Comparing Traditional and Graph-Enhanced Pipelines for Sepsis Prediction Using Omics Data

Joelle ASSY, Silvia TROTTET, Rayane ADAM, Yazid HOBLOS
Master 2 GENIOMHE-AI, Univ. Evry Paris-Saclay

Introduction

Sepsis remains a leading cause of global mortality

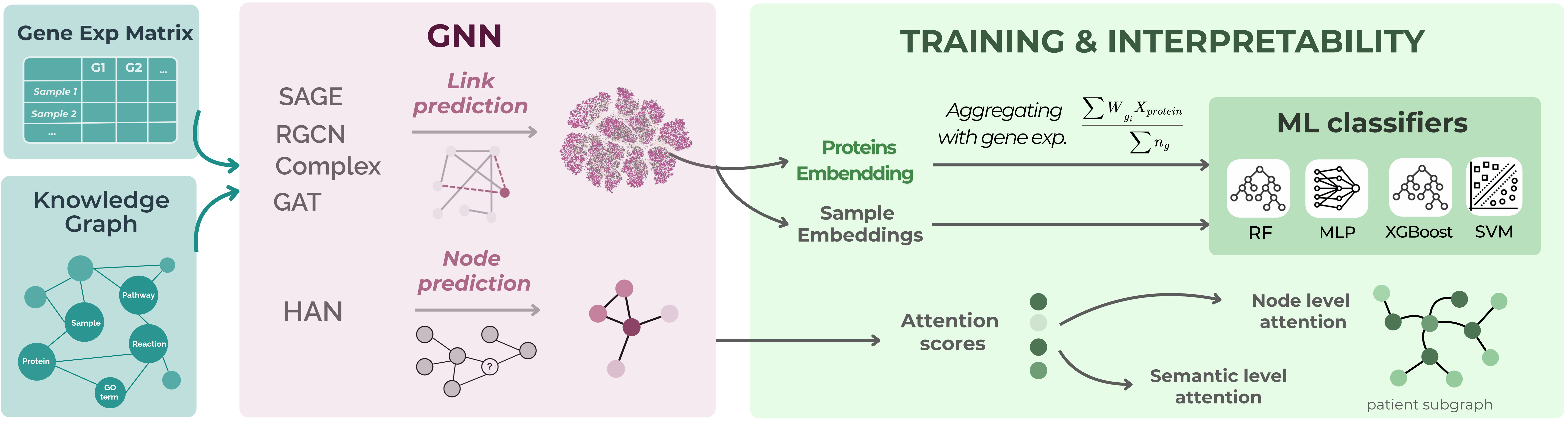
Sepsis is a life-threatening organ dysfunction resulting from a dysregulated host response to infection. Accurate early prediction is vital, yet traditional methods treat genes as independent features. This project investigates if outcomes are better predicted by encoding relational information among proteins, reactions, pathways, and GO terms. We compare a traditional transcriptomic pipeline against a graph-augmented approach that captures structural biological context.

Datasets: (1) Gene Expression (2) Knowledge Graph

(1) GSE54514 dataset with 163 whole blood samples from 26 survivors, 9 non-survivors, 18 controls collected daily for 5 days.

