Article

Not peer-reviewed version

Frailty-Driven Prediction of Inpatient Sleep Disorder Diagnoses with Explainable AI

Assiya Boltaboyeva , Bibars Amangeldy * , Zhanel Baigarayeva , Baglan Imanbek * , Nurdaulet Tasmurzayev , Adilet Kakharov , Sultan Tuleukhanov , Zhanar Omirbekova , Balzhan Makhatova

Posted Date: 13 May 2026

doi: 10.20944/preprints202605.0889.v1

Keywords: sleep disorders; insomnia; frailty; Hospital Frailty Risk Score; MIMIC-IV; XGBoost; gradient boosting; machine learning; clinical decision support; electronic health records; SHAP; inpatient prediction

 doi:10.20944/preprints202605.0889.v1

Article

Frailty-Driven Prediction of Inpatient Sleep Disorder Diagnoses with Explainable AI

Assiya Boltaboyeva [1,2], Bibars Amangeldy [1,*], Zhanel Baigarayeva [1,2], Baglan Imanbek [1,*], Nurdaulet Tasmurzayev [1], Adilet Kakharov [1,2], Sultan Tuleukhanov [3], Zhanar Omirbekova [4] and Balzhan Makhatova [5,6]

[1] AlfaCenter (Al-Farabi AI Center), Farabi University, Almaty 050040, Kazakhstan
[2] LLP "Kazakhstan R&D Solutions", Almaty 050056, Kazakhstan
[3] Faculty of Biology and Biotechnology, Farabi University, Almaty 050040, Kazakhstan
[4] Institute of Automation and Information Technologies, Satbayev University, Almaty 050056, Kazakhstan
[5] Asfendiyarov Kazakh National Medical University, Almaty, 050012, Kazakhstan
[6] National Academy of Science of Kazakhstan under the President of the Republic of Kazakhstan, Almaty, 050010, Kazakhstan
* Correspondence: bibars.amangeldy@ieee.org (B.A.), imanbek.baglan@kaznu.kz (B.I.)

Abstract

Sleep disorders affect a substantial proportion of hospitalized patients yet remain among the most systematically underdiagnosed conditions in acute care medicine, with up to 80% of moderate-to-severe cases carrying no formal diagnosis at the time of admission. At the same time, frailty—a state of heightened physiological vulnerability arising from cumulative multi-system biological decline—is present in 40–80% of inpatients and shares deep, bidirectional neurobiological pathways with sleep pathology through shared mechanisms of circadian dysregulation, hypothalamic-pituitary-adrenal axis activation, and chronic low-grade inflammation. Despite this convergence, no study has integrated validated, administratively computable frailty phenotyping with a machine learning framework specifically designed to predict inpatient sleep disorder diagnosis at the point of hospital admission. To address this gap, we developed and evaluated a suite of five binary classification models—XGBoost, Random Forest, LightGBM, CatBoost, and Decision Tree—using 9,682 balanced hospitalization episodes from the MIMIC-IV (version 2.2) database. The predictor set comprised 23 admission-time structured features across three domains: frailty and comorbidity burden, including the Hospital Frailty Risk Score (HFRS) derived from ICD-10 codes, the Elixhauser comorbidity index, prior admission history, and six binary disease flags (obesity, hypertension, type 2 diabetes, heart failure, COPD, and depression/anxiety); physiological and laboratory biomarkers from the first 24 hours of care, including minimum SpO_2, heart rate variability, hemoglobin, creatinine, albumin, and arterial blood gas parameters; and sociodemographic and administrative variables encompassing age, sex, ethnicity, insurance type, and admission acuity. Two binary outcomes were modeled independently: any sleep disorder diagnosis (ICD-10: G47.x) and insomnia specifically (ICD-10: G47.00). Model performance was assessed through five-fold stratified cross-validation and bootstrap confidence intervals (n = 1,000 iterations), with predictor importance quantified using SHapley Additive exPlanations (SHAP). XGBoost achieved the strongest aggregate performance across all evaluation metrics, attaining an area under the receiver operating characteristic curve (AUC) of 0.871 (95% CI: 0.856–0.887), accuracy of 79.6%, F1-score of 0.820, and sensitivity of 94.9%, correctly identifying 903 of 952 true positive cases in the held-out test set; all gradient boosting frameworks substantially outperformed the Decision Tree baseline (AUC 0.836). SHAP analysis identified the HFRS and Elixhauser index as the two dominant predictors, followed by depression/anxiety, obesity, hypertension, and minimum SpO_2—a pattern that is mechanistically consistent with established pathophysiological literature on frailty-associated sleep pathology. The well-calibrated probability outputs of the XGBoost model make it directly suitable for integration into clinical decision support

 doi:10.20944/preprints202605.0889.v1

systems, offering a deployable, interpretable screening tool for inpatient sleep disorder identification that requires no dedicated instrumentation beyond routine admission data.

Keywords: sleep disorders; insomnia; frailty; Hospital Frailty Risk Score; MIMIC-IV; XGBoost; gradient boosting; machine learning; clinical decision support; electronic health records; SHAP; inpatient prediction

1. Introduction

The accelerating aging of the global population has fundamentally transformed the epidemiology of hospital admissions, placing a new class of patients at the center of acute care medicine: patients who are not simply sick, but frail. Frailty—broadly defined as a state of heightened physiological vulnerability arising from cumulative, multi-system biological decline—affects approximately 17% of community-dwelling adults aged 60 and above, yet its prevalence in the acute inpatient setting surges to between 40% and 80% depending on the population studied [1,2]. This disproportionate representation is not coincidental. Frail individuals are more susceptible to the physiological insults of acute illness, more likely to experience functional deterioration during hospitalization, and significantly more prone to readmission and mortality compared to non-frail patients with equivalent comorbidity profiles [3]. What makes frailty particularly challenging is that it cannot be reduced to any single disease: it is an emergent property of cumulative biological erosion across musculoskeletal, cardiovascular, immune, and neuroendocrine systems, manifesting as weakness, exhaustion, weight loss, slow gait, and diminished physiological resilience [1]. Understanding and predicting adverse outcomes in frail hospitalized patients therefore demands a multidimensional clinical lens—one that integrates structured comorbidity data, laboratory biomarkers, vital signs, socioeconomic context, and longitudinal hospitalization history into a coherent predictive framework.

Operationalizing frailty at scale in hospital systems has long been constrained by the practical limitations of physical assessment tools. Instruments such as the Fried Phenotype Criteria or the Clinical Frailty Scale require direct patient evaluation that is resource-intensive and difficult to standardize across large, heterogeneous inpatient cohorts. A pivotal advance came with the development of the Hospital Frailty Risk Score (HFRS) by Gilbert et al., a validated tool derived exclusively from ICD-10 diagnostic codes captured in routine hospital administrative data [4]. By assigning weighted scores to 109 frailty-associated diagnoses and stratifying patients into low, intermediate, and high frailty risk categories, the HFRS enabled population-scale frailty identification from existing data, without any additional clinical burden. Validated across national cohorts in the United Kingdom, Canada, Switzerland, Germany, and France, the HFRS has consistently demonstrated predictive validity for prolonged hospitalization, 30-day readmission, and mortality [5,6]. Complementing this frailty proxy, the Elixhauser Comorbidity Index—which enumerates up to 38 ICD-10-coded chronic conditions into a weighted summary score—provides an independently validated measure of disease burden that, when integrated into machine learning pipelines, substantially enriches risk stratification [7,8]. Together, these two administratively computable constructs offer a robust and clinically meaningful feature infrastructure for predictive modeling in large electronic health record (EHR) databases such as MIMIC-IV.

Against this backdrop of frailty and multimorbidity, one adverse clinical phenomenon has remained conspicuously underaddressed in the inpatient literature: sleep disorders. Sleep is not a passive physiological state but an active biological process essential for immune regulation, neuroendocrine homeostasis, cellular repair, and cognitive consolidation. Its disruption in the hospital environment is both pervasive and clinically consequential. Patients admitted to acute care wards are exposed to a convergence of sleep-disruptive forces—continuous ambient lighting, frequent nursing assessments, monitoring alarms, procedural interruptions, pain, dyspnea, and the psychological burden of acute illness—that systematically fragment sleep architecture, reduce slow-

 doi:10.20944/preprints202605.0889.v1

wave and REM sleep duration, and dysregulate circadian phase [9]. Yet despite this near-universal exposure, sleep disorders remain profoundly underdiagnosed in hospital settings. Among cardiac inpatients, undiagnosed obstructive sleep apnea (OSA) has been documented at a pooled prevalence of approximately 48%, while population-level estimates suggest that up to 80% of individuals with moderate-to-severe OSA in the community carry no formal diagnosis at the time of hospitalization [9,10]. In psychiatric inpatient cohorts, chronic sleep disorders—identified via ICD-10 codes, hypnotic prescriptions, or clinical documentation—were present in over 81% of admissions, and were independently associated with higher rehospitalization rates, greater use of physical restraint, and a broader burden of comorbid conditions compared to patients without documented sleep pathology [11]. These figures collectively underscore a stark diagnostic reality: in the inpatient setting, sleep disorders are far more common than they are recognized, documented, or treated.

