JMIR Publications
The Leading eHealth Publisher

Risk Stratification for Post-Hospitalization Case Management in Diabetes: Integrating Clinical and Socioeconomic Data

Seung-Yup Lee, Mohammad Saleem, Andrew Land, Erin DeLaney, Alison Garretson, Mahee Patel, Allyson G Hall, Salisa Westrick, Fernando Ovalle, Andrea Cherrington, Jane Banaszak-Holl

Submitted to: JMIR Medical Informatics
on: February 14, 2025

Preprint
JMIR Publications

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Preprint
JMIR Publications

Risk Stratification for Post-Hospitalization Case Management in Diabetes: Integrating Clinical and Socioeconomic Data

Seung-Yup Lee[1] PhD; Mohammad Saleem[1] MS; Andrew Land[2] MD; Erin DeLaney[1] MD; Alison Garretson[2] MBA, RN; Mahee Patel[1] MBA; Allyson G Hall[1] PhD; Salisa Westrick[3] PhD; Fernando Ovalle[1] MD; Andrea Cherrington[1] MD; Jane Banaszak-Holl[1] PhD

[1] University of Alabama at Birmingham Birmingham US
[2] UAB Medicine Birmingham US
[3] Auburn University Auburn US

Corresponding Author:
Seung-Yup Lee PhD

University of Alabama at Birmingham
1720 University Blvd
Birmingham
US

Abstract

Background: Diabetes remains a critical public health issue, affecting millions of individuals and contributing to significant healthcare costs. Patients with diabetes face elevated risks of emergency department (ED) visits post-hospitalization, driven by both clinical factors, such as comorbidities, and social determinants of health (SDoH).

Objective: Effective and interpretable predictive models can help identify high-risk patients with diabetes to improve case management post-hospitalization.

Methods: This study utilized retrospective cohort data from the University of Alabama at Birmingham Medical Center, focusing on patients with diabetes hospitalized between January 2020 and March 2024. Clinical data and SDoH variables were integrated into logistic regression, decision trees, and XGBoost models to predict ED visits within three months post-hospitalization. Performance metrics, such as area under the curve (AUC), precision, sensitivity, and specificity, were used to evaluate the models.

Results: Key predictors of ED visits included past three-month ED visits, insulin usage (aspart and glargine), and SDoH indicators such as the Area Deprivation Index and Social Vulnerability Index. XGBoost demonstrated excellent performance with an AUC of 0.846, outperforming both decision trees and logistic regression, though decision trees provided greater interpretability.

Conclusions: This study highlights the importance of integrating clinical and SDoH factors to predict post-hospitalization ED visits, supporting care management in patients with diabetes. While XGBoost provided superior prediction performance, the trade-off between accuracy and interpretability suggests that decision trees may be better suited for contexts where actionable insights are critical for care providers. Future research should address missing SDoH data and explore the real-world applicability and scalability of these models.

(JMIR Preprints 14/02/2025:72624)
DOI: https://doi.org/10.2196/preprints.72624

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

No. Please do not make my accepted manuscript PDF available to anyone.

Preprint JMIR Publications

Original Manuscript

Preprint
JMIR Publications

Risk Stratification for Post-Hospitalization Case Management in Diabetes: Integrating Clinical and Socioeconomic Data

Abstract

Background: Diabetes remains a critical public health issue, affecting millions of individuals and contributing to significant healthcare costs. Patients with diabetes face elevated risks of emergency department (ED) visits post-hospitalization, driven by both clinical factors, such as comorbidities, and social determinants of health (SDoH).

Objective: Effective and interpretable predictive models can help identify high-risk patients with diabetes to improve case management post-hospitalization.

Methods: This study utilized retrospective cohort data from the University of Alabama at Birmingham Medical Center, focusing on patients with diabetes hospitalized between January 2020 and March 2024. Clinical data and SDoH variables were integrated into logistic regression, decision trees, and XGBoost models to predict ED visits within three months post-hospitalization. Performance metrics, such as area under the curve (AUC), precision, sensitivity, and specificity, were used to evaluate the models.

Results: Key predictors of ED visits included past three-month ED visits, insulin usage (aspart and glargine), and SDoH indicators such as the Area Deprivation Index and Social Vulnerability Index. XGBoost demonstrated excellent performance with an AUC of 0.846, outperforming both decision trees and logistic regression, though decision trees provided greater interpretability.

Discussion: This study highlights the importance of integrating clinical and SDoH factors to predict post-hospitalization ED visits, supporting care management in patients with diabetes. While XGBoost provided superior prediction performance, the trade-off between accuracy and interpretability suggests that decision trees may be better suited for contexts where actionable insights are critical for care providers. Future research should address missing SDoH data and explore the real-world applicability and scalability of these models.

1. Introduction

Diabetes is a major global health issue, affecting 38.4 million people in the United States alone and contributing to an annual healthcare cost of $412.9 billion.[1] Patients with diabetes are at higher risk of hospital admissions and frequent emergency department (ED) visits,[2] often due to comorbidities like heart failure, poor glycemic control, chronic kidney disease, and cardiovascular disease.[3] Effective management during the vulnerable period following hospital discharge is critical for reducing these risks.[4-7]

Diabetes case management is essential during this post-hospitalization period to ensure continuity of care, adherence to medications, and necessary lifestyle adjustments.[5,8] This period is especially crucial because gaps in care, medication mismanagement, or worsening of existing conditions can lead to acute complications.[9,10] Preventing ED visits during this time not only improves patient outcomes but also provides operational and financial benefits, easing the burden on EDs, enabling more efficient use of healthcare resources,[11] and reducing healthcare costs by

preventing repeat ED visits and hospital readmissions.[12]

Identifying patients most at risk for ED visits is crucial to effective post-hospitalization care, and predictive models incorporating clinical factors, such as HbA1c and comorbidities, have become valuable tools in guiding these efforts.[13-20] However, while clinical variables are important, there is a growing recognition of the role of social determinants of health (SDoH)—such as socioeconomic status, access to food, housing stability, and transportation—in influencing healthcare outcomes (such as ED visits) for patients with diabetes, particularly in vulnerable populations.[21] There's a growing recognition of the need for models that incorporate both clinical and SDoH data to better predict adverse events such as ED visits.[22-24] This study integrates both area-level SDoH data, such as zip code-derived indices, and individual-level, self-reported SDoH data with clinical factors. The incorporation of individual-level SDoH leverages the study health system's initiative to collect patient-reported information during their routine clinical encounters. This unique combination allows for more precise modeling and may highlight the importance of collecting SDoH data at various levels.

Recent advancements in machine learning offer a pathway to better prediction of adverse outcomes by combining both clinical and SDoH data and capturing their interactions. As models become more complex, model interpretability is crucial for real-world case management applications.[25,26] Through frameworks like Interpretable Artificial Intelligence (AI), these models not only enhance accuracy but also offer valuable insights into the factors driving ED visits.

This study aims to develop predictive models with different levels of model interpretability for 3-month post-hospitalization ED visits among patients with diabetes by integrating various clinical data and SDoH. The goal is to evaluate both the accuracy and interpretability of models, including logistic regression, decision trees, and XGBoost, while visualizing the results to provide actionable insights to healthcare providers. These insights can support case management to enhance patient outcomes and optimize resource allocation.

2. Methods

2.1. Study setting

This study employs a retrospective cohort design using data from the University of Alabama at Birmingham (UAB) Medical Center, a large academic medical center in the southern United States. Our study cohort includes patients who have had the following conditions, identified by a HbA1c level and/or International Classification of Diseases, 10th Revision (ICD-10)[27] codes: HbA1c ≥ 6.5%, diabetes mellitus due to underlying condition (E08), drug or chemical induced diabetes mellitus (E09), type 1 diabetes mellitus (E10), type 2 diabetes mellitus (E11), other specified diabetes mellitus (E13), elevated blood glucose level (R73), hypoglycemia (E16.2), foot ulcer (L97.509), or gastroparesis (K31.84). Data were extracted from electronic health records (EHR) spanning January 2020 to March 2024, covering 162,063 inpatient encounters. The predictive modeling process involves training and validating statistical and interpretable artificial models using 85% of the data for model training and 15% for validation.

2.2. Outcome variable

The primary outcome is a binary variable indicating whether a patient had an ED visit for diabetes-related concerns within three months post-hospitalization (‘Yes’ for an ED visit, ‘No’ otherwise). This timeframe is clinically and operationally significant as the hospital’s diabetes case manager follows up on discharged patients during this period (frequently identified as the high-risk period in the literature[28-30]). To qualify as a positive outcome, the ED visit must be related to diabetes, as defined by the ICD-10 codes listed in the 2.1 Study Setting section, either as a primary or secondary diagnosis. More details on the rationale for the three-month timeframe and the outcome variable are available in the related research protocol paper.[22]

2.3. Predictors

Our predictors included demographic, clinical, and SDoH variables.

