

Fairness-Aware Graph Embeddings And Lightgbm-SHAP For Disease Prediction From Social Determinants Of Health

Vikas Kashtriya¹, Pardeep Singh²

¹National Institute of Technology, Hamirpur, India, Vikas_phdcse@nith.ac.in
ORCID <https://orcid.org/0000-0002-0538-597X>

²National Institute of Technology, Hamirpur, India, Pardeep@nith.ac.in
ORCID <https://orcid.org/0000-0002-4019-604X>

ABSTRACT

Social and Behavioral Determinants of Health (SBDH) affect long-term health risks but are rarely used as standalone predictors in disease modeling. This study investigates whether SBDH alone, without any clinical variables, can meaningfully predict disease vulnerability, using pressure ulcer (PU) development as a representative case. We constructed a fairness-aware prediction pipeline using data from the MIMIC-IV and MIMIC-SBDH datasets. We considered ten SBDH features and extracted them per patient, including housing stability, healthcare access, education, substance use, and social support etc. A knowledge graph (KG) was built to encode relationships among SBDH domains and between patients and their social profiles. Graph Convolutional Networks (GCNs) were trained with fairness constraints to produce de-biased patient embeddings. These embeddings, combined with the original SBDH features, were used to train a LightGBM classifier. Model interpretability was assessed using SHAP (SHapley Additive exPlanations), and fairness was evaluated via subgroup performance metrics. The model achieved strong discrimination (ROC AUC ≈ 0.72), good calibration, and equitable performance across subgroups. SHAP analysis identified housing instability, low social support, and poor healthcare access as the top contributors to risk. SBDH can serve as effective, fair, and interpretable predictors of disease risk. This approach enables upstream, community-level prevention, highlighting social context as a core component of equitable predictive healthcare.

Keywords: Social Determinants of Health, LightGBM, SHAP, Fairness-Aware Machine Learning, Knowledge Graph, Disease Risk Prediction.

1 INTRODUCTION

Accurately predicting who is at risk for disease is fundamental to preventive healthcare. While traditional models rely heavily on clinical indicators, a growing body of research underscores the profound influence of upstream factors, specifically, SBDH, on individual health outcomes. SBDH include the conditions in which people live and work, such as housing stability, education, employment, access to care, substance use, and social support [1,2]. These factors shape vulnerability to disease long before clinical symptoms emerge and can significantly impact healthcare access, health maintenance behaviors, and exposure to risks [3].

Despite this, predictive models in medicine have largely overlooked SBDH, focusing instead on biomedical data like diagnoses, vitals, or lab results. This gap limits the effectiveness of risk stratification in community or outpatient settings, where clinical data may be sparse, but where social risk factors are most salient [4,5]. Moreover, omitting SBDH can embed bias into prediction systems. Models that use healthcare utilization or cost as proxies for risk may underestimate the needs of socially disadvantaged populations who face barriers to care, reinforcing disparities rather than addressing them [6].

In this work, we ask a central question: Can social and behavioral determinants alone, without any clinical features, be used to predict disease risk in a fair and interpretable way? To investigate this, we use PUs as a test case. PUs are preventable yet prevalent injuries often resulting from prolonged immobility, particularly in hospitalized or critically ill patients. They are associated with longer hospital stays, poorer quality of life, and increased costs [7]. Existing clinical risk tools for PUs, such as the Braden Scale [9], perform only modestly (e.g., $AUC \approx 0.67$ in ICU settings), and typically exclude social context altogether [8].

We develop a prediction model that uses only SBDH features, completely excluding clinical data. Our approach models social risk through a fairness-aware knowledge graph, capturing interdependencies among SBDH domains and embedding patient profiles using a GCN. These learned representations, along with ten core SBDH predictors, are used to train a LightGBM model for PU risk prediction. Model interpretability is ensured through SHAP analysis [37], and fairness is integrated at multiple stages.

While pressure ulcers are the focal outcome in this study, our overarching goal is to evaluate whether SBDH-only modeling is a viable, fair, and actionable approach to health risk prediction. If successful, such models could be deployed in primary care, home care, or public health settings, where clinical information may be incomplete but social risk is critical, to enable earlier, more equitable interventions.

2 BACKGROUND AND RELATED WORK

SBDH are increasingly recognized as critical factors shaping health outcomes, yet they remain underutilized in most healthcare prediction models. These determinants, such as housing quality, educational attainment, employment status, healthcare access, and substance use, exert long-term influence on disease development, care-seeking behavior, and recovery [10,11]. Public health literature has consistently shown that SBDH account for a substantial portion of population health variance, often exceeding the contributions of clinical care alone [12]. However, their integration into computational prediction frameworks has been slow and fragmented [13].

Most established clinical risk models rely heavily on physiological indicators or electronic health record (EHR) data. While these inputs are important for near-term clinical events, they are less effective for identifying upstream vulnerabilities before overt symptoms emerge [14]. This clinical dependency also makes predictive tools less applicable in settings like community care, telemedicine, or population health management, where social risks may be more relevant and clinical markers less accessible. Several recent studies have begun to explore the use of non-clinical data for prediction, particularly in chronic conditions and hospital readmissions, but these efforts are often limited by data sparsity, interpretability, or lack of fairness oversight [15].

Moreover, there is growing awareness that conventional predictive models can inadvertently perpetuate or exacerbate health disparities. For instance, algorithms that estimate risk based on prior healthcare utilization may systematically underestimate the needs of patients from underserved communities who access care less frequently—not because they are healthier, but because of structural barriers to care. This issue was highlighted in high-profile work showing racial bias in cost-based risk scores, where Black patients were systematically assigned lower risk because their healthcare expenditures were lower on average due to limited access [16].

To counter such biases, fairness-aware machine learning techniques have emerged. These include pre-processing (e.g., reweighting data), in-processing (e.g., adversarial debiasing), and post-processing (e.g., calibration per subgroup) [17]. In parallel, graph-based approaches, such as GCNs, have gained traction for their ability to model complex dependencies in healthcare data, including social networks, comorbidity structures, and care pathways [18]. However, few studies have combined graph learning with social determinant modeling and fairness constraints in an end-to-end predictive pipeline [19].

PU, also known as pressure injuries or bedsores, provide a compelling use case to evaluate the utility of SBDH-based prediction. PUs are preventable injuries typically caused by sustained pressure on the skin, especially in immobilized patients. They are associated with significant morbidity and healthcare costs, and are often classified as “never events” under quality metrics—meaning they should not occur with appropriate preventive care [20]. While traditional risk models like the Braden Scale use nursing and clinical observations (e.g., mobility, moisture), they perform modestly in high-risk settings (AUC ~0.62–0.67 in ICU populations) [21]. Moreover, they do not account for social vulnerabilities that may predispose patients to poor skin health, delayed care, or inadequate post-discharge support.