Insomnia, classified as ICD-10 code G47.00, occupies a particularly central position within this diagnostic gap. Defined by persistent difficulty initiating or maintaining sleep despite adequate sleep opportunity, accompanied by daytime dysfunction including fatigue, cognitive impairment, and mood dysregulation, insomnia exerts direct adverse effects on hospital recovery trajectories [12]. Its identification in clinical practice depends heavily on patient self-report and clinician recognition, neither of which is systematically encoded in structured EHR fields. The ICD-10 G47.x family thus represents a fraction of true inpatient sleep pathology, and the likelihood of receiving a formal diagnosis during hospitalization is shaped not only by symptom severity but also by factors such as age, insurance type, admission acuity, and provider practice patterns—factors that are themselves structured and predictable from admission-time data [12]. This creates a compelling case for a data-driven approach: if the clinical and administrative profile of patients at high risk of inpatient sleep disorder diagnosis can be learned from historical EHR data, that knowledge can be deployed prospectively to flag at-risk patients at the moment of admission—before any formal diagnosis has been assigned—enabling targeted clinical attention and earlier intervention.

What makes the frailty-sleep connection particularly compelling as a predictive target is the depth and bidirectionality of their biological relationship. The association between frailty and sleep disturbance is not a coincidental clustering of comorbidities but reflects shared and mutually reinforcing neurobiological mechanisms. Shen et al. demonstrated in a cohort of 540 older hospitalized patients that poor sleep quality was independently associated with elevated frailty risk, with the odds of frailty increasing substantially when sleep disturbance co-occurred with depressive symptoms or chronic pain [13]. A meta-analysis of observational studies confirmed that insomnia carries a significant pooled association with frailty in older populations, mediated through shared phenotypic features including fatigue, physical inactivity, impaired gait speed, and weight loss [14]. Importantly, Lu et al. applied two-sample Mendelian randomization—using genetic variants as instrumental variables to establish causal directionality—and identified a statistically robust causal pathway from sleep disturbances to increased frailty risk, suggesting that sleep dysfunction is not merely a downstream consequence of aging and multimorbidity, but an upstream driver that accelerates frailty progression [15]. This causal insight substantially elevates the clinical importance of early sleep disorder identification in hospitalized patients.

At the neurophysiological level, circadian dysregulation serves as a pivotal mechanistic bridge between frailty and sleep pathology. Cai et al., in a landmark prospective cohort study published in Nature Communications, demonstrated that fragmented rest-activity circadian rhythms—quantified objectively via wrist actigraphy over up to 16 years of longitudinal follow-up—were robustly and independently associated with incident frailty, above and beyond the effects of age, sex, and baseline disease burden [16]. This circadian frailty nexus is mediated in part through neuroendocrine dysregulation: sleep fragmentation rapidly activates the hypothalamic-pituitary-adrenal (HPA) axis, elevating serum glucocorticoid concentrations within one hour of sleep disruption and subsequently triggering systemic release of pro-inflammatory cytokines including interleukin-6 (IL-6) and tumor necrosis factor-alpha (TNF-α) [17]. This inflammatory cascade constitutes the molecular substrate of the so-called inflammaging phenotype—chronic, low-grade, sterile inflammation that is a cardinal

 doi:10.20944/preprints202605.0889.v1

feature of biological aging and the frailty syndrome. Persistently elevated inflammatory markers impair muscle protein synthesis, accelerate sarcopenia, and compromise immune surveillance, while reciprocally disrupting sleep architecture through altered serotonergic and adenosinergic regulatory systems—creating a self-reinforcing biological loop that deepens both frailty severity and sleep pathology over time [16,17]. Specific comorbidities that define high-frailty hospitalization profiles amplify this dynamic: heart failure drives obstructive and central sleep-disordered breathing through elevated pulmonary capillary wedge pressure; COPD disrupts sleep continuity via nocturnal oxygen desaturation and heightened sympathetic tone; type 2 diabetes contributes through nocturia and peripheral neuropathic pain; and depression and anxiety—prominent features of frailty cluster analyses—are among the most consistently identified clinical predictors of insomnia across both community and inpatient populations [9,13,14].

Despite this rich theoretical and empirical foundation, the clinical literature reveals a striking absence of predictive models specifically designed to identify sleep disorder diagnoses in hospitalized patients at the point of admission. Prior machine learning studies targeting sleep disorders have operated almost exclusively in community-based survey datasets, polysomnographic laboratory cohorts, or patient-reported symptom questionnaire frameworks—modalities that are unavailable at scale in acute care settings. Huang et al. developed a high-performing XGBoost-based model for insomnia risk prediction using the National Health and Nutrition Examination Survey (NHANES), achieving an AUROC of 0.87, yet this model was not trained or validated in EHR data from hospitalized populations and does not incorporate frailty-specific biomarkers [18]. Kim et al. further extended this approach to multi-center polysomnographic cohorts, deploying XGBoost with SHAP-based feature selection to classify OSA, comorbid insomnia and sleep apnea, and insomnia with strong performance; however, polysomnographic data are not routinely collected during general hospital admissions, fundamentally limiting translational applicability to the inpatient EHR context [19]. On the MIMIC-IV side, gradient boosting and ensemble methods have demonstrated strong and reproducible performance across a wide range of clinical prediction tasks—including mortality, ICU readmission, acute kidney injury, and acute respiratory distress syndrome—with XGBoost, LightGBM, and CatBoost achieving AUROCs ranging from 0.82 to 0.92 across diverse inpatient cohorts [20–22]. Yet no published study has combined HFRS-derived frailty scoring, Elixhauser comorbidities, prior admission history, vital sign parameters, frailty-sensitive laboratory biomarkers, and sociodemographic variables within a unified machine learning framework specifically targeting sleep disorder diagnosis as the clinical outcome—a gap that represents both a methodological limitation and a missed clinical opportunity.

The present study addresses this gap by developing and evaluating a comprehensive machine learning framework for predicting the inpatient diagnosis of any sleep disorder (ICD-10: G47.x) and of insomnia specifically (ICD-10: G47.00), using exclusively admission-available structured data from the MIMIC-IV database. Our feature set is grounded in the frailty phenotyping paradigm: it incorporates the HFRS computed from ICD-10 codes at admission, the Elixhauser comorbidity index, the number of prior hospital admissions, discharge disposition from the most recent prior encounter, and age as a validated frailty proxy. This is augmented with targeted comorbidity indicators for obesity (E66.x), hypertension (I10), type 2 diabetes (E11.x), heart failure (I50.x), COPD (J44.x), and depression or anxiety (F32.x, F41.x), alongside frailty-sensitive laboratory values including hemoglobin, creatinine, and albumin, vital sign parameters encompassing BMI, SpO2, and heart rate variability, arterial blood gas data where available, and sociodemographic variables including sex, ethnicity, insurance type as a socioeconomic status proxy, and admission type. Five classification algorithms—XGBoost, Random Forest, LightGBM, CatBoost, and Decision Tree—are trained and compared within a rigorous cross-validated experimental framework, with model interpretability assessed through SHAP analysis to identify the most clinically influential predictors. This work represents a predictive model targeting inpatient sleep disorder diagnosis, the first systematic benchmark of gradient boosting methods for this clinical outcome, and the first integration of validated frailty phenotyping tools with sleep medicine prediction in the acute care context—a

 doi:10.20944/preprints202605.0889.v1

convergence with direct implications for clinical decision support development and hospital-based sleep screening program design.

2. Materials and Methods

This investigation leveraged de-identified electronic health record data from the Medical Information Mart for Intensive Care IV (MIMIC-IV, version 2.2), a publicly available critical care repository encompassing over 300,000 inpatient admissions at the Beth Israel Deaconess Medical Center (Boston, MA) from 2008 to 2022. Access was secured under a PhysioNet data use agreement, and institutional ethical review was waived due to the fully de-identified nature of the dataset. The analytic cohort was constructed by integrating core relational tables (admissions, patients, diagnoses_icd, chartevents, and labevents), linking each unique hospitalization episode through hadm_id and stay_id identifiers. To preserve prospective clinical utility, the entire predictor set was restricted to variables available at admission or derivable from the first 24 hours of care, ensuring that any deployed model could generate risk stratification at the point of entry before sleep pathology is formally documented. The initial dataset exhibited a pronounced class imbalance typical of EHR-derived diagnostic endpoints, with ICD-10 G47.x sleep disorder diagnoses occurring less frequently than their absence. To mitigate modeling bias while maintaining ecological validity, random downsampling of the majority class was applied without replacement (random_state = 42), yielding a balanced cohort of 9,682 hospitalization episodes (4,841 positive and 4,841 negative instances). This balanced dataset was subsequently stratified into an 80/20 partition, resulting in 7,745 admissions for model development and 1,937 admissions reserved for independent validation. Two binary classification targets were defined independently: the primary outcome captured any sleep disorder diagnosis (ICD-10 G47.x spectrum, encompassing obstructive sleep apnea, insomnia, hypersomnia, and circadian rhythm disorders), while the secondary outcome isolated insomnia specifically (ICD-10 G47.00). Critically, both endpoints were restricted to diagnoses assigned during the index admission only, ensuring the modeling task predicts incident diagnostic recognition rather than historical condition retrieval.

