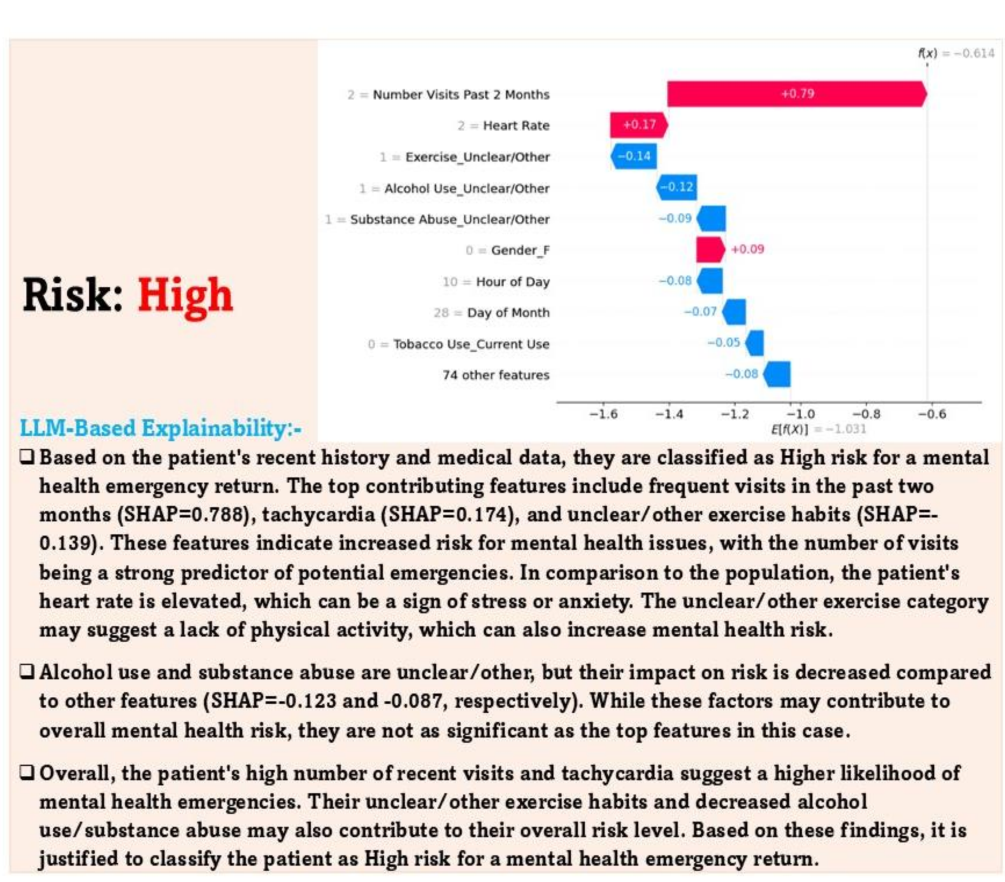
Predictive Performance Gains

Incorporating LLM-extracted features consistently improved performance across all models.