2.3.1 Demographic information

Demographic details included age, gender, race, and marital status, all of which have been extensively associated with diabetes outcomes in the literature.[31,32]

2.3.2 Clinical information

Clinical data were extracted from the EHR, including comorbidities, laboratory test results, medications, and prior admissions.[33,34] Data cleaning involved standardizing formats, coding missing values as 'not measured,' and creating categorical variables. Expert clinical inputs from clinician co-authors and literature guided the transformation of raw data into meaningful predictors (Table 1).

2.3.3 SDoH data

While less commonly applied in predictive modeling, SDoH have been highlighted in explaining diabetes outcomes, including ED utilization.[44-46] We incorporated two types of SDoH data.

Area-level SDoH data: The Area Deprivation Index (ADI), developed by the Health Resources & Services Administration, a.k.a. HRSA, and reported by the Neighborhood Atlas,[47,48] ranks neighborhoods based on socioeconomic disadvantage (0 for least deprived and 100 for most deprived) by taking into account factors such as income, education, employment, and housing quality. The Social Vulnerability Index (SVI), created by the Agency for Toxic Substances and Disease Registry, includes a wider array of factors, covering not just socioeconomic status but also minority status, language, and transportation.[49,50] This index offers area-level insights into community resilience against regional hazards and stressors. Our study used the composite SVI, aggregating data from four key themes: socioeconomic status, household composition and disability, minority status and language, and housing type and transportation, reporting rank scores from 0 (least vulnerable) to

1 (most vulnerable). The Low Income, Low Access (LILA) measure, from the U.S. Department of Agriculture, identifies food deserts, areas where residents have restricted access to affordable and nutritious food options.[51,52] We particularly used the 'LILATRACTS_1AND10' variable, which indicates limited food access if no grocery store is within one mile (urban) or ten miles (rural).[53,54]

Individual-level SDoH data: Insurance status at admission was categorized into five groups: commercial, Medicaid, Medicare, uninsured, and unknown. Self-reported SDoH were collected from the EHR database storing the Protocol for Responding to and Assessing Patient Assets, Risks, and Experience (PRAPARE) survey responses.[55] The PRAPARE survey asks 21 questions, covering personal characteristics, family, resources, and social and emotional health. The tool was implemented at UAB in January 2020, targeting patients attending community health clinics or EDs and integrated in UAB's EHR system.[56] The variables from the PRAPARE survey include veteran status, housing, education, social support, employment status, and unmet needs (e.g., medicine, healthcare, food, utility, and childcare). Due to the high frequency of missing data in self-reported SDoH, we opted for a "not measured" category instead of imputation to avoid bias or complicating result interpretation.[57,58] This approach is especially relevant where missingness might reflect specific patient circumstances.[59,60] Given that the PRAPARE survey was administered in community health clinics and ED settings, patients with PRAPARE entries were generally more socioeconomically vulnerable while missing data may suggest less frequent use of safety-net service due to better access to care.

While the PRAPARE survey was administered for each patient, our predictive modeling was done at the level of individual patient visits. For each visit, we used the most recent PRAPARE data. If no PRAPARE data were available at the time of admission (e.g., an admission prior to the patient's first PRAPARE survey), we used the earliest available survey results. We hypothesized that individual-level SDoH factors remain relatively stable over time for a patient, unlike more dynamic clinical variables.

2.4. Predictive modeling

Multiple predictive modeling approaches were applied. We chose logistic regression, decision trees, and XGBoost methods to perform predictive modeling and check performance metrics. These three modeling approaches represent a traditional statistical inference-based model, an interpretable machine learning model, and an advanced, yet less interpretable machine learning model, respectively. The first 85% of observations were utilized for model training while the remaining 15% of observations, which are new to the trained model, were reserved for testing the performance and generalizability of the trained model.

Logistic regression, a widely used traditional statistical model, is favored for its simplicity and ease of interpretation, making it a mainstay in biomedical research.[61,62] It estimates the probability of an event by fitting data to a logistic curve. Logistic regression assumes a linear relationship between independent variables and the log odds of the dependent variable, unless non-linearity is explicitly accounted for in the model.[63,64]

In contrast, Decision Trees are machine learning algorithms that generates interpretable models. Specifically, we employed the Classification and Regression Trees (CART) method, which recursively splits the data into subsets based on input feature values, forming a tree-like model of decisions or outcomes. This structure facilitates both interpretation and visualization of the decision-making processes, as it illustrates how outcomes are derived. The data splits are optimized to maximize information gain or minimize impurity, resulting in homogeneous categories.[65]

XGBoost, a more sophisticated machine learning algorithm, addresses complex data patterns using an ensemble approach. Through gradient boosting, it sequentially builds decision trees, each refining the errors of the previous ones.[66] This method incrementally combines the predictions of multiple trees to improve accuracy. XGBoost has demonstrated superior performance in domains including healthcare.[67-69]

2.5. Performance measures

We evaluated model performance using sensitivity, precision, specificity, and the area under the curve (AUC). Sensitivity measures the ability to identify true positives, while specificity assesses true negative identification. Precision measures the accuracy of positive predictions (i.e., true positives among all positive predictions). The AUC, widely accepted as a classification performance metric, provides a more comprehensive measure by assessing the trade-off between sensitivity and specificity across various prediction probability thresholds. AUC values above 0.7 are considered acceptable, and values above 0.8 are deemed excellent.[61] These metrics are crucial in healthcare, balancing accurate condition identification and minimizing false alarms, which impacts patient care and resource allocation.

2.6. Feature importance

Feature importance helps interpret model outputs. In this study, we focus on feature importance found from the XGBoost model due to the model's holistic way of processing variables (i.e., non-linearity and interaction across multiple decision trees). While in statistical inference, feature importance is commonly determined by the p-value of a variable's coefficient; in machine learning, it is based on the contribution of each feature to the model's construction or performance. XGBoost uses 'Gain' to assess a feature's contribution to model accuracy,[66] 'Cover' to reflect the number of observations influenced by a feature,[65] and 'Frequency' to show how often a feature is used in splits.[65] Gain is often prioritized for its direct link to accuracy, while Cover and Frequency provide context for comprehensive model interpretation and understanding feature interactions.

3. Results

3.1. Descriptive analysis results

Demographics and Social Determinants of Health.

Table 2 shows results from univariate and bivariate analyses related to patients' ED visit status three months post-hospitalization. Among 162,063 hospitalizations between January 2020 and March 2024 involving patients with a history of diabetes, 6.2% returned to the ED for diabetes-related reasons within 3 months post-hospitalization. The average patient age was 56.1 years (SD = 17.9), with significant differences between those who had an ED visit in the past three months (57.5 years) and those who did not (56.0 years, t=-10.1, p<.001). Most patients were White (54.8%) and male (51.6%), with the majority insured by commercial plans (34.5%), followed by Medicaid (32.2%) and Medicare (24.8%).

The average ADI was 71.7 (SD = 23.1), and the average SVI was 0.57 (SD = 0.29). Of total visits in the study, 21% are from LILA populations. Significant differences were observed by race, gender, and insurance type (all p<.001), as well as by ADI (76.4 vs. 71.4), SVI (0.64 vs. 0.57), and LILA status (31.1% vs. 20.6%, all p<.001) between those with 3-month ED visits and those without. Unmet needs (care, medication, utilities, and food) also showed significant differences between groups (all p<.001).

The overall average frequency of ED visits in the past three months was 0.36 times (SD = 1.4), with those who had visited the ED showing significantly higher visit frequency (0.98 vs. 0.32 times, p<.001).

Diabetes Related Comorbidities.

The most common diabetes-related comorbidity in this study was Type 2 diabetes with hyperglycemia (13.5%), followed by Type 2 diabetes with diabetic chronic kidney disease (11.1%), Type 2 diabetes without complications (9.7%), and hyperglycemia (6.5%). All other comorbidities were observed in less than 5% of patient visits.

Patients who had an ED visit within three months post-hospitalization were significantly more likely to have various diabetes-related comorbidities. Those with Type 2 diabetes with hyperglycemia were nearly twice as likely to be in the three-month ED visit group compared to those without an ED visit

(p<.001). Patients with Type 2 diabetes with chronic kidney disease were 2.5 times more likely to belong to the three-month ED visit group (p<.001). Less common comorbidities, such as diabetic autonomic neuropathy and gastroparesis, showed even higher odds, with ED visitors being over 5 times more likely to have these conditions (both p<.001).

Surgical Procedures.

The most common surgical procedures observed were cardiovascular (27.3%), digestive (13.4%), musculoskeletal (12.6%), and nervous system procedures (10.6%). Patients who did not visit the ED three months post-hospitalization were more likely to have had surgeries during their hospitalization across all systems. Notably, they were almost twice as likely to have undergone nervous system (p<.001) or respiratory system procedures (p<.001) during their hospitalization. Additionally, those without an ED visit post-hospitalization were more than six times as likely to have had a maternity care and delivery procedure (p<.001) and were about 2.5 times more likely to have undergone operating microscope, hemic and lymphatic system, or female genital system procedures (all p<.001) during their hospital stays.

Others.

The majority of patients in this study were obese (41.5%), followed by overweight (26.4%), normal weight (25.2%), and underweight (4.2%) with a higher proportion of patients with obesity among those who had 3-month ED visits post-hospitalization. Most had high glucose levels (52.5%), while most had low (46.2%) or normal (37.2%) eGFR levels with significantly different distributions between the two 3-month ED visit groups (all p<.001).