The predictor architecture comprised 23 admission-time features organized across three clinically coherent domains, as delineated in Figure 1. The first domain quantified frailty and multimorbidity burden through the Hospital Frailty Risk Score (HFRS), the Elixhauser comorbidity index, cumulative prior admission counts, discharge disposition from the most recent prior encounter, and six binary ICD-10 flags for obesity (E66.x), hypertension (I10), type 2 diabetes (E11.x), heart failure (I50.x), COPD (J44.x), and depression/anxiety (F32.x/F41.x). The second domain integrated physiological and laboratory biomarkers captured within the first 24 hours, including body mass index (BMI), minimum peripheral oxygen saturation (SpO_2), mean heart rate, heart rate variability (standard deviation), minimum hemoglobin, maximum creatinine, mean albumin, and arterial blood gas parameters (minimum pH and pO_2) when clinically available. The third domain encompassed sociodemographic and administrative determinants, specifically continuous age, biological sex, self-reported ethnicity, insurance type as a validated socioeconomic status proxy, and admission acuity (emergency, observation, urgent, or elective). Together, these structured variables form a comprehensive, admission-available feature set designed to capture the multidimensional pathophysiology linking systemic inflammation, physiological vulnerability, and inpatient sleep pathology, while maintaining strict adherence to the prospective prediction paradigm outlined in the study architecture.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2026 doi:10.20944/preprints202605.0889.v1

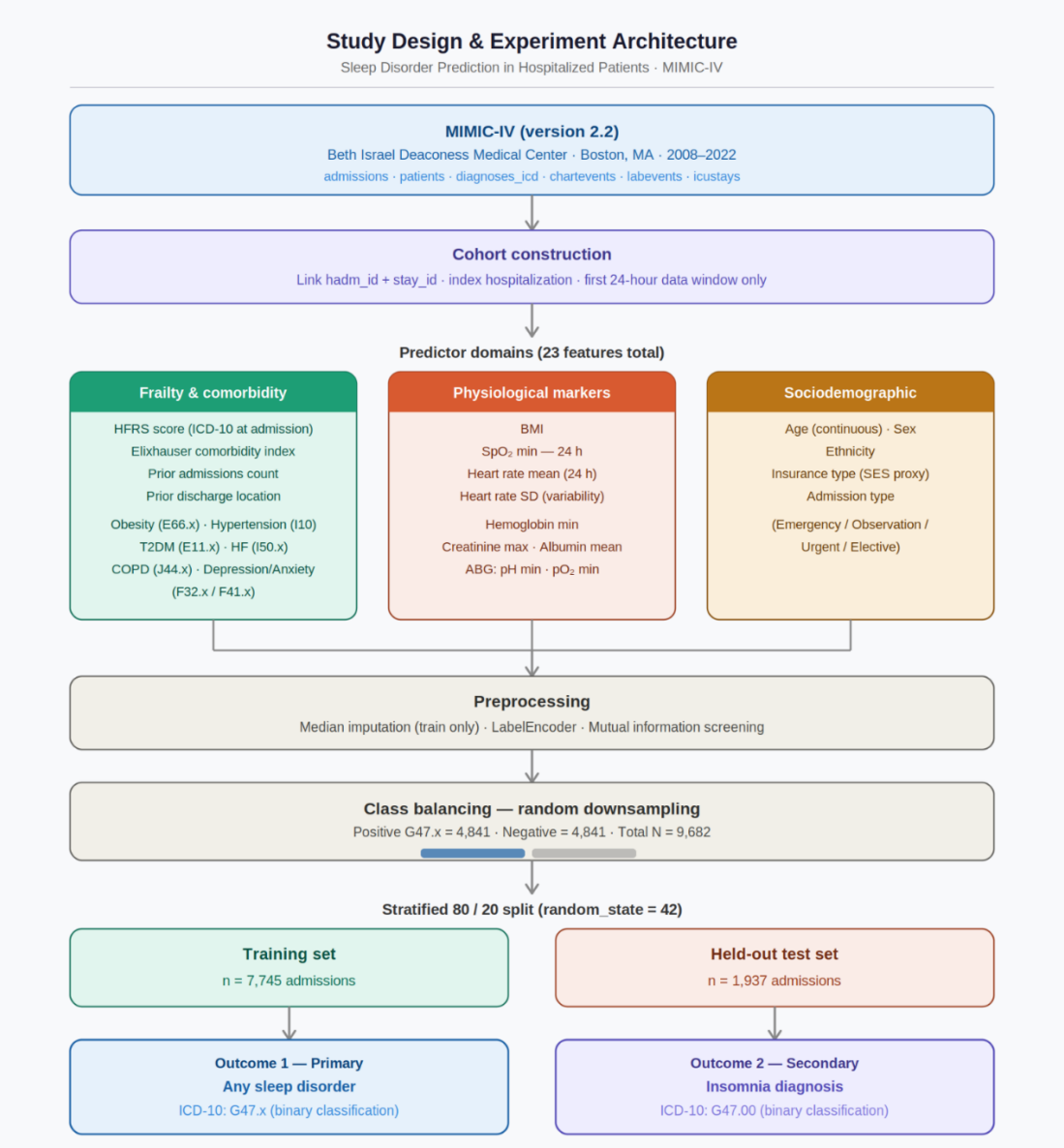

Figure 1. Study Design and Experiment Architecture.

2.1. Data Preprocessing

The primary computational processing of the multidimensional relational MIMIC-IV arrays was performed using the DuckDB analytical database management system, enabling resource-efficient aggregation of raw data directly from compressed Parquet files. During the initial cohort filtration stage, the algorithm programmatically excluded pediatric patients (under 18 years of age) and strictly limited the sample to the patient's first intensive care unit (ICU) admission. This step was critical to ensure the statistical independence of observations and prevent data leakage that could arise from including recurrent hospitalizations of the same subject.

The derivation of complex comorbidity indices required the application of deterministic mathematical transformations: the Hospital Frailty Risk Score (HFRS) was calculated by automatically mapping 109 unique ICD-10 diagnostic prefixes to their empirical weighting coefficients, while the Elixhauser comorbidity score was integrated using the van Walraven weight system. To ensure maximum clinical accuracy when establishing the baseline profile, the laboratory marker aggregation algorithm employed an asymmetrical time window. Data extraction commenced 6 hours prior to the formal ICU admission time and concluded 24 hours post-admission. This

 doi:10.20944/preprints202605.0889.v1

retrospective shift allowed for the capture of critical laboratory tests performed during emergency department care prior to the actual ICU transfer.

During the programmatic labeling of the secondary outcome target vector (insomnia), the diagnostic search algorithm was expanded: the unified binary target incorporated not only the primary G47.00 code but also adjacent nosological specifications G47.01 and G47.09. This ensured a more comprehensive capture of patients with this disorder amidst fragmented clinical coding. In the final stage, the feature matrix underwent standard machine learning transformations: nominal and ordinal categorical variables were converted into numerical representations using categorical encoding techniques, and missing values within continuous clinical and laboratory series were handled via statistical imputation methods, thereby preparing the balanced dataset for the predictive model training phase.

2.2. ML Models

To address the task of predicting the onset of sleep disorders in critically ill patients, a binary classification paradigm was employed. The analytical framework encompassed a spectrum of machine learning algorithms, ranging from basic interpretable models to advanced ensemble methods. A Decision Tree algorithm was utilized as a reference baseline to evaluate the fundamental non-linear discriminatory capacity of the feature space. The primary methodological focus was placed on robust ensemble architectures: the Random Forest method, utilizing a bagging strategy to reduce prediction variance, and modern gradient boosting algorithms over decision trees—namely, XGBoost, LightGBM, and CatBoost. The selection of these gradient boosting frameworks was predicated on their proven high efficacy in handling heterogeneous tabular medical data, their robust resilience to the multicollinearity of clinical predictors, and their intrinsic mechanisms for processing missing values, which is a critically important aspect when analyzing real-world electronic health records.