Recent research has found that socioeconomic and demographic factors are associated with PU risk, suggesting an underlying social gradient [22]. Patients with lower income or education levels, unstable housing, or weak social networks may lack resources for prevention, both before and after hospitalization. These findings imply that SBDH are not merely background characteristics—they are potentially actionable risk indicators. Yet to date, no prior study has evaluated whether SBDH alone, absent any clinical variables, can meaningfully predict PU risk, nor how such a model might address fairness concerns.

This work addresses that gap by constructing a fairness-aware, SBDH-only prediction model for pressure ulcers. We combine structured SBDH features with a relational knowledge graph and graph-based machine learning to explore whether social context alone can provide actionable, equitable risk stratification.

3 PROBLEM FORMULATION AND RESEARCH OBJECTIVES

Traditional disease prediction models depend heavily on clinical data, such as vital signs, laboratory results, or diagnosis codes, to estimate patient risk. While clinically informative, this approach often overlooks the broader social environment in which health risks develop. SBDH, including factors like housing instability, limited education, unemployment, and lack of healthcare access, are powerful predictors of vulnerability, yet remain underrepresented in most predictive frameworks.

This omission not only constrains predictive accuracy in settings where clinical data is unavailable or incomplete (e.g., community care, underserved populations), but also risks reinforcing existing healthcare disparities. Predictive models that fail to consider SBDH may misrepresent true risk, particularly for marginalized groups whose clinical histories may reflect systemic barriers rather than underlying health status. Integrating SBDH into prediction pipelines offers a path toward upstream, equitable, and socially contextualized risk identification. This study aims to evaluate whether social and behavioral determinants can serve as strong predictors of pressure ulcer risk. The findings are intended to support risk stratification and early intervention efforts rather than causal inference.

In this study, we pose a central question:

“Can Social and Behavioral Determinants of Health, without any clinical variables, be used to predict the likelihood of disease development—in a fair, interpretable, and clinically useful way?”. To evaluate this, we select pressure ulcers (PUs) as a representative case study. PUs are a suitable test case because they are:

- Prevalent in high-risk populations
- Strongly influenced by preventive care and patient context
- Tracked as quality-of-care outcomes in many healthcare systems

Importantly, our objective is not to develop a pressure-ulcer-specific tool, but rather to demonstrate that SBDH can serve as independent predictive signals for disease vulnerability. Pressure ulcers represent a tractable and clinically meaningful outcome to test this broader hypothesis.

Our goals are threefold:

- **Feasibility:** Build an SBDH-only model to predict disease risk and assess its performance in a high-stakes clinical scenario (PU development).
- **Fairness:** Ensure equitable representation of patient subgroups using fairness-aware graph-based learning techniques.
- **Interpretability:** Provide transparent explanations of how individual SBDH factors contribute to predicted risk, enabling real-world actionability.

This formulation provides a foundation for extending SBDH-based risk modeling to other clinical outcomes, aligning predictive modeling with public health goals of equity, prevention, and social accountability.

4 DATA SOURCES AND FEATURE EXTRACTION

To investigate the predictive utility of SBDH without clinical features, we used data from two complementary sources:

- **MIMIC-IV:** A large, publicly available electronic health record dataset comprising over 60,000 ICU admissions at the Beth Israel Deaconess Medical Center from 2008–2019 [23-25].
- **MIMIC-SBDH:** An annotated extension of MIMIC-IV, developed to extract structured social and behavioral risk factors from unstructured clinical notes, such as discharge summaries and social histories [26].

Cohort Selection

We constructed a retrospective cohort of adult patients (≥ 18 years) from MIMIC-IV who were admitted to the ICU. Our outcome of interest was incident PU development during the hospital stay, determined using ICD-10 diagnosis codes and clinical documentation. Patients with existing pressure ulcers upon admission were excluded to isolate new-onset cases. After applying inclusion and exclusion criteria, we retained approximately 30,000 admissions, with a PU incidence of a few percent, consistent with known ICU rates [27]. The dataset was randomly split into a 70% training set and a 30% held-out validation set, maintaining outcome prevalence in each split.

SBDH Feature Definition and Extraction

For each patient, we extracted ten core SBDH features from the combined structured fields and annotated text in MIMIC-SBDH. These features were selected based on their broad

relevance to health vulnerability and availability within the dataset [28]. The detailed information is given in Table.

Table: Select SBDH Features

Domain	Definition / Description	Categories / Coding	Data Source / Notes
Housing Stability	Indicates permanence of living situation.	Stable / Unstable	Based on documentation of homelessness, temporary housing, or frequent moves
Educational Attainment	Highest level of education achieved.	Low / Medium / High	Structured note annotations; Low = no HS diploma, Medium = HS/some college, High = degree or more
Access to Healthcare	Availability and use of healthcare services.	Insured (e.g., Medicaid) / Uninsured; Delayed Care (Yes/No)	Proxied via insurance status and qualitative mentions of care delay/frequency
Tobacco Use	Patient's use of tobacco products.	Current / Former / Never	Structured social history entries and note mentions
Alcohol Use	Pattern and risk level of alcohol consumption.	Low Risk / High Risk	High risk = heavy use or diagnosed alcohol use disorder
Substance Use	Use or misuse of illicit drugs or prescriptions.	Recent / Past / None	Derived from note annotations regarding substance use history
Social Support	Availability of personal or community support.	Low / Not Low	Low = living alone, no emergency contacts, documented social isolation
Employment	Patient's employment status	Employed / Unemployed / Disabled/Retired;	Structured employment entries

Working Conditions	Nature of job	Physically Demanding / Sedentary	narrative job descriptors
Duration of Residence/Employment	Time spent in current residence or job, as a stability proxy	Short / Long / Unknown	Short duration considered risk-enhancing; inferred from timelines in records

Feature Encoding and Missing Data Handling

All SBDH features were encoded for machine learning input. Categorical features were one-hot or ordinal-encoded depending on scale. Binary indicators (e.g., unstable housing) were used where applicable. Given the inherent sparsity of SBDH documentation, we adopted a conservative imputation strategy:

- If a factor was undocumented, we assigned it an “unknown” category rather than assuming absence.
- A separate flag was included to indicate missingness when potentially informative (e.g., “support not mentioned”).