Significant differences were found between patients who visited the ED within the three months post-hospitalization than those who did not based on BMI (p<.001), glucose (p<.001), and eGFR levels (p<.001).

Most patients used aspart insulin (57.9%), followed by glargine insulin (32.3%) and regular insulin (16.9%). Patients who visited the ED within the three months post-hospitalization were over

three times as likely to be on detemir insulin (p<.001) or isophane insulin (p<.001), with higher likelihoods for all insulin types.

3.2. Predictive analysis results

Figure 1 is a visualization of the decision tree highlighting key risk factors and interactions among patients with diabetic concerns, aiding in predicting their 3-month ED visit probability. The tree starts by splitting the data based on a primary feature at the top, distinguishing between patients who do not meet a certain threshold and those who do. Subsequent splits refine these groups based on additional factors such as specific thresholds or binary conditions. The numbers in the leaf nodes indicate the counts of no ED visits (left) and ED visits (right) post-hospitalization respectively. The red leaf nodes represent the subgroups that have more ED visits than no ED visits. Insulin status, having a past 3-month ED visit history, and age appear to be the most influential variables. The tree structure helps identify other critical patterns: patients living in socially vulnerable area or with more complex comorbidities (e.g., mental, behavioral, and neurodevelopmental disorders) are flagged for greater ED visit risks. Each pathway in the tree, defined by numerical thresholds or yes/no splits, represents a patient subgroup with specific characteristics, allowing clinicians to target interventions for those most likely to return to the ED within three months of discharge.

[Figure 1. Decision tree structure]

Figure 2 presents the Gain, Cover, and Frequency values for the top 20 variables identified by the XGBoost method, ordered by their Gain measure values. While Figure 1 provided a snapshot of feature interactions from a single decision tree, Figure 2 offers a more comprehensive view of feature importance derived from the ensemble of multiple trees. Overall, the feature importance distribution highlights a mix of clinical, socioeconomic, and demographic factors driving ED visits in this population. The most influential predictor, as shown by the highest gain value, is having a past 3-month ED visit, which suggests that previous ED visits are highly indicative of future visits. This is

followed by the use of aspart insulin and glargine insulin, both showing significant gain and cover values, indicating their strong influence on model accuracy and the number of patients affected.

Socioeconomic factors, such as the SVI and ADI, also play key roles in predicting ED visits, as reflected in their relatively high gain, cover, and frequency values. This highlights the importance of SDoH in understanding healthcare utilization. Age is another prominent predictor, emphasizing the role of demographics in ED visit risk.

[Figure 2. Feature importance from the XGBoost model]

Among clinical factors, high blood glucose levels, diabetic HbA1c values, and specific ICD-10 codes, particularly E11.9 (Type 2 diabetes without complications) and E11.22 (Type 2 diabetes with chronic kidney disease), also contribute notably to the prediction. The model also flags the influence of race, with Black or African American patients having a significant impact on the prediction, further underscoring healthcare disparities.

The individual-level SDoH appear less in the most informative feature list. It is likely due to the prevalence of missing values. Interestingly, the home safety issues predictor was listed in the top 20 list, suggesting that while these factors were recorded for fewer patients, represented by the low Frequency, they are important when they do appear in the data.

Figure 3 reports the performance comparison across the logistic regression, decision trees, and XGBoost models for predicting 3-month ED visits post-hospitalization. The AUC graphs show that XGBoost outperforms the other models with an AUC of 0.846, indicating excellent classification performance, followed by decision trees (0.810) and logistic regression (0.816). This suggests that XGBoost has a stronger ability to discriminate between patients who will and will not visit the ED. Sensitivity is higher in XGBoost (0.296) compared to logistic regression (0.243) and decision trees (0.240). Although all models struggle with this metric, the overall prevalence of the positive cases was around 6%, which partially explain the challenge in prediction. Precision is also higher for XGBoost (0.420) than for logistic regression (0.331) and decision trees (0.335), suggesting XGBoost

is better at reducing false positives, an important factor in healthcare settings. Lastly, Specificity is excellent across all models but highest in XGBoost (0.972), closely followed by decision trees (0.970) and logistic regression (0.967). This means all models are effective at minimizing false negatives, with XGBoost performing marginally better. In summary, XGBoost consistently outperforms both logistic regression and decision trees in terms of AUC, sensitivity, precision, and specificity, making it the strongest model for predicting 3-month ED visits among patients with diabetic concerns post-hospitalization.

[Figure 3. Model performance comparison]

4. Discussion

The strong predictor, the occurrence of past 3-month ED visits, highlights a behavioral component of ED utilization, potential underlying health issues, and the important role of SDoH. This suggests that some patients may habitually use the ED for non-emergency needs, such as refilling prescriptions or addressing routine health concerns, due to potential challenges in accessing primary care or outpatient services. It may also reflect underlying health instability or insufficient outpatient care, where frequent ED visits may indicate unmanaged or poorly controlled chronic conditions that are often exacerbated by socioeconomic factors.

Insulin glargine is a long-acting insulin that provides a steady, basal level of insulin, often prescribed to manage chronically elevated blood glucose levels seen in both Types 1 and 2 diabetes mellitus.[70] In contrast, insulin aspart is a rapid-acting insulin used to control blood glucose spikes after meals.[71] The significance of these two types of insulin as predictors suggests that patients using either one may have more complex or less stable diabetes, making them more prone to acute episodes that could lead to ED visits. The results may also suggest an increased risk of medication management (e.g., nonadherence) associated with these two types of insulins that lead to frequent ED visits.[72-74] More research is needed to understand the implications of these medications on ED visits, as it could represent multiple mechanisms for its contribution such as disease severity,

medication side effects (hypoglycemia), and medication nonadherence related to cost, storage, or dosing frequency.

The use of statins, such as atorvastatin and lovastatin, appeared as significant predictors in the analysis, reflecting their common prescription among patients with diabetes who are at an inherently higher risk for cardiovascular disease.[75,76] The American Diabetes Association standards of care for diabetes state that statin therapy should be initiated in those with diabetes when one or more atherosclerotic cardiovascular disease risk factors is present such as hypertension and tobacco use, to lower the risk of stroke and heart attack.[77] Patients with these risk factors or underlying cardiovascular disease and on statin therapy often have more complex overall health that makes them more likely to experience acute events requiring emergency care.[78,79] These factors compound the complexity of diabetes management, underscoring the intricate intersection of diabetes and cardiovascular health.[80,81]

Socioeconomic factors, represented by the SVI and the ADI, emphasize the impact of SDoH, showing how economic and community-level factors can influence healthcare utilization and potentially overall health, even more so than many traditional clinical variables. Patients from vulnerable or deprived areas may face barriers to accessing regular diabetes care, leading to condition exacerbations and subsequent ED visits. The LILA indicator further underscores the importance of food security in managing diabetes effectively. Notably, "home safety issue" was the only individual-level SDoH in the top 20 predictors, suggesting that environmental stressors like unsafe living conditions can interfere with their ability to manage diabetes, such as lack of a safe space for medication storage, or contribute to psychological stress leading to poor diabetes control and ED visits. Despite the insights into personal living conditions directly influencing a patient’s risk of ED visits, high rates of missing data in individual-level SDoH limited their visibility in the model outcomes.

The presence of specific ICD-10 codes during hospitalization, such as abnormal glucose and

type 2 diabetes with chronic kidney disease, highlighted the range of key diabetes-related conditions that influence ED visit risk. These codes suggest that both patients with underlying glucose abnormalities and those with established complications, such as chronic kidney disease, are at significant risk for acute emergency intervention. Moreover, mental, behavioral disorders, and gastroparesis were identified as important predictors for ED visits.

The significant association of race, particularly being Black or African American, with ED visit risk points to the role of racial disparities in healthcare access and outcomes, which may be influenced by socioeconomic factors, healthcare access, and systemic inequities. Additionally, the presence of histories of respiratory or cardiovascular surgical procedures as predictors highlights the clinical complexity of these patients. Individuals with such medical histories may have underlying health vulnerabilities or interactions with diabetes that increase their likelihood of experiencing complications.

In evaluating models, a trade-off between accuracy and interpretability is observed. Logistic regression, offering the highest interpretability, may lack the capacity to capture complex non-linearity and interactions. Decision trees provide a middle ground, offering more nuanced decision paths while maintaining good interpretability. XGBoost delivers superior accuracy by modeling complex relationships, though its interpretability is lower due to the ensemble nature of the model.

This study has limitations. The high rate of missing data, particularly in individual-level self-reported SDoH variables, may have affected the model’s performance. While we addressed this by introducing a “not measured” category to control for potential bias, the lack of comprehensive SDoH data may have reduced the model's ability to fully capture the influence of these factors. Additionally, this study assumed that individual-level SDoH factors remain relatively constant over time for a patient, which may oversimplify their potential variability. Nonetheless, we leveraged the institution’s initiative to maximize the use of available individual-level SDoH data. It is possible that the study did not capture patients who sought ED care outside the UAB Medical Center. However, we focused on

patients who recently experienced hospitalization within the large healthcare system to reduce bias from those who might seek ED care elsewhere, enabling a system-wide analysis. Future research should expand to include a larger care network to provide broader public health insights. Lastly, the generalizability of the findings may be limited, as the study was conducted in a single academic health system in the southern United States. However, this study setting well represents underserved, diverse populations with high burdens of disease and significant socioeconomic and healthcare access challenges.