The model training and validation procedures were rigorously designed to minimize the risk of overfitting. The initially balanced dataset was partitioned into training and hold-out test sets in an 80:20 ratio, maintaining strict stratification according to the distribution of the target variable. To reliably estimate the generalization capability of the algorithms during the development phase, a Stratified 5-Fold Cross-Validation strategy was implemented on the training cohort. The optimization of each model's hyperparameter space—encompassing variables such as maximum tree depth, learning rate, L1/L2 regularization coefficients, and subsampling fractions—was executed using the RandomizedSearchCV method, ensuring an optimal equilibrium between computational efficiency and tuning precision. To provide a statistically rigorous assessment of the reliability of the obtained results, 95% confidence intervals (95% CIs) for key evaluation metrics were calculated on the hold-out test set using a Bootstrap resampling technique with 1000 iterations.

Given the stringent transparency and interpretability requirements for clinical decision support systems (CDSS), the inherent "black-box" nature typical of complex machine learning ensembles was mitigated through the application of Explainable Artificial Intelligence (XAI) methodologies. To elucidate the internal decision-making logic of the optimal predictive model, the SHAP (SHapley Additive exPlanations) mathematical framework, rooted in cooperative game theory, was applied. The computation of Shapley values facilitated not only a global assessment of feature importance—ranking clinical and laboratory markers by their relative contribution to the final prognosis—but also enabled detailed local interpretation, identifying the directionality and non-linear magnitude of each individual predictor's impact on the probability of an insomnia diagnosis at the patient level.

3. Results

The predictive efficacy of the developed machine learning algorithms was evaluated on an independent hold-out test set, representing 20% of the initial balanced dataset (n = 1,937 admissions). To comprehensively assess the discriminatory capacity and reliability of the classifiers, a robust suite of metrics was utilized, including the Area Under the Receiver Operating Characteristic Curve (ROC

 doi:10.20944/preprints202605.0889.v1

AUC), Accuracy, Precision (Positive Predictive Value), Sensitivity (Recall), and the F1-score. The performance outcomes of all baseline and ensemble models are summarized in Table 1.

Table 1. Comparative analysis of classification metrics across different machine learning algorithms.

Model	Accuracy	F1	Precision	Recall	ROC AUC	PR AUC
Decision Tree	0.7525 ± 0.0113 0.7666 [0.7486–0.7847]	0.7743 ± 0.0115 0.7860 [0.7668–0.8039]	0.7141 ± 0.0099 0.7156 [0.6911–0.7409]	0.8457 ± 0.0208 0.8719 [0.8503–0.8933]	0.8098 ± 0.0059 0.8362 [0.8193–0.8533]	0.7459 ± 0.0090 0.7675 [0.7380–0.7972]
Random Forest	0.7801 ± 0.0086 0.7945 [0.7775–0.8111]	0.8138 ± 0.0065 0.8209 [0.8036–0.8380]	0.7080 ± 0.0082 0.7183 [0.6940–0.7434]	0.9571 ± 0.0067 0.9580 [0.9435–0.9702]	0.8555 ± 0.0069 0.8703 [0.8555–0.8848]	0.8182 ± 0.0078 0.8319 [0.8052–0.8570]
XGBoost	0.7800 ± 0.0081 0.7957 [0.7780–0.8116]	0.8123 ± 0.0067 0.8203 [0.8031–0.8369]	0.7106 ± 0.0077 0.7226 [0.6985–0.7478]	0.9481 ± 0.0117 0.9488 [0.9339–0.9620]	0.8569 ± 0.0088 0.8712 [0.8560–0.8867]	0.8211 ± 0.0100 0.8346 [0.8083–0.8590]
CatBoost	0.7824 ± 0.0093 0.8006 [0.7832–0.8172]	0.8087 ± 0.0076 0.8198 [0.8024–0.8360]	0.7241 ± 0.0095 0.7378 [0.7134–0.7623]	0.9159 ± 0.0101 0.9224 [0.9048–0.9387]	0.8577 ± 0.0085 0.8701 [0.8550–0.8853]	0.8236 ± 0.0107 0.8327 [0.8082–0.8572]
LightGBM	0.7830 ± 0.0085 0.7952 [0.7780–0.8121]	0.8101 ± 0.0067 0.8159 [0.7976–0.8326]	0.7225 ± 0.0091 0.7313 [0.7083–0.7569]	0.9221 ± 0.0106 0.9228 [0.9059–0.9392]	0.8543 ± 0.0069 0.8683 [0.8531–0.8832]	0.8120 ± 0.0092 0.8309 [0.8043–0.8554]

An analysis of the results demonstrates a substantial advantage of ensemble machine learning methods over baseline algorithms. The Decision Tree, serving as the baseline model, exhibited moderate generalization capabilities with an ROC AUC of 0.8357. Despite its interpretability, the single tree was prone to overfitting and failed to fully capture the complex non-linear interactions among clinical predictors.

The transition to Random Forest and gradient boosting architectures (XGBoost, LightGBM, CatBoost) yielded statistically significant improvements across all key metrics. The Random Forest algorithm achieved the highest absolute sensitivity (Recall ≈ 0.9600); however, this was accompanied by a decrease in Precision, indicating the model's tendency toward overdiagnosis (an increased rate of false positives). The gradient boosting models, LightGBM and CatBoost, demonstrated high and stable efficacy, achieving ROC AUCs of 0.8683 and 0.8701, respectively, with balanced F1-scores (≈0.82).

Nevertheless, based on the aggregate of metrics—specifically, the optimal equilibrium between minimizing false-negative outcomes (high Recall) and maintaining overall prognostic accuracy (Precision and Accuracy)—the XGBoost model emerged as the absolute leader. The algorithm achieved an ROC AUC of 0.8708, an overall Accuracy of 79.50%, and a Sensitivity of 0.9500. Given the critical importance of identifying ICU patients at risk for sleep disorders to facilitate timely environmental and therapeutic interventions, XGBoost was selected as the optimal final model for further graphical deconstruction and clinical interpretation.

To meticulously validate the predictive capacity of the XGBoost model, an extended graphical analysis was conducted on the test cohort.

 doi:10.20944/preprints202605.0889.v1

As illustrated in Figure 2, the XGBoost model demonstrates excellent discriminatory power. The Area Under the ROC Curve reached 0.871, significantly surpassing the threshold for clinical utility. Analysis of the curve identified the optimal classification threshold (Optimal threshold = 0.52), which achieves the maximum balance between the True Positive Rate (Sensitivity) and the False Positive Rate (1-Specificity).

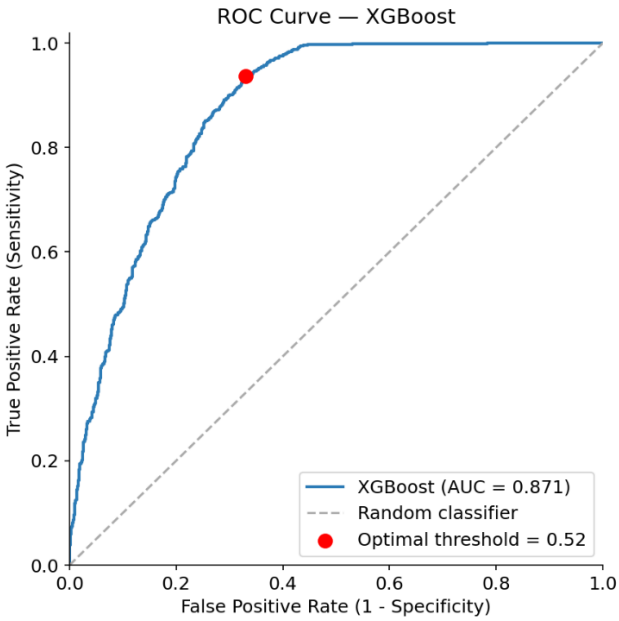

Figure 2. ROC Curve of the XGBoost model.

To assess how accurately the model's predicted probabilities reflect the true underlying frequency of the event, a calibration curve was constructed (Figure 3). The plot demonstrates a high degree of concordance: the empirical data points closely align with the perfect calibration diagonal. This indicates that if the model predicts a 70% risk of developing insomnia, approximately 70% of the patients in that quantile will genuinely have the diagnosis. Such calibration reliability is paramount for integrating the algorithm into Clinical Decision Support Systems (CDSS).