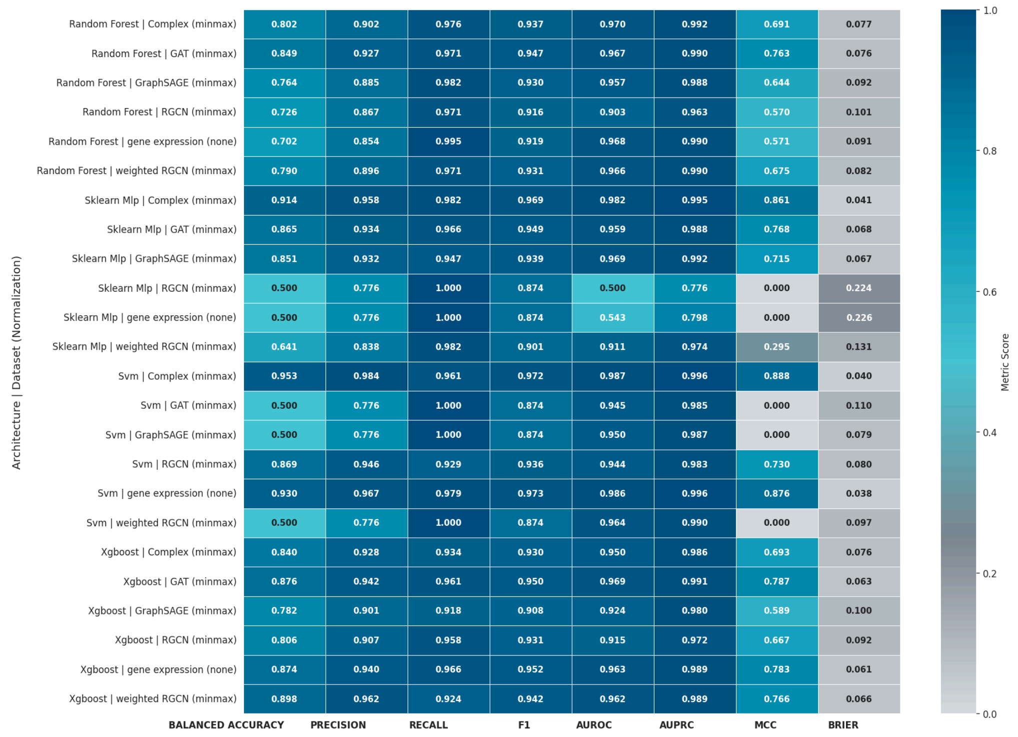
(2) Multi-relational network mapping Patient Samples, Proteins, GO Terms, Reactions, and Pathways, by integrating curated data from 8 sources: GO, Pathway Ontology, PPI, Entrez, BioGRID, UniProt, STRING, and Reactome.