5. Conclusion

This study developed predictive models for 3-month post-hospitalization ED visits in patients with diabetes, integrating both clinical and SDoH data. Medication prescriptions, area-level SDoH, such as SVI and ADI, and past ED visit patterns, emerged as strong predictors. The selection of the appropriate model depends on the clinical context: decision trees may be preferred when interpretability is crucial for decision-making, balancing interpretability with accuracy, while XGBoost may be more suitable when accuracy is prioritized despite lower interpretability. Future work should address missing data in individual-level SDoH and focus on implementing the most suitable model to assess its applicability and usability in real-world settings.

Figure Legends

Figure 1. Decision tree structure; SVI: Social vulnerability index; BMI: Body mass index; ICD-10, F: Mental, behavioral, and neurodevelopmental disorders; ICD-10, K31.84: Gastroparesis.

Figure 2. Feature importance from the XGBoost model; SVI: Social vulnerability index; ADI: Area deprivation index; ICD-10, R73.9: Hyperglycemia; ICD-10, E11.9: Type 2 diabetes without complications; ICD-10, E11.22: Type 2 diabetes with diabetic chronic kidney disease; ICD10, F: Mental, behavioral, and neurodevelopmental disorders; LILA: Low income, low access; ICD-10, O: Pregnancy, childbirth and the puerperium; ICD-10, E11.42: Type 2 diabetes with diabetic polyneuropathy.

Figure 3. Model performance comparison; AUC: Area under the curve.

Funding Statement:

This work was supported by 2023-2024 Interdisciplinary Collaborative Research Grant in the School

of Health Professions at the University of Alabama at Birmingham.

Conflict of Interest:

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics Approval Statement:

The University of Alabama at Birmingham Office of the Institutional Review Board for Human Use (UAB Office of IRB) has waived ethics approval. Also, the UAB Office of IRB has waived informed consent to participate. The UAB Office of IRB determined this project is not subject to FDA regulations and is not Human Subjects Research (De-identified data).

Acknowledgements:

The authors thank the University of Alabama at Birmingham, Department of Biomedical Informatics and Data Science for their data extraction and transformation expertise.

References

1. W.H.O. Diabetes. Available from: https://www.who.int/health-topics/diabetes#tab=tab_1.
2. Hudon C, Courteau J, Krieg C, et al. Factors associated with chronic frequent emergency department utilization in a population with diabetes living in metropolitan areas: a population-based retrospective cohort study. BMC Health Serv Res 2017;17(1):525, doi:10.1186/s12913-017-2453-3
3. Shamoon H, Duffy H, Fleischer N, et al. The Effect of Intensive Treatment of Diabetes on the Development and Progression of Long-Term Complications in Insulin-Dependent Diabetes-Mellitus. New Engl J Med 1993;329(14):977-986, doi:DOI 10.1056/nejm199309303291401
4. Drincic A, Pfeffer E, Luo J, et al. The effect of diabetes case management and Diabetes Resource Nurse program on readmissions of patients with diabetes mellitus. Journal of clinical & translational endocrinology 2017;8(29-34
5. McCants KM, Reid KB, Williams I, et al. The impact of case management on reducing readmission for patients diagnosed with heart failure and diabetes. Professional case management 2019;24(4):177-193
6. Norris SL, Nichols PJ, Caspersen CJ, et al. The effectiveness of disease and case management for people with diabetes: a systematic review. American journal of preventive medicine 2002;22(4):15-38
7. Rubin DJ, Shah AA. Predicting and preventing acute care re-utilization by patients with diabetes. Current Diabetes Reports 2021;21(1-13
8. Black RL, Duval C. Diabetes discharge planning and transitions of care: a focused review. Current diabetes reviews 2019;15(2):111-117
9. Garnica P. Transition of care for patients with diabetes. Current diabetes reviews 2017;13(3):263-279
10. Demidowich AP, Batty K, Zilbermint M. Instituting a successful discharge plan for patients with

type 2 diabetes: challenges and solutions. Diabetes Spectrum 2022;35(4):440-451

11. Wang J, Geiss LS, Williams DE, et al. Trends in emergency department visit rates for hypoglycemia and hyperglycemic crisis among adults with diabetes, United States, 2006-2011. PloS one 2015;10(8):e0134917
12. Association AD. Economic costs of diabetes in the US in 2017. Diabetes care 2018;41(5):917-928
13. Karter AJ, Warton EM, Lipska KJ, et al. Development and validation of a tool to identify patients with type 2 diabetes at high risk of hypoglycemia-related emergency department or hospital use. JAMA internal medicine 2017;177(10):1461-1470
14. Shin J, Kim J, Lee C, et al. Development of various diabetes prediction models using machine learning techniques. Diabetes & Metabolism Journal 2022;46(4):650-657
15. Chiu YM, Courteau J, Dufour I, et al. Machine learning to improve frequent emergency department use prediction: a retrospective cohort study. Scientific Reports 2023;13(1):1981
16. O'Connor PJ, Sperl-Hillen JM, Rush WA, et al. Impact of electronic health record clinical decision support on diabetes care: a randomized trial. Ann Fam Med 2011;9(1):12-21, doi:10.1370/afm.1196
17. Wang L, Li X, Wang Z, et al. Trends in prevalence of diabetes and control of risk factors in diabetes among US adults, 1999-2018. Jama 2021;326(8):704-716
18. Jeffery R, Iserman E, Haynes RB, et al. Can computerized clinical decision support systems improve diabetes management? A systematic review and meta-analysis. Diabetic Med 2013;30(6):739-745, doi:10.1111/dme.12087
19. Cleveringa FG, Gorter KJ, van den Donk M, et al. Computerized decision support systems in primary care for type 2 diabetes patients only improve patients' outcomes when combined with feedback on performance and case management: a systematic review. Diabetes Technol Ther 2013;15(2):180-92, doi:10.1089/dia.2012.0201
20. Li X, Yu S, Zhang Z, et al. Predictive modeling of hypoglycemia for clinical decision support in evaluating outpatients with diabetes mellitus. Current Medical Research and Opinion 2019;35(11):1885-1891
21. Berkowitz SA, Orr CJ. Three Lessons About Diabetes and the Social Determinants of Health. Diabetes Care 2023;dci230045
22. Lee S-Y, Hayes LW, Ozaydin B, et al. Integrating Social Determinants of Health in Machine Learning–Driven Decision Support for Diabetes Case Management: Protocol for a Sequential Mixed Methods Study. JMIR Research Protocols 2024;13(1):e56049
23. Basu S, Narayanaswamy R. A prediction model for uncontrolled type 2 diabetes mellitus incorporating area-level social determinants of health. Medical care 2019;57(8):592-600
24. Chi W, Andreyeva E, Zhang Y, et al. Neighborhood-level Social Determinants of Health Improve Prediction of Preventable Hospitalization and Emergency Department Visits Beyond Claims History. Population Health Management 2021;24(6):701-709
25. Carvalho DV, Pereira EM, Cardoso JS. Machine learning interpretability: A survey on methods and metrics. Electronics 2019;8(8):832
26. Gilpin LH, Bau D, Yuan BZ, et al. Explaining explanations: An overview of interpretability of machine learning. IEEE: 2018.

27. Organization WH. International Statistical Classification of Diseases and related health problems: Alphabetical index. World Health Organization: 2004.

28. Saxena FE, Bierman AS, Glazier RH, et al. Association of early physician follow-up with readmission among patients hospitalized for acute myocardial infarction, congestive heart failure, or chronic obstructive pulmonary disease. JAMA Network Open 2022;5(7):e2222056-e2222056

29. Davison BA, Metra M, Senger S, et al. Patient journey after admission for acute heart failure: length of stay, 30-day readmission and 90-day mortality. European journal of heart failure 2016;18(8):1041-1050

30. Culler SD, Kugelmass AD, Cohen DJ, et al. Understanding Readmissions in Medicare Beneficiaries during the 90-Day Follow-up Period of an Acute Myocardial Infarction Admission. Journal of the American Heart Association 2019;8(21):e013513

31. Nicklett EJ. Socioeconomic status and race/ethnicity independently predict health decline among older diabetics. BMC Public Health 2011;11(1-8

32. Ramezankhani A, Azizi F, Hadaegh F. Associations of marital status with diabetes, hypertension, cardiovascular disease and all-cause mortality: a long term follow-up study. PloS one 2019;14(4):e0215593

33. Yan JW, Gushulak KM, Columbus MP, et al. Risk factors for recurrent emergency department visits for hyperglycemia in patients with diabetes mellitus. International journal of emergency medicine 2017;10(1-8

34. Selya A, Anshutz D, Griese E, et al. Predicting unplanned medical visits among patients with diabetes: translation from machine learning to clinical implementation. BMC Medical Informatics and Decision Making 2021;21(1-11

35. (WHO) WHO. International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10). 2016. Available from: https://www.who.int/classifications/icd/en/.