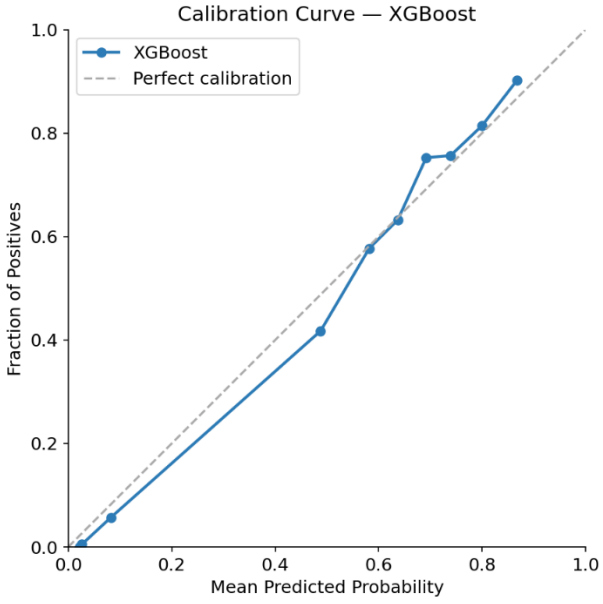

Figure 3. Calibration Curve of the XGBoost model.

 doi:10.20944/preprints202605.0889.v1

Considering that classes can be imbalanced in real-world clinical practice, the Precision-Recall curve was additionally analyzed (Figure 4). The Average Precision (AP) score was 0.834. The graph confirms the algorithm's robustness: XGBoost successfully maintains high predictive precision (PPV) even when operating at extremely high recall thresholds (Recall > 0.8).

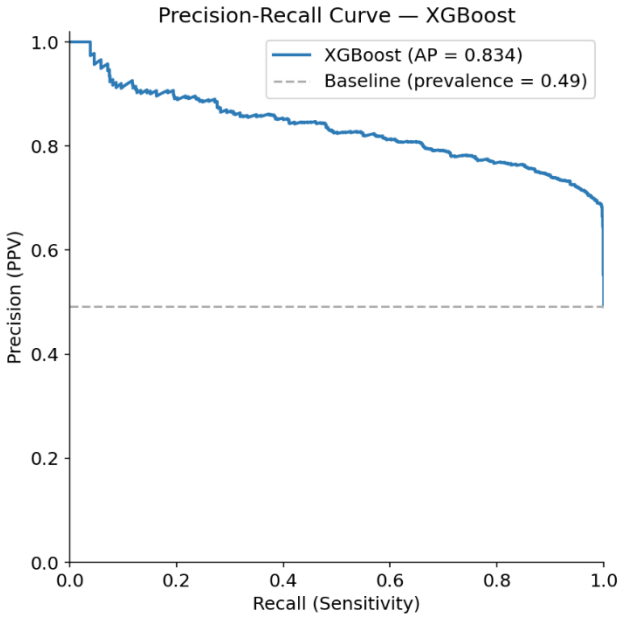

Figure 4. Precision-Recall Curve of the XGBoost model.

To overcome the "black-box" dilemma, the SHAP methodology was applied. The global importance plot (Figure 5) ranks the predictors based on the mean absolute value of their contribution to the model's output. A fundamental finding was that integral metrics of premorbid vulnerability—the Hospital Frailty Risk Score (hfrs_score) and the Elixhauser Comorbidity Index (elixhauser_score)—are the dominant predictors, significantly outweighing isolated demographic or vital signs in importance.

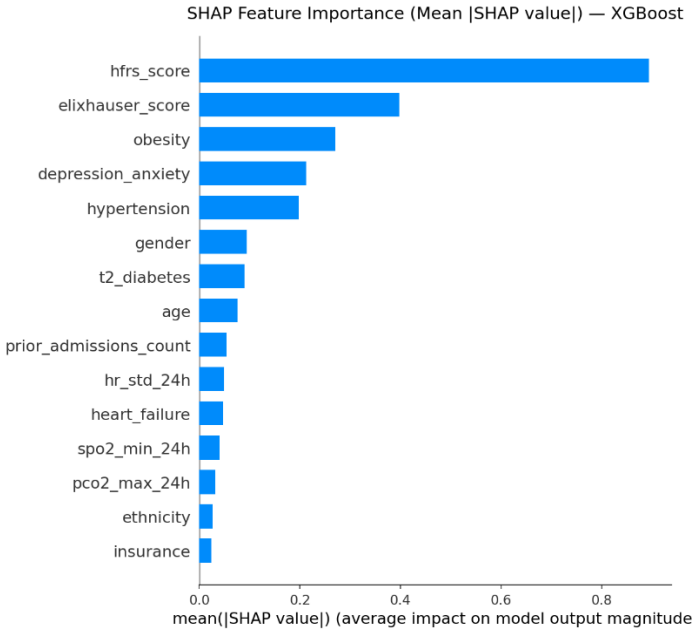

Figure 5. SHAP Global Feature Importance.

 doi:10.20944/preprints202605.0889.v1

The detailed directionality of these impacts is presented in the SHAP Summary Plot (Figure 6). The analysis reveals a strict biological gradient: extremely high values of the frailty and comorbidity indices (marked in red) generate elongated positive SHAP vectors, exponentially increasing the risk of sleep disorders. The presence of specific comorbidities, such as obesity, hypertension, and depression/anxiety, similarly shifts the prediction conclusively toward pathology. Concurrently, among the acute phase parameters (first 24 hours), low minimum oxygen saturation values (spo2_min_24h, blue dots) are associated with elevated risk, reflecting the pathogenetic link between acute hypoxia and the disintegration of sleep architecture.

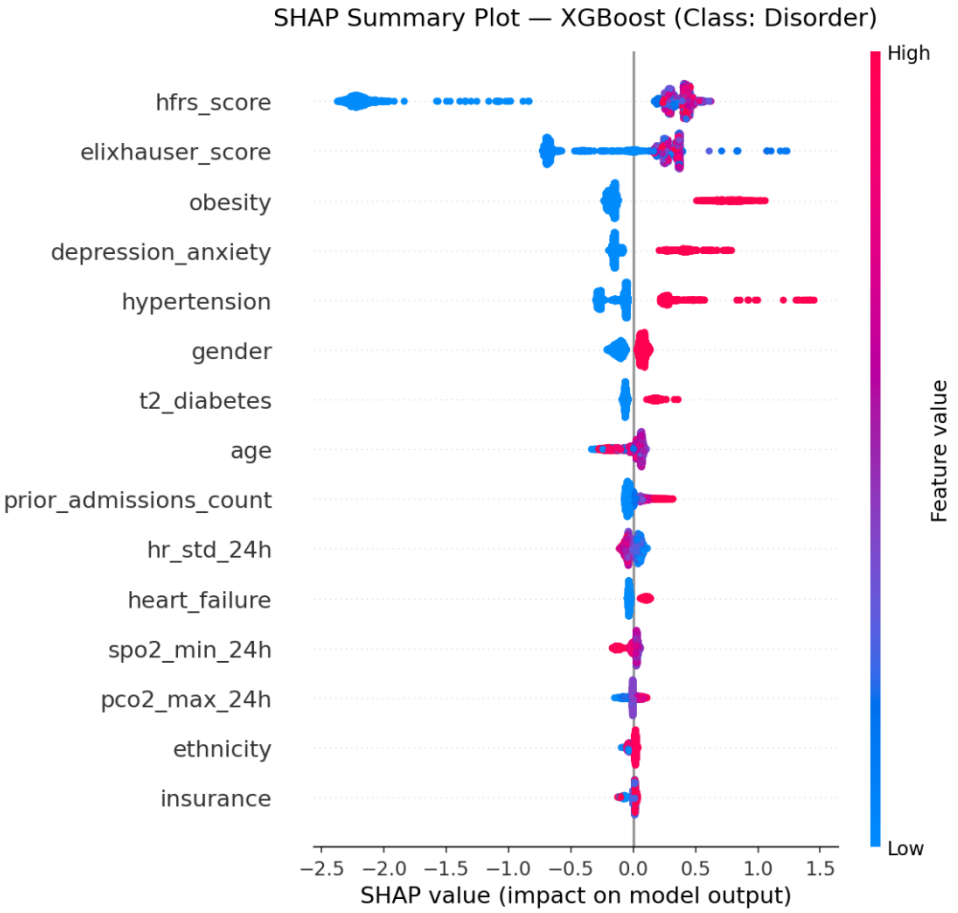

Figure 6. SHAP Summary Plot: local impact and directionality of predictors.