Overall Architecture



Results

ComplEx embeddings enhance sepsis prediction



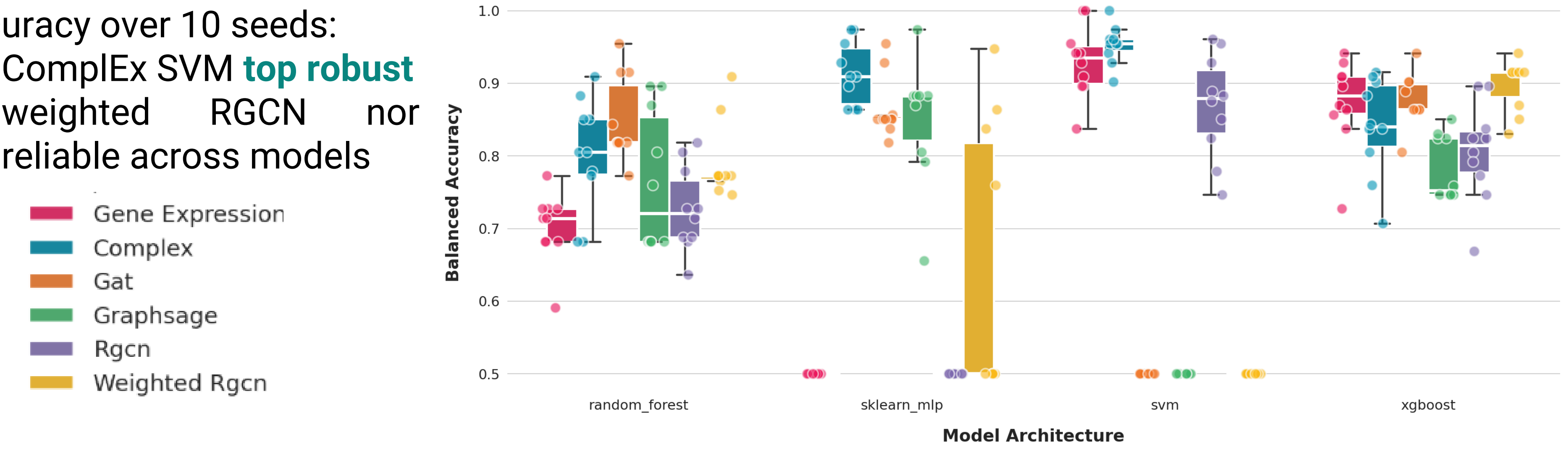
Architecture | Dataset (Normalization)	BALANCED ACCURACY	PRECISION	RECALL	F1	AUROC	AUPRC	MCC	BRIER
Random Forest | Complex (minmax)	0.802	0.902	0.976	0.937	0.970	0.992	0.691	0.077
Random Forest | GAT (minmax)	0.849	0.927	0.971	0.947	0.967	0.990	0.763	0.076
Random Forest | GraphSAGE (minmax)	0.764	0.885	0.982	0.930	0.957	0.988	0.644	0.092
Random Forest | RGCN (minmax)	0.726	0.867	0.971	0.916	0.903	0.963	0.570	0.101
Random Forest | gene expression (none)	0.702	0.854	0.995	0.919	0.968	0.990	0.571	0.091
Random Forest | weighted RGCN (minmax)	0.790	0.896	0.971	0.931	0.966	0.990	0.675	0.082
Sklearn Mlp | Complex (minmax)	0.914	0.958	0.982	0.969	0.982	0.995	0.861	0.041
Sklearn Mlp | GAT (minmax)	0.865	0.934	0.966	0.949	0.959	0.988	0.768	0.068
Sklearn Mlp | GraphSAGE (minmax)	0.851	0.932	0.947	0.939	0.969	0.992	0.715	0.067
Sklearn Mlp | RGCN (minmax)	0.500	0.776	1.000	0.874	0.500	0.776	0.000	0.224
Sklearn Mlp | gene expression (none)	0.500	0.776	1.000	0.874	0.543	0.798	0.000	0.226
Sklearn Mlp | weighted RGCN (minmax)	0.641	0.838	0.982	0.901	0.911	0.974	0.295	0.131
Svm | Complex (minmax)	0.953	0.984	0.961	0.972	0.987	0.996	0.888	0.040
Svm | GAT (minmax)	0.500	0.776	1.000	0.874	0.945	0.985	0.000	0.110
Svm | GraphSAGE (minmax)	0.500	0.776	1.000	0.874	0.950	0.987	0.000	0.079
Svm | RGCN (minmax)	0.869	0.946	0.929	0.936	0.944	0.983	0.730	0.080
Svm | gene expression (none)	0.930	0.967	0.979	0.973	0.986	0.996	0.876	0.038
Svm | weighted RGCN (minmax)	0.500	0.776	1.000	0.874	0.964	0.990	0.000	0.097
Xgboost | Complex (minmax)	0.840	0.928	0.934	0.930	0.950	0.986	0.693	0.076
Xgboost | GAT (minmax)	0.876	0.942	0.961	0.950	0.969	0.991	0.787	0.063
Xgboost | GraphSAGE (minmax)	0.782	0.901	0.918	0.908	0.924	0.980	0.589	0.100
Xgboost | RGCN (minmax)	0.806	0.907	0.958	0.931	0.915	0.972	0.667	0.092
Xgboost | gene expression (none)	0.874	0.940	0.966	0.952	0.963	0.989	0.783	0.061
Xgboost | weighted RGCN (minmax)	0.898	0.962	0.924	0.942	0.962	0.989	0.766	0.066

Graph protein embeddings are mostly better than gene expression across all metrics