36. Association AM. Current Procedural Terminology: CPT 2024. American Medical Association: Chicago, IL; 2023.

37. Committee TIE. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. Diabetes care 2009;32(7):1327

38. Jeon JY, Ko S-H, Kwon H-S, et al. Prevalence of diabetes and prediabetes according to fasting plasma glucose and HbA1c. Diabetes & metabolism journal 2013;37(5):349

39. Borrell LN, Samuel L. Body mass index categories and mortality risk in US adults: the effect of overweight and obesity on advancing death. American journal of public health 2014;104(3):512-519

40. Nichols GA, Hillier TA, Brown JB. Normal fasting plasma glucose and risk of type 2 diabetes diagnosis. The American journal of medicine 2008;121(6):519-524

41. Brambilla P, La Valle E, Falbo R, et al. Normal fasting plasma glucose and risk of type 2 diabetes. Diabetes care 2011;34(6):1372-1374

42. Webster AC, Nagler EV, Morton RL, et al. Chronic kidney disease. The lancet 2017;389(10075):1238-1252

43. ElSayed NA, Aleppo G, Aroda VR, et al. 9. Pharmacologic approaches to glycemic treatment: standards of care in diabetes—2023. Diabetes care 2023;46(Supplement_1):S140-S157

44. Walker RJ, Williams JS, Egede LE. Influence of race, ethnicity and social determinants of health on diabetes outcomes. The American journal of the medical sciences 2016;351(4):366-373
45. Hill-Briggs F, Adler NE, Berkowitz SA, et al. Social determinants of health and diabetes: a scientific review. Diabetes care 2020;44(1):258
46. Walker RJ, Smalls BL, Campbell JA, et al. Impact of social determinants of health on outcomes for type 2 diabetes: a systematic review. Endocrine 2014;47(29-48
47. Knighton AJ, Savitz L, Belnap T, et al. Introduction of an area deprivation index measuring patient socioeconomic status in an integrated health system: implications for population health. EGEMs 2016;4(3):
48. Messer LC, Laraia BA, Kaufman JS, et al. The development of a standardized neighborhood deprivation index. Journal of Urban Health 2006;83(1041-1062
49. Spielman SE, Tuccillo J, Folch DC, et al. Evaluating social vulnerability indicators: criteria and their application to the Social Vulnerability Index. Natural hazards 2020;100(417-436
50. Tate E. Social vulnerability indices: a comparative assessment using uncertainty and sensitivity analysis. Natural Hazards 2012;63(325-347
51. Rhone A, Ver Ploeg M, Williams R, et al. Understanding low-income and low-access census tracts across the nation: subnational and subpopulation estimates of access to healthy food. 2019;
52. DeVoe JE, Baez A, Angier H, et al. Insurance+ access≠ health care: typology of barriers to health care access for low-income families. The Annals of Family Medicine 2007;5(6):511-518
53. Bailey J, Baker E, Schechter MS, et al. Food insecurity screening and local food access: contributions to nutritional outcomes among children and adults with cystic fibrosis in the United States. Journal of Cystic Fibrosis 2024;23(3):524-531
54. Chenarides L, Cho C, Nayga Jr RM, et al. Dollar stores and food deserts. Applied Geography 2021;134(102497
55. LaForge K, Gold R, Cottrell E, et al. How 6 organizations developed tools and processes for social determinants of health screening in primary care: an overview. The Journal of ambulatory care management 2018;41(1):2-14
56. Howell CR, Bradley H, Zhang L, et al. Real-world integration of the protocol for responding to and assessing patients' assets, risks, and experiences tool to assess social determinants of health in the electronic medical record at an academic medical center. Digital health 2023;9(20552076231176652
57. Lin W-C, Tsai C-F. Missing value imputation: a review and analysis of the literature (2006–2017). Artificial Intelligence Review 2020;53(1487-1509
58. Mansoor H, Ali S, Alam S, et al. Impact of missing data imputation on the fairness and accuracy of graph node classifiers. IEEE: 2022.
59. Henry AJ, Hevelone ND, Lipsitz S, et al. Comparative methods for handling missing data in large databases. Journal of vascular surgery 2013;58(5):1353-1359. e6
60. Hughes RA, Heron J, Sterne JA, et al. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. International journal of epidemiology 2019;48(4):1294-1304
61. Hosmer DW, Lemeshow S, Sturdivant RX, et al. Applied logistic regression. Wiley: Hoboken,

New Jersey; 2013.

62. Boateng EYA, D. A. A Review of the Logistic Regression Model with Emphasis on Medical Research. Journal of Data Analysis and Information Processing 2019;7(190 - 207
63. Omer DM, A. B. Modelling logistic regression using multivariable fractional polynomials. Imperial Journal of Interdisciplinary Research 2017;3(11):8-16
64. Bache-Mathiesen LK, Andersen TE, Dalen-Lorentsen T, et al. Not straightforward: modelling non-linearity in training load and injury research. BMJ Open Sport Exerc Med 2021;7(3):e001119, doi:10.1136/bmjsem-2021-001119
65. Hastie T, Friedman JH, Tibshirani R. The elements of statistical learning : data mining, inference, and prediction. Springer: New York ; London; 2009.
66. Chen TG, C. . A scalable tree boosting system. Association for Computing Machinery: San Francisco, CA; 2016.
67. Liu J, Wu J, Liu S, et al. Predicting mortality of patients with acute kidney injury in the ICU using XGBoost model. PLoS One 2021;16(2):e0246306, doi:10.1371/journal.pone.0246306
68. Ogunleye A, Wang QG. XGBoost Model for Chronic Kidney Disease Diagnosis. Ieee Acm T Comput Bi 2020;17(6):2131-2140, doi:10.1109/Tcbb.2019.2911071
69. Asselman A, Khaldi M, Aammou S. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interact Learn Envir 2023;31(6):3360-3379, doi:10.1080/10494820.2021.1928235
70. Bolli GB, Owens DR. Insulin glargine. The Lancet 2000;356(9228):443-445
71. Mudaliar SR, Lindberg FA, Joyce M, et al. Insulin aspart (B28 asp-insulin): a fast-acting analog of human insulin: absorption kinetics and action profile compared with regular human insulin in healthy nondiabetic subjects. Diabetes care 1999;22(9):1501-1506
72. Dankers M, Van den Berk-Bulsink M, van Dalfsen-Slingerland M, et al. Non-adherence to guideline recommendations for insulins: a qualitative study amongst primary care practitioners. BMC primary care 2022;23(1):150
73. Haycox A. Insulin aspart: an evidence-based medicine review. Clinical drug investigation 2004;24(695-717
74. Evans M, Wilkinson M, Giannpolou A. Fast-acting insulin aspart: the rationale for a new mealtime insulin. Diabetes Therapy 2019;10(1793-1800
75. Stern MP. Diabetes and cardiovascular disease: the “common soil” hypothesis. Diabetes 1995;44(4):369-374
76. Grundy SM, Benjamin IJ, Burke GL, et al. Diabetes and cardiovascular disease: a statement for healthcare professionals from the American Heart Association. Circulation 1999;100(10):1134-1146
77. Committee ADAPP. 10. Cardiovascular Disease and Risk Management: Standards of Care in Diabetes—2024. Diabetes Care 2023;47(Supplement_1):S179-S218, doi:10.2337/dc24-S010
78. Mansi IA, Chansard M, Lingvay I, et al. Association of statin therapy initiation with diabetes progression: a retrospective matched-cohort study. JAMA internal medicine 2021;181(12):1562-1574
79. Grundy SM, Stone NJ, Blumenthal RS, et al. High-intensity statins benefit high-risk patients: why and how to do better. Elsevier: 2021.

80. Casula M, Mozzanica F, Scotti L, et al. Statin use and risk of new-onset diabetes: A meta-analysis of observational studies. Nutrition, metabolism and cardiovascular diseases 2017;27(5):396-406

81. Crandall JP, Mather K, Rajpathak SN, et al. Statin use and risk of developing diabetes: results from the Diabetes Prevention Program. BMJ Open Diabetes Research and Care 2017;5(1):e000438

Preprint JMIR Publications

Table 1. Clinical data items and coding methods; DM: Diabetes mellitus.