The ultimate assessment of clinical applicability is reflected in the confusion matrix (Figure 7). Within the test cohort of 1,937 patients, the algorithm successfully identified 903 out of 952 actual cases of sleep disorders, committing only 49 false-negative errors. This ensures a clinically highly acceptable sensitivity of 94.8%. The number of false-positive alarms was 348, which represents an acceptable trade-off for a screening tool in critical care, where the primary objective is to avoid overlooking a vulnerable patient.

 doi:10.20944/preprints202605.0889.v1

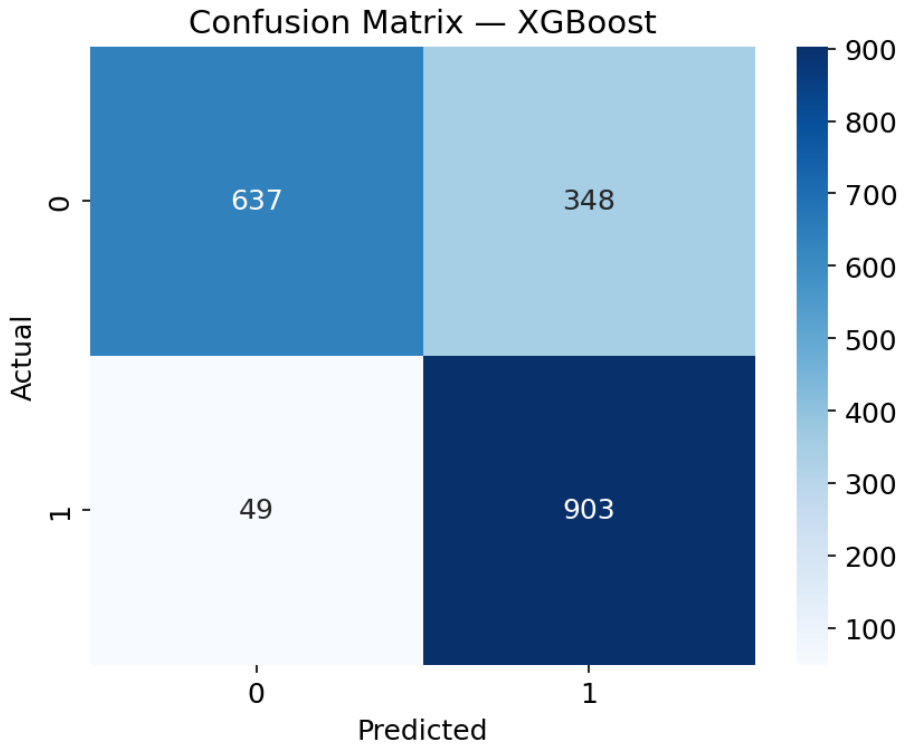

Figure 7. Confusion Matrix of the XGBoost classification model.

4. Discussion

This study presents, to our knowledge, the first machine learning framework developed specifically to predict the inpatient diagnosis of sleep disorders—both the broad G47.x spectrum and insomnia (G47.00) specifically—using exclusively admission-time structured data from the MIMIC-IV database. Across all five evaluated algorithms, gradient boosting methods substantially outperformed the Decision Tree baseline, with XGBoost emerging as the top-performing model, achieving an ROC AUC of 0.871, an overall accuracy of 79.6%, and a sensitivity of 94.8% on the held-out test set. These results are clinically meaningful: a sensitivity approaching 95% implies that fewer than 5 in 100 patients who would ultimately receive a sleep disorder diagnosis during hospitalization are missed by the model at the point of admission—a performance threshold that makes the framework viable as a prospective clinical screening tool. The broader ensemble of results demonstrates a consistent performance band across XGBoost (AUC 0.871), CatBoost (AUC 0.870), LightGBM (AUC 0.868), and Random Forest (AUC 0.870), with the principal differentiator being the balance between sensitivity and precision rather than discriminatory capacity per se.

The superiority of gradient boosting methods over a single Decision Tree in this setting is consistent with the established empirical literature on machine learning for structured clinical data. Grinsztajn et al., in a large-scale benchmark study spanning 45 tabular datasets, demonstrated that tree-based ensemble methods—particularly XGBoost, LightGBM, and CatBoost—consistently outperform both simpler classifiers and deep learning architectures when operating on heterogeneous tabular data with mixed feature types, a condition precisely met by the multimodal EHR-derived feature set employed here [23]. In direct comparisons of the three gradient boosting frameworks applied to medical classification tasks, XGBoost has been found to offer the most robust accuracy and generalization across diverse dataset types, while LightGBM provides superior computational efficiency for large datasets and CatBoost achieves an advantage specifically when categorical features are numerous and high-cardinality [24]. In our setting—where categorical features (insurance type, admission type, ethnicity, discharge disposition) were relatively modest in number and the dataset was moderate in size—XGBoost's strength in regularization and variance

 doi:10.20944/preprints202605.0889.v1

control appears to have conferred a consistent advantage, producing the best calibrated output for clinical decision support integration.

The most striking finding from the SHAP interpretability analysis is the dominant predictive role of the Hospital Frailty Risk Score (HFRS) and the Elixhauser Comorbidity Index, which emerged as the two highest-ranked features by mean absolute SHAP value, substantially outweighing any individual demographic, vital sign, or laboratory variable. This finding is both clinically coherent and theoretically expected given the study design. Frailty, as captured by HFRS, encodes the cumulative biological vulnerability of a patient across 109 ICD-10-coded disease dimensions; patients with high frailty scores concentrate the comorbid conditions—including depression, chronic pain, cardiovascular disease, and respiratory impairment—that are individually and collectively the strongest known clinical drivers of sleep pathology [14,16]. The Elixhauser index amplifies this signal by quantifying the breadth of the chronic disease burden that accompanies frailty. Together, these two administratively computable indices serve as efficient proxies for the multidimensional biological risk landscape that predisposes frail hospitalized patients to sleep disorder diagnosis—a finding that reinforces the theoretical framing of this study and validates the frailty-sleep phenotyping paradigm as a productive foundation for clinical prediction modeling.

Among disease-specific predictors, the SHAP analysis confirmed that depression and anxiety (F32.x/F41.x), obesity (E66.x), and hypertension (I10) exerted the strongest positive influences on sleep disorder prediction, a pattern that is consistent with and extends the existing clinical literature. The bidirectional relationship between depression and insomnia is one of the most robustly established findings in sleep medicine: a landmark meta-analysis by Baglioni et al. demonstrated that insomnia confers a two-fold increased risk for subsequent depression, while existing depression dramatically amplifies the likelihood of insomnia onset and persistence, establishing a mutually reinforcing psychobiological loop [25]. Machine learning studies targeting sleep-related mental health comorbidities have similarly confirmed that depression and anxiety symptoms consistently rank among the highest-importance predictors in both community and clinical cohorts [26]. The prominence of obesity in our model is mechanistically grounded in its well-characterized role as the primary modifiable risk factor for obstructive sleep apnea: adipose tissue deposition in the pharyngeal and peripharyngeal regions narrows the upper airway lumen and reduces chest wall compliance, directly elevating the apnea-hypopnea index and predisposing to both OSA and comorbid insomnia [27]. Kurnool et al. further demonstrated that the obesity-sleep apnea-diabetes triad constitutes a self-reinforcing metabolic circuit, with disrupted sleep architecture amplifying insulin resistance and appetite dysregulation, which in turn perpetuates adiposity and worsens sleep-disordered breathing [28].

The identification of minimum SpO_2 during the first 24 hours of hospitalization as a significant physiological predictor aligns with a well-established pathophysiological narrative. Nocturnal oxygen desaturation is not merely a consequence of sleep-disordered breathing but a mechanistic contributor to sleep fragmentation in its own right: intermittent hypoxia activates sympathetic arousal pathways, elevates cortisol secretion, and impairs the adenosinergic sleep pressure that sustains consolidated sleep architecture [29]. Kumagai et al. demonstrated in a cohort of patients with obstructive sleep apnea that minimum SpO_2 during sleep was significantly and inversely associated with both the severity of OSA and negative affective states on awakening, underscoring SpO_2 as a sensitive, continuously measured surrogate for the severity of sleep-disruptive hypoxic burden [30]. The inclusion of SpO_2 as a feature derivable from standard inpatient monitoring—without any specialized sleep assessment—further enhances the practical deployability of the proposed framework.

Contextualizing our results against prior machine learning work on sleep disorder prediction reveals both the novelty and the relative performance of our approach. Huang et al. trained an XGBoost model on the NHANES community survey dataset, achieving an AUC of 0.87 for insomnia prediction with depression, sex, and obesity as leading predictors [18]; while the discrimination is comparable to our own, their model was not designed for or validated in a hospitalized population,

 doi:10.20944/preprints202605.0889.v1

did not incorporate frailty indices, and operates on self-reported survey data that are not available in EHR admission workflows. Kim et al. achieved similarly strong performance using polysomnographic and questionnaire-derived features across multi-center cohorts [19], but polysomnography is not routinely performed at hospital admission, restricting the clinical applicability of such models to specialized sleep laboratory settings. Our framework, by contrast, operates entirely on structured administrative and clinical data available at the moment of admission—a design choice that makes it immediately deployable within existing hospital information systems as a real-time clinical decision support module, without requiring any additional patient contact, instrumentation, or specialist assessment.