Metric	XGBoost (Without LLM)	XGBoost (With LLM)
AUC	0.74	0.76
AUC-PR	0.58	0.61

Interpretability and Feature Importance

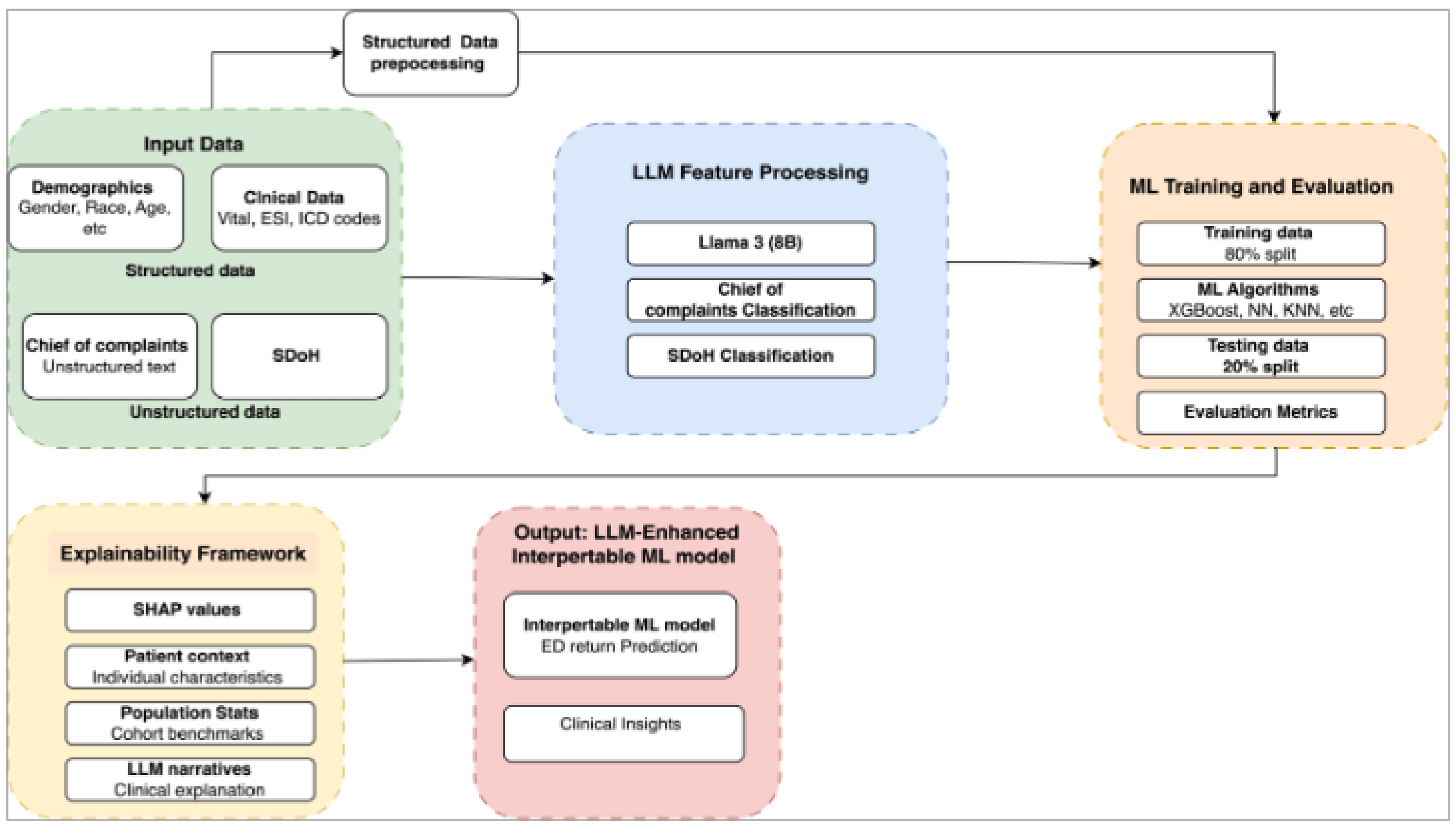




METHODS

We developed a framework that integrates structured and unstructured ED data for interpretable predictive modeling, consisting of two main steps:

- LLM Feature Processing: Unstructured text (chief complaints and SDOH) was processed using Llama-3 (8B) to classify and convert it into structured features, combined with demographics, vitals, and clinical codes.
- Model Training and Explainability: Combined features were used to train ML models (XGBoost, Random Forest, SVM, BlueBERT), with SHAP-based and LLM-generated explanations providing transparent, patient-level insights for ED outcome prediction.



Discussion & Conclusion

- LLM integration is a viable and effective strategy to leverage unstructured data for improved prediction and interpretability in healthcare.
- While predictive gains were modest, the LLM-extracted features captured critical risk factors (substance use, housing instability) that were otherwise absent from structured data.
- The high accuracy of LLM-generated explanations enables the creation of actionable, personalized risk narratives for clinicians, which is crucial for preventative intervention and patient safety.

Future Directions

- Validate the framework on external datasets.
- Explore advanced LLM prompt engineering techniques for further performance gains.

Abdulaziz Ahmed, aahmed@uab.edu