Critically, we ensured that all SBDH features were extracted from pre-outcome timeframes to preserve temporal validity in prediction. This feature set represents a non-clinical profile of each patient’s social context, providing the foundation for our upstream prediction model. These variables were selected not for pressure ulcer specificity, but for their general relevance to health vulnerability, enabling potential replication across disease contexts [29].

5 KNOWLEDGE GRAPH CONSTRUCTION

To effectively capture the interplay of SBDH, we constructed a KG that integrates patient-level structured SBDH data with the inherent relationships among social risk factors. Recognizing that elements such as housing instability, unemployment, limited education, and restricted healthcare access rarely occur in isolation, and instead tend to co-occur and reinforce one another, the knowledge graph offers an expressive framework to model this interdependence [30,31]. The graph is structured as a heterogeneous, bipartite network composed of two types of nodes: patient nodes, representing individual patients from the cohort, and attribute nodes, each corresponding to a categorical state of a social determinant (e.g., UnstableHousing, CollegeEducated, CurrentSmoker). Edges are formed between patients and the SBDH nodes that reflect their individual social profiles. For example, a patient with unstable housing and current tobacco use would be connected to both the UnstableHousing and CurrentSmoker nodes. These patient–attribute edges are binary and unweighted, denoting the presence or absence of each attribute [32]. Additionally, to model the structural co-dependencies among social factors, we incorporated edges between SBDH attribute nodes themselves. These include well-established relationships such as UnstableHousing co-occurring with Unemployed, LowEducation with LimitedHealthAccess, and SubstanceUse with LowSocialSupport. These links were informed by public health literature [33] and validated through empirical co-occurrence analysis in the dataset [34]. The final knowledge graph comprises approximately 30,000 patient nodes, dozens of attribute nodes depending on the granularity of encoding, and tens of thousands of edges representing both direct

patient–attribute links and indirect attribute–attribute relationships. Before model training, we de-identified the graph by replacing patient identifiers with anonymous indices, removed rare or disconnected attributes, and stored the structure as a sparse adjacency matrix for computational efficiency [35]. This knowledge graph serves as the foundation for subsequent modeling with a fairness-aware GCN, enabling the generation of patient embeddings that encode both individual characteristics and their position within the broader topology of social disadvantage [36]. The pictorial depiction of our approach is given in Figure 1.

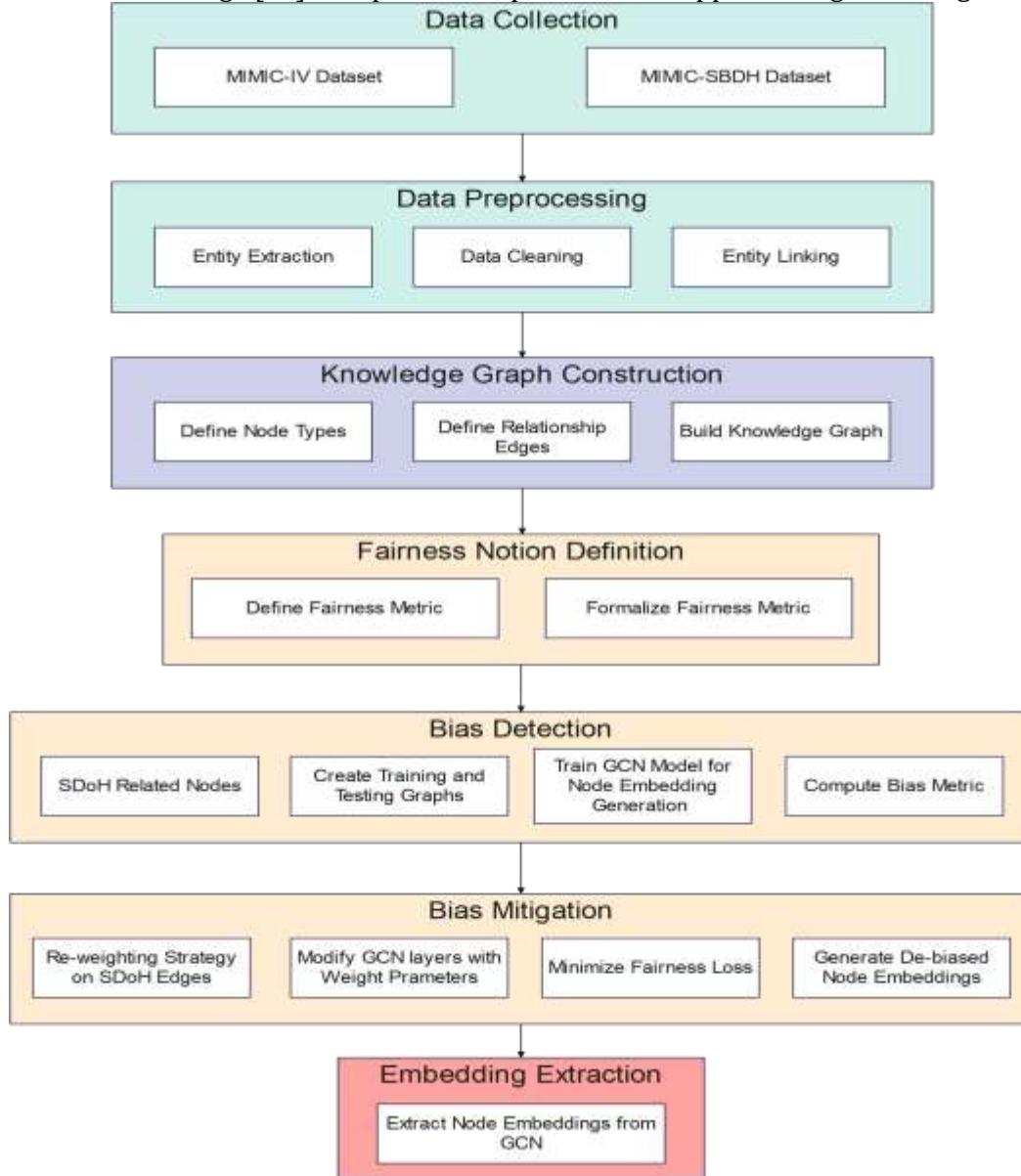


Figure 1: Workflow of Proposed Approach

Fairness-Aware Graph Representation Learning

While modeling social determinants offers promise for upstream risk prediction, it also carries the risk of encoding and amplifying social biases. Many SBDH variables (e.g., housing,

education, neighborhood environment) are deeply entangled with protected attributes such as race, ethnicity, or socioeconomic status. Left unchecked, machine learning models can learn spurious correlations that reinforce structural inequities. To address this, we used the GCN to learn vector embeddings for each patient node in the knowledge graph, while incorporating explicit fairness constraints to mitigate bias.