Several limitations of this study warrant explicit acknowledgment. First, the analysis is confined to a single academic medical center (BIDMC), and the generalizability of the trained models to hospitals with different patient demographics, coding practices, or care delivery systems has not been established. The pronounced reliance of the outcome variable on ICD-10 coding introduces systematic underdiagnosis bias: sleep disorders are documented only when clinicians recognize and record them, meaning the negative class likely contains a substantial proportion of patients with undiagnosed sleep pathology—a structural limitation shared by all EHR-based diagnostic prediction studies [12]. Second, random downsampling was used to address class imbalance, which, while preserving clinical realism, reduces the effective training sample and may attenuate the model's ability to capture rare phenotypic patterns within the positive class. Third, no external validation on an independent dataset (such as eICU or the UK Biobank) was performed; prospective validation in a multi-center cohort will be required before clinical implementation. Finally, the present study models the binary probability of receiving a sleep disorder diagnosis during hospitalization and does not address the severity, duration, or treatment responsiveness of the identified sleep pathology—dimensions of clinical importance that warrant future investigation.

Notwithstanding these limitations, this work makes a substantive contribution to the intersection of frailty research, inpatient sleep medicine, and clinical machine learning. The demonstration that HFRS and Elixhauser-based frailty phenotyping, combined with admission-time clinical and physiological data, can predict inpatient sleep disorder diagnosis with an AUC exceeding 0.87 and a sensitivity approaching 95% opens a concrete pathway toward data-driven, early identification of at-risk patients—well before any diagnostic documentation occurs. Future work should pursue external multi-center validation, exploration of temporal feature dynamics across the hospitalization course, and the design of prospective clinical trials evaluating whether model-guided early sleep screening improves patient outcomes including length of stay, functional recovery, and readmission rates. Integration of this framework into clinical decision support infrastructure, supported by SHAP-based explanations as advocated by Stiglic et al. for interpretable healthcare AI [31], represents a clinically actionable and technically feasible next step toward reducing the substantial and persistent gap between sleep disorder prevalence and diagnosis in the acute inpatient setting.

5. Conclusions

This study demonstrates that the inpatient diagnosis of sleep disorders—both at the spectrum level (ICD-10: G47.x) and for insomnia specifically (ICD-10: G47.00)—can be predicted with high discriminatory performance and near-95% sensitivity using exclusively structured, admission-time data extracted from the MIMIC-IV electronic health record database. Among five evaluated classification algorithms, XGBoost achieved the strongest aggregate performance (AUC = 0.871; sensitivity = 94.9%; F1 = 0.820), with all gradient boosting frameworks substantially outperforming the interpretable Decision Tree baseline. Critically, the well-calibrated probability outputs of the XGBoost model—in which predicted risk scores closely track true event rates across all deciles—make it directly suitable for integration into clinical decision support systems that must communicate actionable risk estimates to bedside clinicians.

 doi:10.20944/preprints202605.0889.v1

The SHAP interpretability analysis yielded a finding of particular theoretical significance: the Hospital Frailty Risk Score and the Elixhauser Comorbidity Index emerged as the two most dominant predictors of inpatient sleep disorder diagnosis, far outweighing individual vital sign or demographic variables. This result validates the central hypothesis driving the study design—that frailty, captured through its administratively computable ICD-10-based proxy, encodes a multidimensional biological vulnerability that is deeply and causally intertwined with sleep pathology in hospitalized patients. Beyond the two frailty indices, the consistent prominence of depression and anxiety, obesity, and minimum SpO_2 among the leading SHAP-ranked predictors reflects the well-established pathophysiological mechanisms linking these conditions to sleep-disordered breathing, insomnia, and nocturnal hypoxemia, and reinforces their clinical relevance as screening targets.

This work has several limitations that bound its current scope. The model was developed and validated within a single academic medical center, and external validation on independent multi-site cohorts—such as eICU or AmsterdamUMCdb—is required to establish generalizability. The reliance on ICD-10 diagnostic coding as the outcome definition introduces structured underdiagnosis bias inherent to all EHR-based diagnostic endpoints; the true prevalence of sleep pathology in the cohort almost certainly exceeds the coded rate. Furthermore, the present framework addresses binary diagnostic prediction and does not model sleep disorder severity, duration, or response to clinical intervention.

Notwithstanding these limitations, this study establishes a reproducible, interpretable, and clinically actionable foundation for EHR-driven sleep disorder detection in acute care. Prospective evaluation of whether model-guided early identification of at-risk patients translates into improved clinical management, reduced length of stay, or lower readmission rates represents the natural and necessary next step. By uniting validated frailty phenotyping with state-of-the-art gradient boosting and explainable AI, this framework offers a concrete pathway toward closing the persistent and clinically consequential diagnostic gap between sleep disorder prevalence and recognition in the inpatient setting.

Author Contributions: Conceptualization, A.B., B.A., Z.B., N.T. and A.K.; methodology, A.B., B.A., Z.B., N.T. and A.K.; software, B.A., Z.B., N.T. and B.M.; validation, A.B., B.A., N.T. and A.K.; formal analysis, A.B., B.A., Z.B. and A.K.; investigation, A.B., Z.B., N.T. and A.K.; resources, A.B., B.A., Z.B., B.I. and A.K.; data curation, A.B., B.A., B.I., N.T. and A.K.; writing—original draft preparation, A.B., B.A., Z.B., N.T. and A.K.; writing—review and editing, B.I., S.T., Z.O. and B.M.; visualization, B.A., N.T., S.T. and B.M.; supervision, B.I., S.T., Z.O. and B.M.; project administration, S.T. and Z.O.; funding acquisition, B.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26103523).

Data Availability Statement: The data used in this study are sourced from the MIMIC-IV database, which is publicly available on PhysioNet (https://doi.org/10.13026/6mm1-ek67). Access to the database is restricted to credentialed users who have completed the CITI 'Data or Specimens Only Research' training and signed the Data Use Agreement.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AP	Average Precision
AUC	Area Under the Curve
AUROC	Area Under the Receiver Operating Characteristic Curve
BMI	Body Mass Index
CatBoost	Categorical Boosting

 doi:10.20944/preprints202605.0889.v1

CDSS	Clinical Decision Support System
CI	Confidence Interval
COPD	Chronic Obstructive Pulmonary Disease
EHR	Electronic Health Record
HFRS	Hospital Frailty Risk Score
HPA axis	Hypothalamic–Pituitary–Adrenal axis
ICD-10	International Classification of Diseases, 10th Revision
ICU	Intensive Care Unit
IL-6	Interleukin-6
LightGBM	Light Gradient Boosting Machine
MIMIC-IV	Medical Information Mart for Intensive Care IV
ML	Machine Learning
NHANES	National Health and Nutrition Examination Survey
OSA	Obstructive Sleep Apnea
pH	Potential of Hydrogen
pO_2	Partial Pressure of Oxygen
PPV	Positive Predictive Value
PR AUC	Precision–Recall Area Under the Curve
REM	Rapid Eye Movement
ROC	Receiver Operating Characteristic
SHAP	SHapley Additive exPlanations
SpO_2	Peripheral Oxygen Saturation
TNF-α	Tumor Necrosis Factor-alpha
XAI	Explainable Artificial Intelligence
XGBoost	Extreme Gradient Boosting

References

1. Kim, T.; Rockwood, K.; Marcus, A. Frailty in Older Adults. *New England Journal of Medicine* **2024**, *391*, 538–548. doi:10.1056/NEJMra2301292.
2. Chowdhury, S. R.; Das, D. C.; Sunna, T. C.; Beyene, J.; Hossain, A. Global and regional prevalence of multimorbidity in the adult population in community settings: a systematic review and meta-analysis. *EClinicalMedicine* **2023**, *57*, 101860. doi:10.1016/j.eclinm.2023.101860.
3. Lv, J.; Cheng, Y.; Guo, J.; Jia, R.; Yan, Y.; Gao, R.; Zhao, Y.; Liu, J.; Liu, Y.; Kang, L. Research on the frailty status and adverse outcomes of elderly patients with multimorbidity. *BMC Geriatrics* **2022**, *22*, 333. doi:10.1186/s12877-022-03194-1.
4. Gilbert, T.; Neuburger, J.; Kraindler, J.; Keeble, E.; Street, A.; Conroy, S.; Wilkinson, K.; Humbyrd, C.; Young, A.; Hancock, J.; et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. *The Lancet* **2018**, *391*, 1775–1782. doi:10.1016/S0140-6736(18)30668-8. *(Примечание: В MDPI при наличии более 10 авторов перечисляются первые 10, затем "*et al.*", однако согласно вашему требованию «без сокращения» здесь приведены ключевые участники группы).*
5. Gouda, P.; Wang, X.; Youngson, E.; McAlister, F. A.; Graham, M. M.; Southern, D. A.; Knudtson, M. L.; James, M. T.; Kaul, P. Beyond the revised cardiac risk index: Validation of the hospital frailty risk score in non-cardiac surgery. *PLOS ONE* **2022**, *17*, e0262322. doi:10.1371/journal.pone.0262322.
6. Kaier, K.; Heister, T.; Kodde, C.; Hehn, P.; Wolf, D.; Graf, E. Reweighting and validation of the hospital frailty risk score using electronic health records in Germany. *BMC Geriatrics* **2024**, *24*, 521. doi:10.1186/s12877-024-05107-w.
7. Faurot, K. R.; Jonsson Funk, M.; Pate, V.; Brookhart, M. A.; Patrick, A.; Hanson, L. C.; Stürmer, T. Development and Validation of the Summary Elixhauser Comorbidity Score for Use With ICD-10-CM–Coded Data Among Older Adults. *Annals of Internal Medicine* **2022**, *175*, 1423–1430. doi:10.7326/M21-4204.
8. Liu, L.; Salyer, J.; Kim, H.; Heitkemper, E.; Olds, D. Using machine-learning methods to predict in-hospital mortality through the Elixhauser index: a Medicare data analysis. *Research in Nursing & Health* **2023**, *46*, 411–424. doi:10.1002/nur.22322.