Data category	Description and coding	Reference
Diabetes-related comorbidities related to hospitalization		
DM due to underlying condition (E08)	Binary variable indicating whether each hospitalization is associated with a specific diagnosis code as either the primary or secondary diagnosis.	ICD-10 Code[35]
Drug or chemical induced DM (E09)		
Type 1 DM (E10)		
Type 2 DM (E11)		
Other specified DM (E13)		
Elevated blood glucose level (R73)		
Hypoglycemia (E16.2)		
Foot ulcer (L97.509)		
Gastroparesis (K31.84)		
Surgical procedures during hospitalization		
General surgical procedure	Binary variable indicating whether the patient underwent a specific type of surgery during their hospitalization.	CPT Code[36]
Integumentary system		
Musculoskeletal system		
Respiratory system		
Cardiovascular system		
Hemic and lymphatic systems		
Mediastinum and diaphragm		
Digestive system		
Urinary system		
Male genital system		
Reproductive system		
Intersex surgery		
Female genital system		
Maternity care and delivery		
Endocrine system		
Nervous system		
Eye and ocular adnexa		
Auditory system		
Operating microscope procedure		
Laboratory test/measure		
Hemoglobin A1c (HbA1c) (%)	Measured for 77.7% patient visits. Categorical variable coded as normal (5.6 and lower), prediabetic (5.7 ~ 6.4), diabetic (6.5 and higher), and not measured.	International Expert Committee (2009)[37], Jeon et al. (2013)[38]
Body mass index (BMI) (kg/m^2)	Measured for 97.3% patient visits. Categorical variable coded as underweight (18.4 and lower), normal	Borrell and Samuel (2014)[39]

	(18.5 ~ 24.9), overweight (25.0 ~ 29.9), obesity (30.0 and higher), and not measured.	
Fasting blood glucose (mg/dL)	Measured for 63.7% patient visits. Categorical variable coded as low (69.9 and below), normal (70.0 ~ 100.0), elevated (100.1 ~ 125.0), high (125.1 and higher), and not measured.	Nichols et al (2008)[40], Brambilla et al. (2011)[41]
Estimated glomerular filtration rate (eGFR) (mL/min/1.73m^2)	Measured for 95.0% patient visits. Categorical variable coded as normal (90 and higher), decreased (30.0 ~ 89.9), severely decreased (15.0 ~ 29.9), failure (14.9 and lower), and not measured.	Webster et al. (2017)[42]
Medication		
Insulin, regular	Binary variable indicating whether the medication has been prescribed to the patient within the last year from their hospitalization.	ElSayed et al. (2023)[43]
Insulin, lispro		
Insulin, detemir		
Insulin, isophane		
Insulin, degludec		
Insulin, with fluid		
Insulin, other		
Statin, atorva		
Statin, simva		
Statin, prava		
Statin, rosuva		
Statin, lova		
Statin, other		

Table 2. Univariate and bivariate analysis and inferential statistical analysis results; DM: Diabetes mellitus.

	All	3-Month ED Visit – No	3-Month ED Visit – Yes	*t*-test / Chi-square
Count (%)	162,063 (100)	151,938 (93.8)	10,125 (6.2)	
Demographics				
Age, mean (SD)	56.1 (17.9)	56.0 (18.0)	57.5 (15.2)	***t* = -10.1, *p*<.001**
Race/ethnicity, N (% of total)				**χ^2=1426.0, *p*<.001**
White	88,794 (54.8)	84,885 (55.9)	3,909 (38.6)	
Black	64,777 (40.0)	59,038 (38.9)	5,739 (56.7)	
Asian	3,165 (2.0)	2,866 (1.9)	299 (3.0)	
Hispanic	2,717 (1.7)	2,598 (1.7)	128 (1.3)	
American Indian or Alaska Native	218 (0.1)	206 (0.1)	12 (0.1)	
Native Hawaiian or Other Pacific Islander	17 (0.01)	17 (0.01)	0 (0)	
Other	323 (0.2)	319 (0.2)	4 (0.04)	
Gender, N (% of total)				**χ^2=114.9, *p*<.001**
Female	78,377 (48.4)	72,958 (48.0)	5,419 (53.5)	
Male	83,686 (51.6)	78,980 (52.0)	4,706 (46.5)	
Healthcare Utilization				
Past 3-month ED Visit Frequency, mean (SD)	0.36 (1.4)	0.32 (1.3)	0.98 (2.4)	***t* = -27.3, *p*<.001**
Social Determinants of Health				
Payer type, N (% of total)				**χ^2=183.4, *p*<.001**
Commercial	55,895 (34.5)	52,733 (34.7)	3,162 (31.2)	
Medicaid	52,254 (32.2)	48,421 (31.9)	3,833 (37.9)	
Medicare	40,202 (24.8)	37,742 (24.8)	2,460 (24.3)	
Uninsured/Self-pay	13,712 (8.5)	13,042 (8.6)	670 (6.6)	
ADI, mean (SD)	71.7 (23.1)	71.4 (23.1)	76.4 (21.5)	***t* = -21.2, *p*<.001**
SVI, mean (SD)	0.57 (0.29)	0.57 (0.29)	0.64 (0.28)	***t* = -22.9, *p*<.001**
LILA, N (% of total)	34,387 (21.1)	31,234 (20.6)	3,153 (31.1)	**χ^2=462.5, *p*<.001**
Unmet needs – Care, N (% of total)	3,671 (2.3)	3,297 (2.2)	374 (3.7)	**χ^2=98.9, *p*<.001**
Unmet needs – Medication, N (% of total)	3,981 (2.5)	3,530 (2.3)	451 (4.5)	**χ^2=179.0, *p*<.001**

Unmet needs – Utilities, N (% of total)	1,602 (1.0)	1,366 (0.90)	236 (2.3)	**χ^2=197.4, *p*<.001**
Unmet needs – Food, N (% of total)	377 (0.23)	314 (0.21)	63 (0.62)	**χ^2=68.9, *p*<.001**
Housing Instability				**χ^2=1729.7, *p*<.001**
No	788 (0.5)	724 (0.5)	64 (0.6)	
Yes	27,435 (16.9)	24,207 (15.9)	3,228 (31.9)	
Refuse	121 (0.1)	113 (0.1)	8 (0.1)	
Not measured	133,719 (82.5)	126,894 (83.5)	6,825 (67.4)	
Stress				**χ^2=1790.9, *p*<.001**
None	6,092 (3.6)	5,366 (3.5)	726 (7.2)	
Little	7,128 (4.4)	6,383 (4.2)	745 (7.4)	
Somewhat	6,457 (4.0)	5,655 (3.7)	802 (7.9)	
Quite	3,202 (2.0)	2,736 (1.8)	466 (4.6)	
Much	1,901 (1.2)	1,659 (1.1)	242 (2.4)	
Refuse	1,332 (0.8)	1,199 (0.8)	133 (1.3)	
Not measured	135,951 (87.6)	128,940 (84.9)	7,011 (69.2)	
Transportation Issue				**χ^2=1720.5, *p*<.001**
No issue	22,612 (14.0)	20,000 (13.2)	2,612 (25.8)	
Issue with medical appointment/getting medications	1,995 (1.2)	1,770 (1.2)	225 (2.2)	
Issue with non-medical meetings, work, or necessities	1,720 (1.1)	1,481 (1.0)	239 (2.4)	
Other issue	967 (0.6)	835 (0.5)	132 (1.3)	
Not measured				
Education				**χ^2=1732.7, *p*<.001**
Less than high school	2,739 (1.7)	2,377 (1.6)	362 (3.6)	
High school	11,043 (6.8)	9,649 (6.4)	1,394 (13.8)	
More than high school	7,405 (4.6)	6,520 (4.3)	885 (8.7)	
Refuse	4,302 (2.7)	3,904 (2.6)	398 (3.9)	
Not measured	136,574 (84.2)	129,488 (85.1)	7,086 (70.0)	
Unmet household needs past year (multiple choice)				

No	14,980 (9.24)	13,341 (8.8)	1,639 (16.2)	**χ^2=619.9, *p*<.001**
Care	3,671 (2.3)	3,297 (2.2)	374 (3.7)	**χ^2=98.9, *p*<.001**
Medication	3,981 (2.5)	3,530 (2.3)	451 (4.5)	**χ^2=179.0, *p*<.001**
Utilities	1,602 (1.0)	1,366 (0.9)	236 (2.3)	**χ^2=197.4, *p*<.001**
Food	377 (0.2)	314 (0.2)	63 (0.6)	**χ^2=68.9, *p*<.001**
Cloth	80 (0.1)	64 (0.1)	16 (0.2)	**χ^2=23.5, *p*<.001**
Phone	208 (0.1)	197 (0.1)	11 (0.1)	χ^2=0.2, *p*=.67
Childcare	78 (0.1)	69 (0.1)	9 (0.1)	χ^2=2.9, *p=.09*
Not measured	140,701 (86.8)	133,002 (87.5)	7,699 (76.0)	**χ^2=1095.5, *p*<.001**
Work status/employment				**χ^2=1954.1, *p*<.001**
Unemployed	12,421 (7.7)	10,960 (7.2)	1,461 (14.4)	
Employed	15,631 (9.6)	13,880 (9.1)	1,751 (17.3)	
Retire	3,752 (2.3)	3,386 (2.2)	366 (3.6)	
Disable	5,752 (3.5)	4,834 (3.2)	918 (9.1)	
Part time	939 (0.6)	829 (0.5)	110 (1.1)	
Refuse	956 (0.6)	849 (0.6)	107 (1.1)	
Not measured	122,612 (75.7)	117,200 (77.1)	5,412 (53.4)	
Physical and emotional safety				**χ^2=1838.3, *p*<.001**
No	830 (0.5)	715 (0.5)	115 (1.1)	
Unsure	301 (0.2)	261 (0.2)	40 (0.4)	
Yes	23,151 (14.3)	20,303 (13.4)	2,848 (28.1)	
Refuse	2,586 (1.6)	2,417 (1.6)	169 (1.7)	
Not measured				
Diabetes-Related Comorbidities				
DM due to underlying conditions (E08)	520 (0.32)	467 (0.31)	53 (0.52)	**χ^2=13.2, *p*<.001**
Drug or chemical induced DM (E09)	371 (0.23)	347 (0.23)	24 (0.24)	χ^2=0.01, *p*=.94
Type 1 DM (E10)	3,579 (2.2)	2,876 (1.9)	703 (6.9)	**χ^2=1118.7, *p*<.001**
Type 2 DM with ketoacidosis (E11.1)	755 (0.47)	688 (0.45)	67 (0.66)	**χ^2=8.5, *p*<.01**
Type 2 DM with diabetic nephropathy (E11.21)	918 (0.57)	759 (0.50)	159 (1.6)	**χ^2=191.4, *p*<.001**
Type 2 DM with diabetic chronic kidney disease (E11.22)	17,947 (11.1)	15,512 (10.2)	2,435 (24.0)	**χ^2=1844.9, *p*<.001**