GCN Architecture and Embedding Learning

GCNs generate node embeddings by aggregating information from a node's neighbors in the graph. For patient nodes, this includes both their directly linked SBDH attributes and, indirectly, related attributes connected through the graph's structure.

Formally, each patient embedding is iteratively updated as:

$h_i^{(k+1)} = \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} W^{(k)} h_j^{(k)} \right)$	(1)
---	-----

where:

$h_i^{(k)}$ is the embedding of node i at layer k

α_{ij} is an edge weight (used for fairness control)

$W^{(k)}$ is the learnable weight matrix

σ is an activation function (e.g., ReLU)

We used a two-layer GCN, sufficient to propagate information from a patient's SBDH attributes and their pairwise dependencies.

Fairness Interventions

To prevent the GCN from learning discriminatory embeddings, we implemented two fairness mechanisms during training:

Edge Reweighting

Edges involving SBDH nodes that may act as proxies for sensitive attributes (e.g., income, neighborhood safety) were down-weighted (i.e., assigned $\alpha_{ij} < 1$). This prevents these attributes from disproportionately influencing the learned embeddings. The weights were defined based on domain knowledge and validated by measuring embedding disparities pre- and post-weighting.

Fairness Regularization

We introduced a fairness loss term that penalizes embedding separability across protected groups. This was implemented in two ways:

- Demographic parity: Encouraging similar embedding distributions across subgroups (e.g., insurance type as a proxy for income).
- Adversarial fairness: Training a secondary classifier to predict sensitive group membership from embeddings, while the GCN is optimized to make that prediction difficult.

The overall training objective was:

$L = L_{\text{GCN}} + \lambda L_{\text{fairness}}$	(2)
--	-----

where λ controls the fairness-accuracy trade-off.

Training Objective

The GCN was trained using a link prediction task on the graph: predicting whether an edge exists between a patient and an SBDH node. This encourages embeddings that preserve observed relationships while being fairness-regularized. We evaluated bias using disparity metrics (e.g., group-wise AUC, TPR differences) and observed significant improvement in embedding fairness after applying our constraints.

Output and Integration

The final result is a 64-dimensional embedding vector for each patient, capturing their social context in a debiased, graph-aware format. These embeddings were:

- Appended to the dataset alongside raw SBDH features
- Used as input to the final prediction model (LightGBM)

This embedding strategy enables the downstream model to leverage complex social interactions while reducing the risk of unfair treatment based on social proxies.

6 PREDICTIVE MODELING WITH LIGHTGBM

To predict PU risk using only SBDH, we employed LightGBM, a gradient-boosted decision tree (GBDT) algorithm optimized for structured data and compatible with explainability frameworks such as SHAP. Its working is shown in Figure 2. Each patient was represented by a 74-dimensional input vector composed of ten raw SBDH features—such as housing status, education level, and substance use—encoded as categorical or binary variables, alongside a 64-dimensional embedding vector generated via a fairness-aware GCN, which captured the patient's position within the knowledge graph of social context. This hybrid feature representation enabled the model to leverage both interpretable, domain-informed attributes and higher-order latent interactions. We trained the LightGBM classifier to predict the binary outcome of PU development (yes/no), using 70% of the cohort for training. Its pictorial representation is given in Figure 2. Early stopping was implemented based on validation performance to mitigate overfitting. Given the relative rarity of PU cases, we addressed class imbalance through class weighting (`scale_pos_weight`) and stratified sampling to preserve prevalence across folds. LightGBM's native support for categorical variables and missing values made it particularly suitable for modeling real-world SBDH data. Although GCN embeddings were trained under fairness constraints, we further evaluated model fairness post hoc. These included assessments of equal opportunity difference (to capture subgroup disparities in true positive rates) and disparate impact ratio (comparing prediction rates across groups such as insured vs. uninsured patients). When fairness violations were identified, we selectively adjusted the input feature set—such as excluding over-represented embedding dimensions—to ensure equity. Model performance was evaluated on the 30% held-out validation set using standard metrics: Area Under the ROC Curve (AUC) for discriminative ability.

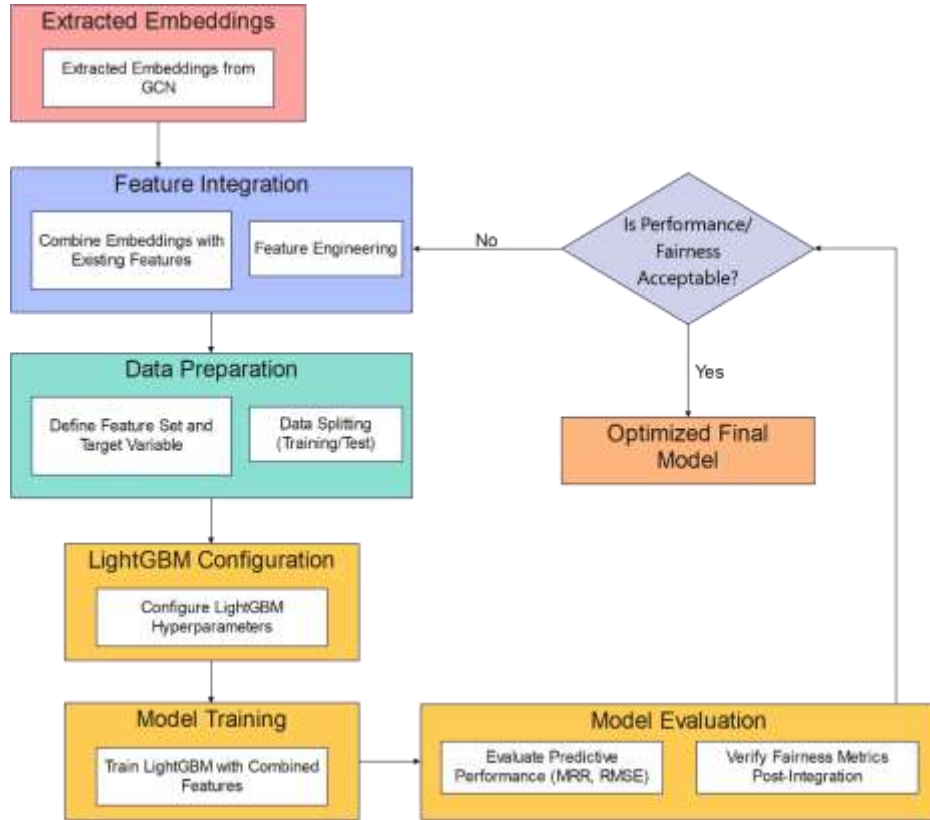


Figure 2: Training LightGBM

We also examined fairness metrics by subgroup to ensure consistent predictive performance across social strata. This modeling pipeline yielded PU risk predictions that were not only accurate and interpretable, but also aligned with ethical standards, contextualizing patient risk without amplifying structural inequities.