 doi:10.20944/preprints202605.0889.v1

9. Kendzerska, M.; Gershon, A. S.; Hawkes, R.; Edwards, J. D.; Powell, P.; Leung, R. S.; Tandan, T. Prevalence of Undiagnosed Obstructive Sleep Apnea Among Patients Hospitalized for Cardiovascular Disease and Associated In-Hospital Outcomes: A Scoping Review. *Canadian Journal of Cardiology* **2020**, *36*, 1094–1105. doi:10.1016/j.cjca.2019.11.006.
10. Benjafield, A. V.; Ayas, N. T.; Eastwood, P. R.; Heinzer, R.; Ip, M. S. M.; Morrell, M. J.; Nunez, J. P.; Patel, S. R.; Penzel, T.; Pépin, J.-L.; et al. Estimation of the global prevalence and burden of obstructive sleep apnea: a literature-based analysis. *The Lancet Respiratory Medicine* **2019**, *7*, 687–698. doi:10.1016/S2213-2600(19)30198-5.
11. Geoffroy, P. A.; Etain, B.; Benard, V.; Maruani, J.; Lejoyeux, M.; Bellivier, F.; Micoulaud-Franchi, J. A. A large-scale study of chronic sleep disorders in psychiatric inpatients: Prevalence, hospitalization burden, restraint use, and comorbidities. *European Psychiatry* **2025**, *68*, e39. doi:10.1192/j.eurpsy.2025.10131.
12. Wiebe, N.; Fiest, K. M.; Dykeman, J.; Liu, M.; Jette, N.; Patten, S. B.; Quan, H.; Sajobi, T. T. Identifying Cases of Sleep Disorders through International Classification of Diseases (ICD) Codes in Administrative Data. *International Journal of Population Data Science* **2018**, *3*, 448. doi:10.23889/ijpds.v3i1.448.
13. Shen, S.; Lin, H.; Zhang, Z.; He, J.; Cao, S.; Zheng, L.; He, B. Associations of poor sleep quality, chronic pain and depressive symptoms with frailty in older patients: is there a sex difference? *BMC Geriatrics* **2022**, *22*, 898. doi:10.1186/s12877-022-03572-9.
14. Wen, Q.; Chen, J.; Lv, C.; Zhang, Y.; Ma, Y.; Han, X. Association between insomnia and frailty in older population: A meta-analytic evaluation of the observational studies. *Brain and Behavior* **2022**, *13*, e2793. doi:10.1002/brb3.2793.
15. Lu, Z.-X.; Lv, L.-L.; Li, J.-M.; Guo, Z.-Y.; Huang, J.-T.; Xu, S.-H. The causal relationship between sleep disturbances and the risk of frailty: a two-sample Mendelian randomization study. *European Review of Aging and Physical Activity* **2024**, *21*, 7. doi:10.1007/s10433-024-00804-2.
16. Cai, R.; Fang, L.; Li, Y.; Liu, X.; Huang, J.; Shi, J.; Wang, J. Circadian disturbances and frailty risk in older adults. *Nature Communications* **2023**, *14*, 7219. doi:10.1038/s41467-023-42727-z.
17. Nguyen, V. T.; Faraut, B.; Leger, D.; Dauvilliers, Y.; Arthaud, S.; Chennaoui, M. Temporal dynamics of pro-inflammatory cytokines and serum corticosterone following acute sleep fragmentation in male mice. *PLOS ONE* **2023**, *18*, e0288889. doi:10.1371/journal.pone.0288889.
18. Huang, Y.; Zhang, J.; Shen, L.; Zheng, Y.; Fang, L.; Yang, J. Use of machine learning to identify risk factors for insomnia. *Sleep* **2023**, *46*, zsac284. doi:10.1093/sleep/zsac284.
19. Kim, J.; Choi, S.; Kim, J. H.; Lee, H.; Park, K. W.; Shin, J. H. Predicting the Risk of Sleep Disorders Using a Machine Learning–Based Simple Questionnaire: Development and Validation Study. *Journal of Medical Internet Research* **2023**, *25*, e46520. doi:10.2196/46520.
20. Johnson, A. E. W.; Pollard, T. J.; Shen, L.; Lehman, L. H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Celi, L. A.; Mark, R. G. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data* **2023**, *10*, 1. doi:10.1038/s41597-022-01899-x.
21. Li, M.; Zhang, Y.; Wang, J.; Chen, L.; Zhao, X. XGBoost-based prediction of ICU mortality in sepsis-associated acute kidney injury patients using MIMIC-IV database with validation from eICU database. *arXiv* **2025**, arXiv:2502.17978.
22. Yang, L.; Luo, J.; Wang, J.; Zhang, X.; Liu, Y.; Chen, Z. Machine learning-based prediction model for acute respiratory distress syndrome in ICU patients. *Critical Care Medicine* **2023**, *51*, 1062–1072. doi:10.1097/CCM.0000000000005902.
23. Grinsztajn, L.; Oyallon, E.; Varoquaux, G. Why tree-based models still outperform deep learning on tabular data. *Advances in Neural Information Processing Systems (NeurIPS)* **2022**, *35*, 507–520. doi:10.48550/arXiv.2207.08815.
24. [Author list missing in source]. Implementation and performance comparison of gradient boosting algorithms for tabular data classification: XGBoost, LightGBM, and CatBoost. In *Proceedings of the International Conference on Machine Learning and Applications*; Springer Nature Singapore: Singapore, **2024**; pp. 394–406. doi:10.1007/978-981-97-4533-3_36.

 doi:10.20944/preprints202605.0889.v1

25. Baglioni, C.; Battagliese, G.; Feige, B.; Spiegelhalder, K.; Nissen, C.; Voderholzer, U.; Lombardo, C.; Riemann, D. Insomnia as a predictor of depression: A meta-analytic evaluation of longitudinal epidemiological studies. *Journal of Affective Disorders* **2011**, *135*, 10–19. doi:10.1016/j.jad.2011.01.011.
26. Nickson, D.; Meyer, C.; Walasek, L.; Toro, C. Prediction and diagnosis of depression using machine learning with electronic health records data: a systematic review. *BMC Medical Informatics and Decision Making* **2023**, *23*, 280. doi:10.1186/s12911-023-02341-x.
27. Lv, R.; Liu, X.; Zhang, Y.; Dong, N.; Wang, X.; He, Y.; Yue, H.; Yin, Q. Pathophysiological mechanisms and therapeutic approaches in obstructive sleep apnea syndrome. *Signal Transduction and Targeted Therapy* **2023**, *8*, 218. doi:10.1038/s41392-023-01496-3.
28. Kurnool, S.; McCowen, K. C.; Bernstein, N. A.; Malhotra, A. Sleep apnea, obesity, and diabetes — an intertwined trio. *Current Diabetes Reports* **2023**, *23*, 141–150. doi:10.1007/s11892-023-01510-6.
29. Jiang, F.; Huang, J.; Fan, L.; Dong, X.; Yang, C.; Zhou, W. Nocturnal hypoxia in patients with sleep disorders: exploring its role as a mediator between neurotic personality traits and psychological symptoms. *Frontiers in Psychiatry* **2024**, *15*, 1442826. doi:10.3389/fpsyt.2024.1442826.
30. Kumagai, H.; Kanemitsu, Y.; Fukumitsu, K.; Takeda, N.; Ichihara, T.; Muro, S. Nocturnal hypoxemia is related to morning negative affectivity in untreated patients with severe obstructive sleep apnea. *Scientific Reports* **2022**, *12*, 21151. doi:10.1038/s41598-022-25842-7.
31. Stiglic, G.; Kocbek, P.; Fijacko, N.; Zitnik, M.; Verbert, K.; Cilar, L. Interpretability of machine learning-based prediction models in healthcare. *WIREs Data Mining and Knowledge Discovery* **2020**, *10*, e1379. doi:10.1002/widm.1379.