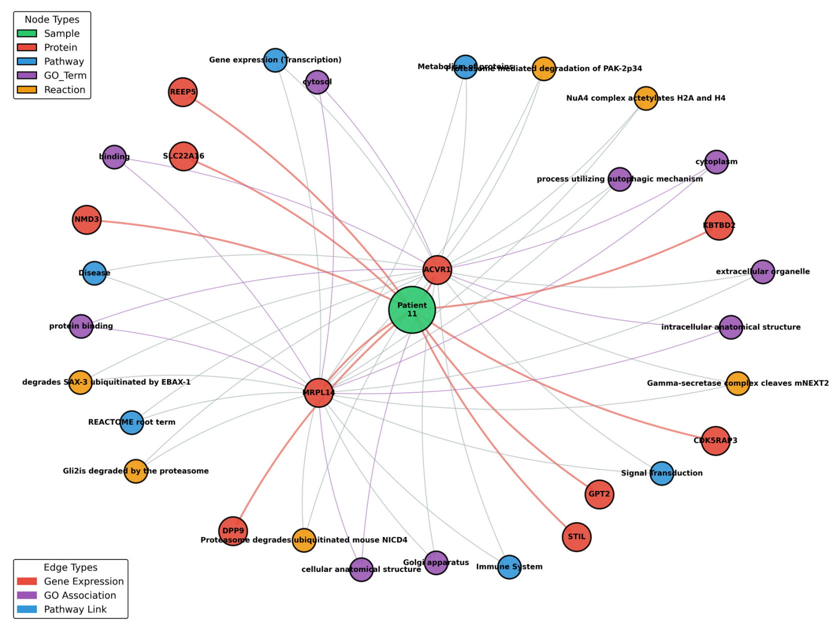
- ComplEx embeddings are top ranked
- ComplEx, GraphSAGE and GAT perform well for MLP
- SVM gene expression (93% acc) outperformed by ComplEx (95.3%)
- Recall of 100% and precision 77.6% are due to small dataset with imbalance
- RGCN protein embeddings are generally not great for prediction in different models

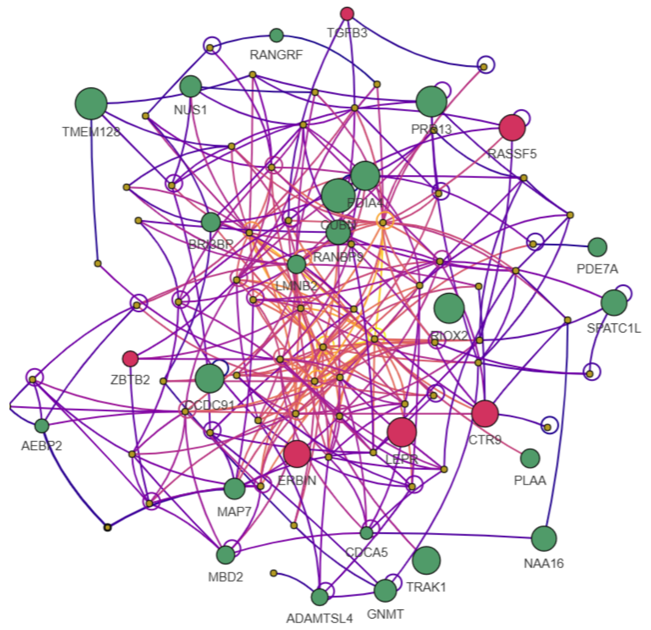
Accuracy over 10 seeds:

- ComplEx SVM **top robust**
- weighted RGCN nor reliable across models



HAN patient-level and SHAP-ComplEx population-level explanations provide complementary insights





HAN-based patient-specific explanation

HAN-derived heterogeneous subgraphs highlight proteins with highest gradient-based influence on sepsis prediction, propagating influence through functional (GO) and pathway-level contexts to drive individual predictions.

ComplEx-based consensus PPI drivers

SHAP analysis across all ComplEx ML models yields a PPI network enriched in hub proteins, suggesting shared regulation.

Discussion

- **Data scarcity:** dataset limited to 53 patients undermines training and lead to testing instability.
- **Class imbalance:** Only 22% healthy samples introduces bias toward septic phenotypes.
- **Temporal nature:** 163 samples spanning multiple time points require specialized modeling.
- **Unoptimized knowledge graph:** KG construction used no filtering, resulting in dense PPIs, pathway-dominated topology, and high reaction embeddedness.
- **KG structural bias:** Underrepresented edge types and highly connected pathway nodes may dominate message passing and attribution signals.

Future Directions

- **Adaptive KG refinement:** Filter and simplify KG, optimize architecture, link to interpretability, then retrain embeddings to compare performance.
- **Temporal modeling:** account for data temporality.
- **Interpretability convergence:** assess the interpretability approaches in more depth, reapply to optimized KG, and check convergent patterns.
- **Advanced visualization:** Highlight identified biomarkers and interactive exploration of insights.