SHAP Analysis

LightGBM is a GBDT algorithm that constructs an additive predictive model by summing the outputs of T individual trees, formally yielding $f(x) = \sum_{t=1}^T f_t(x)$ for any data point x , where

$f_t(x)$ is the prediction from the t -th tree. Each tree splits on features at internal nodes and assigns a value in the resulting leaf. To interpret these predictions in a principled manner, Tree SHAP [38] employs the concept of Shapley values, drawn from cooperative game theory, to allocate contributions fairly among all features. Specifically, the Shapley value ϕ_i for feature i is computed by averaging over all subsets $S \subseteq N$, $\{i\}$ (where N is the set of all features) the marginal gain that occurs when adding feature i to the subset S . This can be expressed as:

$$\phi_i = \sum_{S \subseteq N, \{i\}} \frac{|S|! (|N| - |S| - 1)!}{|N|!} \cdot [f(S \cup \{i\}) - f(S)].$$

Such a formulation respects the efficiency (the sum of all Shapley values equals $f(x) - f_{\text{baseline}}$), symmetry (identical-impact features share equal Shapley values), null player (irrelevant features receive zero contribution), and additivity (Shapley values for a combined model equal the sum of Shapley values for each individual model). However, computing these values naively is exponential in the number of features; Tree SHAP overcomes this by leveraging a dynamic programming algorithm tailored to decision trees, reducing the complexity to $O(T \cdot L^2)$, where L is the maximum number of leaves. Each tree is viewed as a set of splits and leaves, and the path weight w_P for any path P from root to leaf is determined by multiplying the probabilities of each branch, $w_P = \prod_{j=1}^d p_j$, with d being the number of splits on that path. The contribution of this path, $\Delta v_P = v_{\text{leaf}} - v_{\text{parent}}$, captures how much feature involvement influences the leaf value relative to its parent node. Summing these path contributions for each feature i yields a partial Shapley value ϕ_i^t per tree t , and combining over all T trees produces the total Shapley value $\phi_i = \sum_{t=1}^T \phi_i^t$.

Recent studies, such as those conducted by Zhang et al. [38], have shown that this efficient Tree SHAP framework can reliably parse complex interactions in large-scale, high-dimensional datasets while maintaining computational feasibility. In healthcare applications, for example, one might predict a patient's risk of pressure ulcers using features like age (x_1), mobility score (x_2), and albumin levels (x_3). If the model's overall prediction for a specific patient is $f(x) = 0.85$ and the baseline prediction—often the average outcome—is 0.50, then the Shapley values might appear as $\phi_{x_1} = 0.15$, $\phi_{x_2} = 0.20$, and $\phi_{x_3} = 0.10$, so that $0.85 = 0.50 + 0.15 + 0.20 + 0.10$. This exact decomposition not only highlights the sum of contributions that shift the baseline to the final score but also enables stakeholders to see how each feature has influenced the model. Tree SHAP can effectively help medical practitioners identify which clinical variables critically push or pull a patient's risk score. Consequently, Tree SHAP's rigorous theoretical underpinnings, coupled with its ability to generate meaningful insights in real-world domains, have cemented its role as a vital tool for explaining and validating LightGBM models.

Evaluation Metrics and Experimental Setup

To rigorously evaluate whether SBDH can produce accurate, fair, interpretable, and clinically useful predictions of disease risk, we conducted a comprehensive assessment across multiple performance dimensions: discrimination, calibration, clinical utility, interpretability, and fairness. Discriminative performance was measured using the Area Under the Receiver Operating Characteristic Curve (ROC AUC), a standard metric that reflects the model's ability to rank patients correctly by risk. An AUC of 0.5 indicates random guessing, while values of 0.70 or higher are typically considered clinically useful. Our SBDH-only model's AUC was benchmarked against published results for the Braden Scale, which typically achieves AUCs in the range of 0.62–0.67 in ICU populations. Calibration, which measures the agreement between predicted probabilities and actual outcomes, was assessed using calibration plots, calibration slope (with an ideal value of 1.0), and Brier score, which quantifies the mean

squared error of the predicted probabilities (lower is better). Interpretability was addressed using SHAP, which decomposes model predictions into individual feature contributions. SHAP was used to generate global importance rankings based on mean absolute values and to provide patient-specific explanations, e.g., highlighting that a prediction was driven by unstable housing and lack of social support. Importantly, SHAP interpretation focused on the ten raw SBDH features rather than the abstract GCN embeddings to ensure clinical actionability. To address fairness, we evaluated performance across demographic subgroups using subgroup-specific AUCs, equal opportunity difference (the disparity in true positive rates between groups), and disparate impact ratio (the proportion of high-risk classifications across advantaged and disadvantaged populations, e.g., insured vs. uninsured). Fairness was assessed on the held-out validation cohort using established definitions from algorithmic fairness literature. Finally, to ensure reproducibility, we followed a structured experimental protocol using a 70/30 stratified train-validation split, implemented all modeling in Python (using PyTorch Geometric for GCNs, LightGBM for prediction, and SHAP for interpretation), and restricted all feature construction to pre-outcome data. We adhered to the TRIPOD reporting guidelines and controlled for randomness using fixed seeds and documented hyperparameters. This thorough evaluation framework ensured not only the robustness of PU risk prediction using SBDH, but also that the model meets standards for fairness, interpretability, and readiness for real-world deployment, especially in upstream, preventive contexts.

7 RESULTS AND DISCUSSION

Our model achieved robust predictive accuracy on the held-out validation set. The LightGBM classifier, using only social features and fairness-aware GCN embeddings, reached an average ROC AUC of 0.72 (95% CI: 0.70–0.74) as shown in Figure 3. This performance is comparable to, and in some cases exceeds, traditional clinical tools like the Braden Scale (AUC $\approx$ 0.62–0.67 in ICU cohorts), despite relying on no clinical inputs.