Type 2 DM with ophthalmic complications (E11.3)	2,398 (1.5)	1,964 (1.3)	434 (4.3)	**χ^2=581.6, *p*<.001**
Type 2 DM with diabetic neuropathy (E11.40)	7,231 (4.5)	6,060 (4.0)	1,171 (11.6)	**χ^2=1276.7, *p*<.001**
Type 2 DM with diabetic polyneuropathy (E11.42)	4,985 (3.1)	4,080 (2.7)	905 (8.9)	**χ^2=1242.8, *p*<.001**
Type 2 DM with diabetic autonomic (poly)neuropathy (E11.43)	1,374 (0.85)	1,016 (0.67)	358 (3.5)	**χ^2=1924.8, *p*<.001**
Type 2 DM with circulatory complications (E11.5)	7,013 (4.3)	6,029 (4.0)	984 (9.7)	**χ^2=756.8, *p*<.001**
Type 2 DM with diabetic arthropathy (E11.61)	243 (0.15)	312 (0.14)	30 (0.30)	**χ^2=14.4, *p*<.001**
Type 2 DM with skin complications (E11.62)	1,792 (1.1)	1,461 (0.96)	331 (3.3)	**χ^2=460.1, *p*<.001**
Type 2 DM with hypoglycemia (E11.64)	4,311 (2.7)	3,766 (2.5)	545 (5.4)	**χ^2=308.1, *p*<.001**
Type 2 DM with hyperglycemia (E11.65)	21,802 (13.5)	19,337 (12.7)	2,465 (24.3)	**χ^2=1099.6, *p*<.001**
Type 2 DM with other specified complication (E11.69)	2,241 (1.4)	1,864 (1.2)	377 (3.7)	**χ^2=432.1, *p*<.001**
Type 2 DM without complications (E11.9)	15,767 (9.7)	14,281 (9.4)	1,486 (14.7)	**χ^2=300.4, *p*<.001**
Other specified DM (E13)	237 (0.15)	198 (0.13)	39 (0.39)	**χ^2=40.5, *p*<.001**
Hyperglycemia (R73.9)	10,500 (6.5)	10,452 (6.9)	48 (0.47)	**χ^2=641.6, *p*<.001**
Hypoglycemia (E16.2)	2,453 (1.5)	2,418 (1.6)	35 (0.35)	**χ^2=98.0, *p*<.001**
Gastroparesis (K31.84)	2,296 (1.4)	1,747 (1.1)	549 (5.4)	**χ^2=1237.5, *p*<.001**
Surgical Procedures				
General	213 (0.13)	200 (0.13)	13 (0.13)	χ^2=0.0, *p*=1
Integumentary system	13,067 (8.1)	12,406 (8.1)	661 (6.5)	**χ^2=34.1, *p*<.001**
Musculoskeletal system	20,484 (12.6)	19,543 (12.9)	941 (9.3)	**χ^2=109.2, *p*<.001**
Respiratory system	14,705 (9.1)	14,224 (9.4)	481 (4.8)	**χ^2=244.1, *p*<.001**
Cardiovascular system	44,299 (27.3)	42,190 (27.8)	2,109 (20.8)	**χ^2=229.7, *p*<.001**
Hemic and lymphatic systems	2,791 (1.7)	2,715 (1.8)	76 (0.75)	**χ^2=59.6, *p*<.001**
Digestive system	21,641 (13.4)	20,425 (13.4)	1,216 (12.0)	**χ^2=16.7, *p*<.001**
Urinary system	4,999 (3.1)	4,725 (3.1)	274 (2.7)	**χ^2=5.0, *p*<.05**
Male genital system	287 (0.18)	276 (0.18)	11 (0.11)	χ^2=2.5, *p*=.12
Female genital system	2,156 (1.3)	2,095 (1.4)	61 (0.60)	**χ^2=43.0, *p*<.001**

Maternity care and delivery	5,259 (3.2)	5,204 (3.4)	35 (0.54)	**χ^2=250.2, *p*<.001**
Endocrine system	335 (0.21)	323 (0.21)	12 (0.12)	χ^2=3.6, *p*=.06
Nervous system	17,246 (10.6)	16,616 (10.9)	630 (6.2)	**χ^2=221.3, *p*<.001**
Auditory system	156 (0.10)	148 (0.10)	8 (0.08)	χ^2=0.2, *p*=.68
Operating microscope procedures	672 (0.41)	655 (0.43)	17 (0.17)	**χ^2=15.3, *p*<.001**
BMI				**χ^2=221.5, *p*<.001**
Normal	40,768 (25.2)	38,532 (25.4)	2,236 (22.1)	
Underweight	6,863 (4.2)	6,503 (4.3)	360 (3.6)	
Overweight	42,821 (26.4)	40,282 (26.5)	2,539 (25.1)	
Obesity	67,293 (41.5)	62,451 (41.1)	4,842 (47.8)	
Not measured	4,318 (2.7)	4,170 (2.7)	148 (1.5)	
HbA1c				**χ^2=5548.5, *p*<.001**
Normal	43,211 (26.7)	42,001 (27.6)	1,210 (12.0)	
Prediabetic	36,248 (22.4)	34,268 (22.6)	1,980 (19.6)	
Diabetic	46,495 (28.7)	40,414 (26.6)	6,081 (60.1)	
Not measured	36,109 (22.3)	35,255 (23.2)	854 (8.4)	
Blood Glucose				**χ^2=3839.2, *p*<.001**
Normal	17,483 (10.9)	16,350 (10.8)	1,133 (11.2)	
Low	589 (0.36)	553 (0.36)	36 (0.36)	
High	85,032 (52.5)	76,946 (50.6)	8,086 (79.9)	
Critical	199 (0.12)	193 (0.13)	6 (0.06)	
Not measured	58,760 (36.3)	57,896 (38.1)	864 (8.5)	
eGFR				**χ^2=1297.8, *p*<.001**
Normal	60,271 (37.2)	57,409 (37.8)	2,862 (28.3)	
Low	74,889 (46.2)	69,999 (46.1)	4,890 (48.3)	
Severely low	9,686 (6.0)	8,787 (5.8)	899 (8.9)	
Failure	9,145 (5.6)	7,929 (5.2)	1,216 (12.0)	
Not measured	8,072 (5.0)	7,814 (5.1)	258 (2.5)	
Insulin				
Insulin Regular	27,314 (16.9)	23,281 (15.3)	4,033	**χ^2=4067.4,**

			(39.8)	***p*<.001**
Insulin Isophane	11,961 (7.4)	9,933 (6.5)	2,028 (20.0)	**χ^2=2525.9, *p*<.001**
Insulin Aspart	93,836 (57.9)	84,717 (55.8)	9,119 (90.1)	**χ^2=4581.9, *p*<.001**
Insulin Glargine	52,332 (32.3)	45,495 (29.9)	6,837 (67.5)	**χ^2=6130.6, *p*<.001**
Insulin Detemir	4,233 (2.6)	3,451 (2.3)	782 (7.7)	**χ^2=1107.1, *p*<.001**