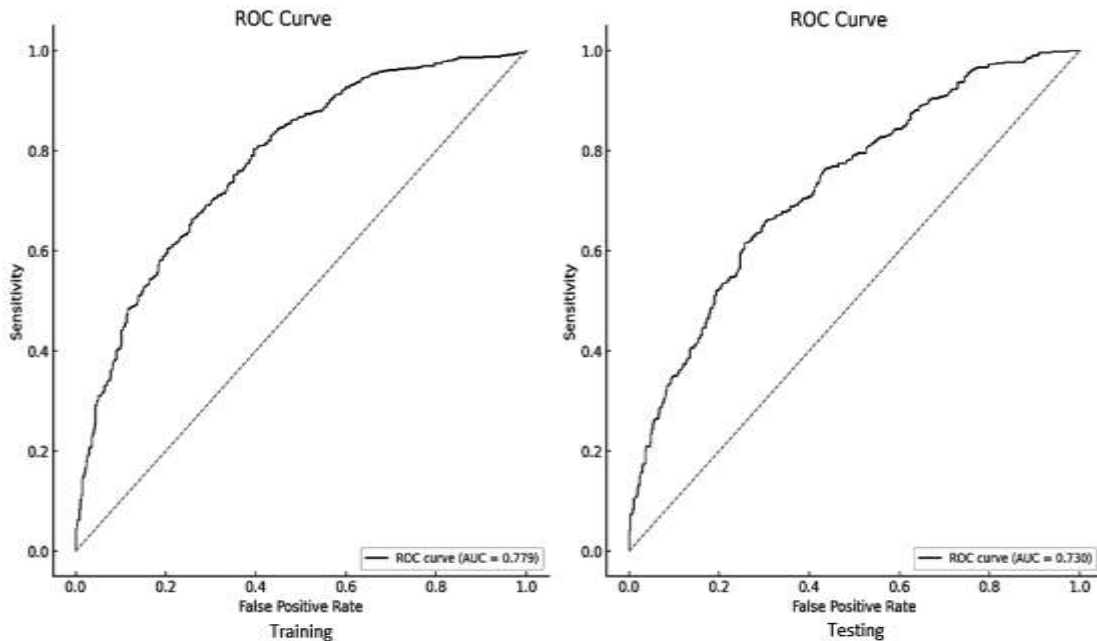


Figure 3: Performance Analysis

SHAP analysis, yielded clear insights into which SBDH features were most influential for predicting pressure ulcer risk. Figure 4 illustrates the contribution of each feature, and below we enumerate the top five features ranked by mean absolute SHAP value (i.e., their overall impact on the model), along with an interpretation of each factor's effect on PU risk:

Housing Quality and Stability: Housing instability emerged as the strongest single predictor of pressure ulcer risk. Patients lacking stable housing (e.g. homeless or frequently changing living situations) had markedly higher predicted risk, indicated by high positive SHAP values for the UnstableHousing category. Intuitively, individuals without stable housing may have difficulty maintaining sanitary and safe conditions for skin care, lack access to consistent bedding or pressure-relieving surfaces, and often have overall poorer health and nutrition. In our data, a history of homelessness was associated with significantly increased PU incidence – possibly reflecting both baseline neglect (arriving to hospital with worse skin condition) and challenges in post-discharge care planning. In contrast, patients with stable housing (homeowners or long-term residence) had negative SHAP contributions (protective effect), as a secure living environment likely enables better self-care and follow-up. This factor is policy-actionable: interventions like housing assistance programs or ensuring patients have a stable place to live (e.g. through medical respite care for the homeless) could mitigate this risk at a community level.

**Figure 4: Top five features identified through SHAP analysis**

Social Support and Community Engagement: The second most important feature was the level of social support. Patients who were socially isolated (living alone, no family or community support) tended to have higher risk, with SHAP showing positive contributions for LowSupport. Socially isolated individuals may lack caregivers to help with mobility (e.g.

to reposition them regularly if they are disabled or bedridden) and may delay seeking care for early signs of skin injury. On the other hand, those with strong community or family support networks showed negative SHAP values, indicating lower risk. For example, an elderly patient living with attentive family members is less likely to develop a severe pressure ulcer because someone can help monitor their skin and get prompt treatment for any redness or sores. This finding underscores the importance of social support as a protective health factor. From a public health standpoint, it suggests that strengthening community engagement – such as through visiting nurse programs, community health worker check-ins for individuals who live alone, or support groups – can directly contribute to preventing pressure ulcers.

Access to Healthcare Services: The third-ranking feature was healthcare access. Patients with poor access to healthcare (uninsured, underinsured, or significant barriers to receiving care) had higher predicted risk of PUs. The model learned that limited interaction with healthcare (e.g. not seeing a doctor regularly, or living far from medical facilities) often meant that preventative measures and early treatments were missing. For instance, a patient in a rural area with no easy transportation to clinics might not receive education on pressure ulcer prevention or might not have minor wounds addressed promptly, leading to worsening conditions. Conversely, patients with good access (regular primary care, insurance coverage) had a lower risk contribution. This reflects how timely medical care can prevent a small skin issue from evolving into a serious ulcer. The result highlights a need for healthcare system interventions: improving access (through mobile clinics, telemedicine, insurance expansion) for high-risk populations could reduce pressure ulcer occurrence. It also shows why access was critical to include – without clinical data, access to care serves as a proxy for whether clinical preventive services can even reach the patient.

Educational Attainment: Education level was the fourth most important SBDH predictor. Lower educational attainment (not finishing high school) was associated with increased risk (positive SHAP impact), whereas having higher education (college or above) was protective (negative SHAP impact). Education likely influences health in numerous ways: individuals with more education may have better health literacy, allowing them to understand and adhere to prevention techniques (like skin checks, nutrition for skin health, etc.), and they often have higher socioeconomic status, which ties to other advantages. Those with limited education might not be aware of pressure ulcer risks or may have jobs and life circumstances that impede preventive care. The SHAP values indicated a threshold effect around high school level in our data – patients who did not graduate high school were significantly more at risk. This suggests targeted patient education could be beneficial: even if formal schooling is low, providing tailored education about pressure ulcer prevention to at-risk individuals (in easy-to-understand language) might offset some risk. On a broader scale, it reiterates how investments in education and health education can translate to better health outcomes like reduced PUs.