Preprint JMIR Publications

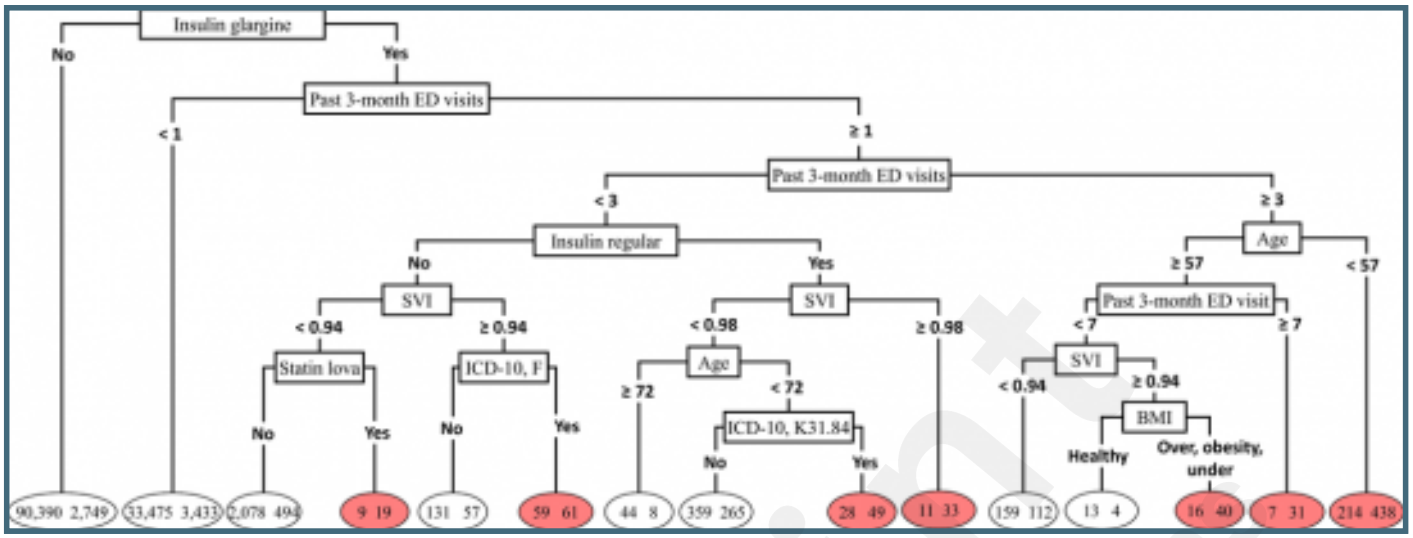
Supplementary Files

Preprint JMIR Publications

Figures

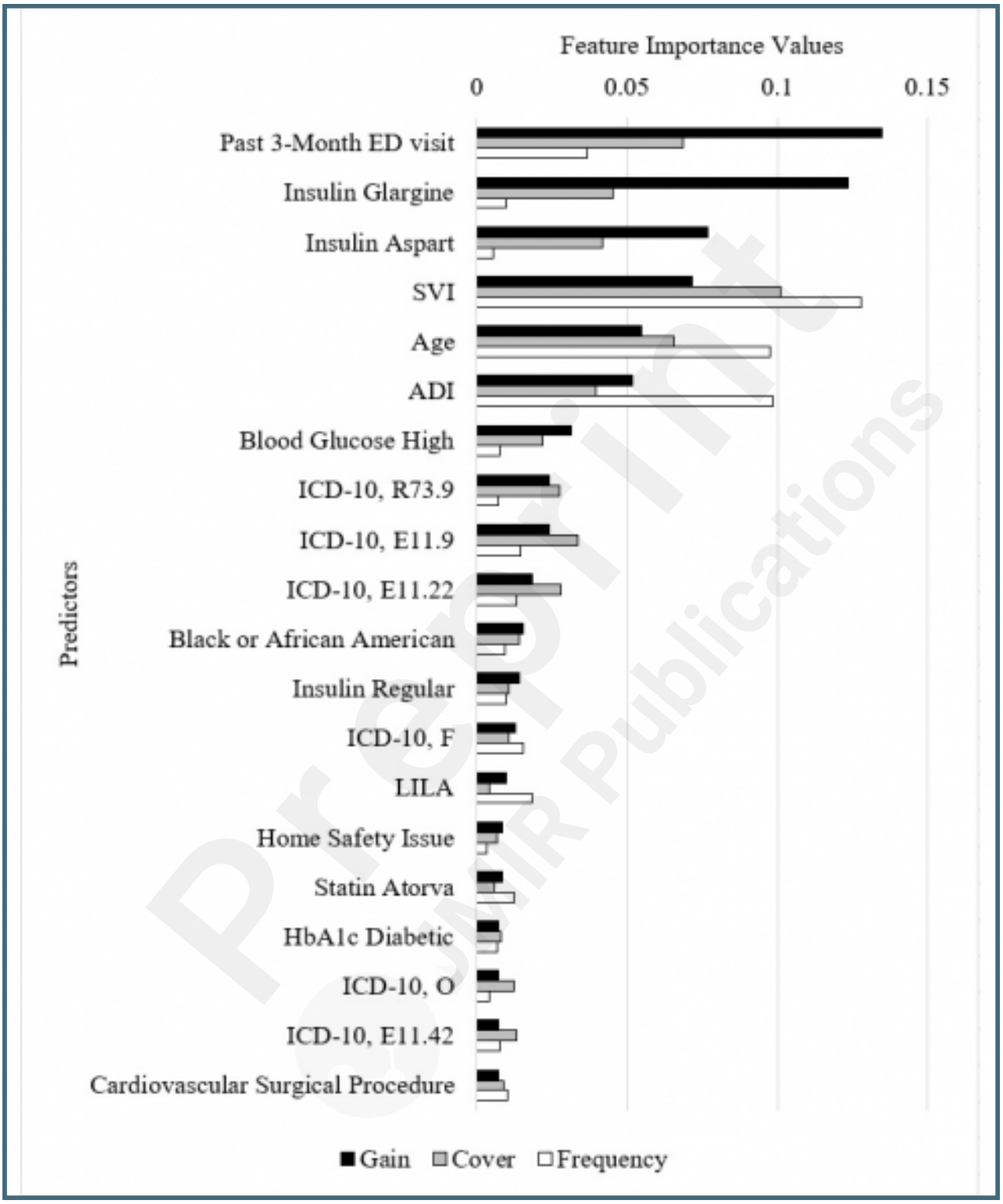
Preprint JMIR Publications

Decision tree structure.



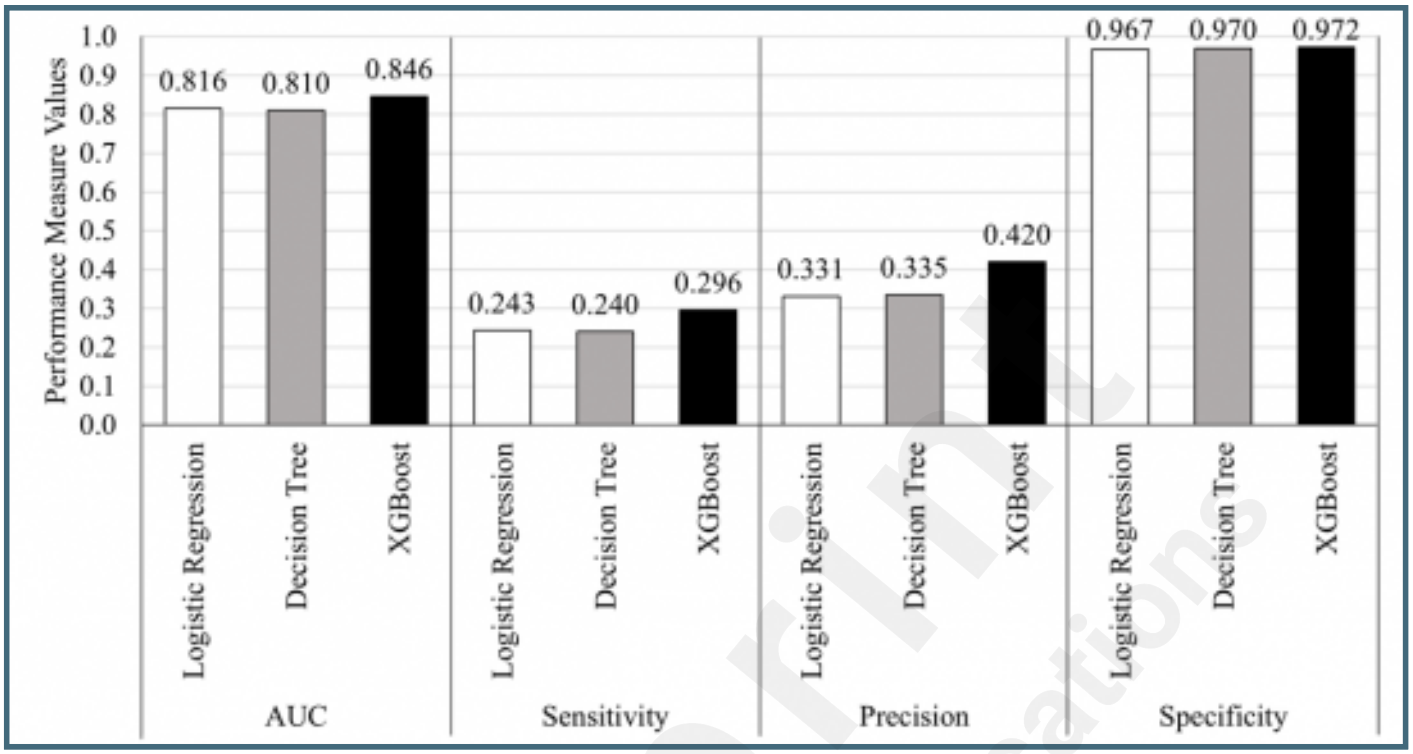
Preprint JMIR Publications

Feature importance from the XGBoost model.



Preprint JMIR Publications

Model performance comparison.



Preprint JMIR Publications