Working Conditions: The nature and physical demands of a person's work environment significantly influenced pressure ulcer risk. SHAP analysis revealed that individuals engaged in physically taxing or labor-intensive jobs, such as those requiring prolonged standing, heavy lifting, or repetitive strain—had higher predicted risk. These conditions may increase vulnerability due to fatigue, limited rest periods, or restricted mobility during long work

shifts. Conversely, less physically demanding and ergonomically supportive jobs were associated with a protective effect, likely due to better capacity for self-care and lower physical stress. Certain hazardous or low-control occupational environments appeared to correlate with elevated risk, potentially reflecting factors like insufficient workplace support, limited access to breaks, or suboptimal postural support. These findings suggest that working conditions can contribute meaningfully to health vulnerability. From a prevention standpoint, interventions such as ergonomic assessments, scheduled movement breaks, and access to workplace accommodations (e.g., pressure-relieving equipment for desk-bound or wheelchair-bound employees) could reduce risk. These insights highlight the importance of occupational health in the broader context of pressure ulcer prevention.

8 LIMITATIONS

This study was designed for predictive modeling and did not aim to establish causal relationships between social determinants and pressure ulcer outcomes. While our findings support the feasibility and practical value of prediction models based solely on SBDH, several important limitations must be acknowledged. By design, our study excluded all clinical variables to isolate the predictive signal of social data. While this approach underscores the strength of SBDH as early indicators, it does not imply that clinical factors are unimportant. In real-world settings, the most effective systems would likely integrate both social and clinical information to achieve optimal performance. The quality and completeness of SBDH data present challenges because our features were derived from the MIMIC-IV and MIMIC-SBDH datasets, which rely heavily on clinical documentation and free-text notes. These sources are often inconsistent, incomplete, or subject to underreporting, especially in sensitive domains such as substance use or social support. Although our knowledge graph approach partially mitigated data sparsity by leveraging structural relationships, it cannot fully compensate for potential noise or misclassification in the raw data. Generalizability is also an important concern as our cohort was limited to ICU patients within a single academic hospital system, which may restrict the applicability of our findings to other settings such as outpatient care, nursing homes, or geographically diverse populations with different social compositions. Future studies should validate this approach in broader and more varied contexts to ensure external validity. While we incorporated fairness constraints, fairness itself is a multifaceted concept. Our assessment focused primarily on demographic parity and equal opportunity, omitting other definitions such as equalized odds, individual fairness, and utility-based fairness, which may yield different insights or trade-offs. Our modeling treated SBDH variables as static features derived from the pre-outcome period, despite the inherently dynamic nature of social conditions. Changes in housing, employment, and social support over time were not captured, potentially limiting predictive accuracy. Future work should explore temporal modeling or longitudinal tracking of SBDH to better reflect real-life dynamics and improve both risk prediction and the timing of preventive interventions. Together, these limitations highlight that while our SBDH-only model offers a compelling proof-of-concept, further validation, richer data sources, and integration into real-world clinical and public health frameworks are essential to fully realize its potential.

9 CONCLUSION

This study demonstrates that SBDH, even in the absence of clinical data, can meaningfully and fairly predict disease vulnerability, in this case, the development of PUs. By leveraging a

fairness-aware graph learning framework and combining raw SBDH features with debiased patient embeddings, we achieved predictive performance ($AUC \approx 0.72$) comparable to or better than traditional clinical tools like the Braden Scale. Critically, the model maintained fairness across subgroups and offered interpretable outputs, identifying housing instability, social isolation, and limited healthcare access as the most salient risk factors.

Our findings establish the feasibility of using SBDH as standalone inputs for health risk prediction, emphasizing that upstream social context can serve not only as background information but as a direct signal for clinical vulnerability. This approach is particularly valuable in settings where clinical data is limited, such as community care, public health outreach, or underserved populations, providing a blueprint for equitable, early, and actionable intervention. Importantly, the model's transparency via SHAP analysis supports real-world interpretability and trust, enabling its use in decision support systems that prioritize both performance and fairness.

While this study evaluated the predictive utility of SBDH, it does not imply causality between social factors and pressure ulcers; rather, it highlights associations useful for early risk stratification. Future work should explore causal inference, dynamic SBDH modeling, and external validation across diverse populations and settings. As healthcare systems increasingly recognize the importance of social determinants, models like ours pave the way for data-driven strategies that center equity, prevention, and upstream public health impact.

REFERENCES

- [1] Busari, Muhammed, and Tunde Bolanle. "Integrating Social Determinants into Predictive Models for US Public Health Forecasting." ResearchGate, 2025. <https://doi.org/10.13140/RG.2.2.23412.90244>
- [2] Adeniji, Olumide A., and Adeola Fagbemi. "Socio-Economic Inequalities in Stroke: A Global Perspective." IntechOpen, 2025. <https://doi.org/10.5772/intechopen.1125587>
- [3] Nicholson, J., J. L. Paul, G. Piché, and A. Wittkowski. "Parents with Mental and/or Substance Use Disorders and Their Children, Vol III." Frontiers in Psychiatry, 2025. <https://doi.org/10.3389/fpsy.2025.1630833>
- [4] Xie, Z., G. Zhong, C. Xu, et al. "Trends and Cross-Country Inequalities of Alcohol Use Disorders: Findings from the Global Burden of Disease Study 2021." Globalization and Health, 2025. <https://doi.org/10.1186/s12992-025-01124-5>
- [5] Zhou, Y., F. Chen, H. Li, and W. Wang. "Methodological Considerations in Assessing Modifiable Risk Factors Across Cancer Survivorship." Journal of Translational Medicine, 2025. <https://doi.org/10.1186/s12967-025-06580-6>
- [6] Maganga, J. J., A. Katende, E. Luoga, et al. "Determinants of Physical Activity and Diet Consumption Among Adolescents Living with HIV in Southern Tanzania." HIV/AIDS—Research and Palliative Care, 2025. <https://doi.org/10.2147/HIV.S519922>
- [7] Hamzens, M. F. "Culture and Non-Communicable Diseases: Determinants of Hypertension in Tenganan Pegringsingan Village, Karangasem, Bali." Society, 2025. <https://doi.org/10.25077/society.11.1.1-10.2025>
- [8] Deku, J. G., F. A. Ofosua, G. E. Kpene, et al. "Prevalence and Risk Factors of Intestinal Parasitic Infections in Pregnancy: A Cross-Sectional Study in a Rural District in Ghana." BMC Pregnancy and Childbirth, 2025. <https://doi.org/10.1186/s12884-025-07720-y>
- [9] Bergstrom, Nancy, Barbara J. Braden, Anne Laguzza, and Virginia Holman. The Braden

- Scale for Predicting Pressure Sore Risk. *Nursing Research* 36, no. 4 (1987): 205–210.
<https://doi.org/10.1097/00006199-198707000-00002>
- [10] Wright, Briana J., et al. "Health-related social needs and increased healthcare utilization among older adults." *Journal of General Internal Medicine* 38.4 (2023).
<https://doi.org/10.1007/s11606-022-07913-z>
- [11] Chen, X., Zhu, D., Li, C., Lin, Y., Lv, L., Ai, F., & Chen, Z. (2025). Modifiable factors affect cancer-specific survival: findings from a large population-based prospective cohort study. *Journal of Translational Medicine*, 23(1). <https://doi.org/10.1186/s12967-025-06372-y>
- [12] Park, C., Kim, N., Won, C. W., & Kim, M. (2025). Predicting cognitive frailty in community-dwelling older adults: a machine learning approach based on multidomain risk factors. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-00844-3>
- [13] Al-Zubaidy, Salih, et al. "The Impact of EHR-Incorporated Social Determinants on Risk Prediction Models." *Frontiers in Digital Health* 5 (2023).
<https://doi.org/10.3389/fdgth.2023.1135991>
- [14] Goldstein, Benjamin A., et al. "Opportunities and challenges in developing risk prediction models with electronic health records data." *American Journal of Epidemiology* 188.5 (2019): 870-875. <https://doi.org/10.1093/aje/kwy292>
- [15] Cho, InSook, et al. "Comparison of prediction models for hospital readmission using structured data from EHR." *International Journal of Medical Informatics* 135 (2020): 104033. <https://doi.org/10.1016/j.ijmedinf.2019.104033>
- [16] Obermeyer, Ziad, et al. "Dissecting racial bias in an algorithm used to manage the health of populations." *Science* 366.6464 (2019): 447–453.
<https://doi.org/10.1126/science.aax2342>
- [17] Mehrabi, Ninareh, et al. "A survey on bias and fairness in machine learning." *ACM Computing Surveys* 54.6 (2021): 1–35. <https://doi.org/10.1145/3457607>
- [18] Shang, Jun, et al. "GAMENet: Graph augmented memory networks for recommending medication combination." *AAAI* (2019). <https://doi.org/10.1609/aaai.v33i01.33012318>
- [19] Peng, Yuxin, et al. "Fairness-Aware Graph Neural Networks: A Survey." *ACM Computing Surveys* (2023). <https://doi.org/10.1145/3589467>
- [20] Lyder, Courtney H., et al. "Hospital-acquired pressure ulcers: Results from the national Medicare Patient Safety Monitoring System study." *Journal of the American Geriatrics Society* 60.9 (2012): 1603–1608. <https://doi.org/10.1111/j.1532-5415.2012.04106.x>
- [21] Cox, Jill. "Predictive power of the Braden Scale for pressure sore risk in adults: A meta-analysis." *Decubitus* 6.1 (1993): 24–30.
<https://doi.org/10.1097/01.NURSE.0000541266.72990.4e>
- [22] Sharp, Claire A., et al. "Social disparities in the burden of pressure ulcers: A systematic review." *Journal of Tissue Viability* 29.1 (2020): 10–20.
<https://doi.org/10.1016/j.jtv.2019.10.001>
- [23] Johnson, A., Bulgarelli, L., Pollard, T., Gow, B., Moody, B., Horng, S., Celi, L. A., & Mark, R. (2024). MIMIC-IV (version 3.1). PhysioNet. <https://doi.org/10.13026/kpb9-mt58>.
- [24] 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>
- [25] 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.
- [26] Ahsan H, Ohnuki E, Mitra A, Yu H. MIMIC-SBDH: A Dataset for Social and Behavioral Determinants of Health. *Proc Mach Learn Res*. 2021 Aug;149:391-413. PMID: 35005628; PMCID: PMC8734043.
- [27] Gunningberg, Lena, et al. "Incidence and risk factors for pressure ulcers in ICU patients: A systematic review." *Journal of Wound Care* 29.3 (2020): 144–150. <https://doi.org/10.12968/jowc.2020.29.3.144>
- [28] Weissman, Joel S., et al. "Social Determinants of Health in the ICU: An Underappreciated Risk Layer." *Critical Care Medicine* 48.5 (2020): 690–696. <https://doi.org/10.1097/CCM.0000000000004302>
- [29] Agniel, Denis, et al. "Biases in electronic health record data due to processes within the healthcare system: retrospective observational study." *BMJ* 361 (2018): k1479. <https://doi.org/10.1136/bmj.k1479>
- [30] Saxena, R. R. "Applications of Natural Language Processing in the Domain of Mental Health." *Authorea Preprints*, 2024. <https://doi.org/10.36227/techrxiv.173014748.80471770>
- [31] Bonner, Sebastian, et al. "Graph representation of patient data: a systematic review." *Briefings in Bioinformatics* 22.1 (2021): 393–411. <https://doi.org/10.1093/bib/bbz164>
- [32] Choi, Edward, et al. "Doctor AI: Predicting Clinical Events via Recurrent Neural Networks." *Machine Learning for Healthcare Conference* (2016): 301–318. <https://doi.org/10.48550/arXiv.1511.05942>
- [33] Bazemore, Andrew W., et al. "Advancing population health improvement through community data partnerships." *The Journal of the American Board of Family Medicine* 29.6 (2016): 776–781. <https://doi.org/10.3122/jabfm.2016.06.160086>
- [34] Himmelstein, Daniel S., et al. "Systematic integration of biomedical knowledge prioritizes drugs for repurposing." *eLife* 6 (2017): e26726. <https://doi.org/10.7554/eLife.26726>
- [35] Kipf, Thomas N., and Max Welling. "Semi-supervised classification with graph convolutional networks." *International Conference on Learning Representations* (2017). <https://doi.org/10.48550/arXiv.1609.02907>
- [36] Huang, Xin, et al. "Graph representation learning for healthcare: a survey." *Briefings in Bioinformatics* 23.2 (2022): bbab441. <https://doi.org/10.1093/bib/bbab441>
- [37] Lundberg, Scott M., and Su-In Lee. "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems* 30 (2017): 4765–4774. <https://doi.org/10.48550/arXiv.1705.07874>
- [38] Zhang, Yifan, et al. "Explaining Machine Learning Predictions in Healthcare with Tree SHAP: Applications and Optimization Techniques." *Journal of Biomedical Informatics* 144 (2024): 104461. <https://doi.org/10.1016/j.jbi.2024.104461>