Nursing Outlook 73 (2025) 102431

Contents lists available at ScienceDirect

Nursing Outlook

journal homepage: www.nursingoutlook.org

Comparing the influence of social risk factors on machine learning model performance across racial and ethnic groups in home healthcare

Mollie Hobensack, PhD, RN [a,*], Anahita Davoudi, PhD [b], Jiyoun Song, PhD, APRN [c], Kenrick Cato, PhD, RN [c,d], Kathryn H. Bowles, PhD, RN [b,c], Maxim Topaz, PhD, RN [b,e,f]

[a] Icahn School of Medicine at Mount Sinai, New York, NY
[b] Center for Home Care Policy & Research, VNS Health, New York, NY
[c] University of Pennsylvania School of Nursing, Philadelphia, PA
[d] Children's Hospital of Philadelphia, Philadelphia, PA
[e] Columbia University School of Nursing, New York, NY
[f] Data Science Institute, Columbia University, New York, NY

ARTICLE INFO

Article history:
Received 21 August 2024
Received in revised form 12 April 2025
Accepted 15 April 2025
Available online xxxx

Keywords:
Home healthcare
Hospitalization
Social determinants of health
Machine learning
Nursing informatics

ABSTRACT

This study examined the impact of social risk factors on machine learning model performance for predicting hospitalization and emergency department visits in home healthcare. Using retrospective data from one U.S. home healthcare agency, four models were developed with unstructured social information documented in clinical notes. Performance was compared with and without social factors. A subgroup analyses was conducted by race and ethnicity to assess for fairness. LightGBM performed best overall. Social factors had a modest effect, but findings highlight the feasibility of integrating unstructured social information into machine learning models and the importance of fairness evaluation in home healthcare.

Introduction

In the United States, many patients exhibit a preference for receiving healthcare services at home rather than in a hospital setting, valuing the convenience, enhanced quality of life, and personalized care this pathway offers (Arieli et al., 2023; Le et al., 2022). Home healthcare (HHC) embodies this preference by allowing patients transitioning from acute care or referred from the community to access healthcare services in their own homes for periods ranging from 30 to 60 days. During this time, patients benefit from a comprehensive array of interdisciplinary services, including skilled nursing, physical therapy, occupational therapy, speech therapy, and social work services (Centers for Medicare and Medicaid Services, 2003). Despite strong evidence supporting the advantages of HHC (Admi et al., 2015; Howard et al., 2019; Romagnoli et al., 2013), about 20% of patients admitted to HHC are hospitalized or visit the emergency department (ED) (Busby et al., 2015; CMS, n.d.).

To decrease rates of hospitalization and ED visits, research has leveraged machine learning to identify patients at risk for deterioration (Huang et al., 2021). Machine learning can predict high-risk patients by analyzing routinely collected, clinician-generated data from the electronic health records (EHR). Previous findings from the team indicate that machine learning can effectively identify HHC patients at risk for hospitalization or ED visits (Hobensack et al., 2023b). However, one limitation of machine learning is it can perpetuate existing health disparities reflected within data (Rajkomar et al., 2018). Recent research has advocated for evaluating machine learning model fairness by comparing performance across subgroups (e.g., F-score), which can reveal biases that occur when a machine learning model systematically favors one social group over another (Barton et al., 2023; Huang et al., 2022). Bias within model performance can lead to disparities in treatment and interventions, exacerbating inequities in health delivery (Obermeyer et al., 2019).

The Biopsychosocial Model emphasizes a holistic approach to understanding health outcomes by integrating biological, psychological, and social factors (Borell-Carrió et al., 2004; Engel, 1977; Lehman et al., 2017). The biological dimension encompasses physiological characteristics; the psychological dimension

* Corresponding author: M. Hobensack, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place Box 1070, New York, NY 10029.
E-mail address: mollie.hobensack@mssm.edu (M. Hobensack).

https://doi.org/10.1016/j.outlook.2025.102431

includes emotions and health behaviors; and the social dimension includes socioeconomic circumstances (Lehman et al., 2017). A prior study reported improved machine learning performance when patient characteristics from all three dimensions were included, compared with models that relied on a single dimension alone (Chan et al., 2023). Yet, within HHC, only a limited number of studies have employed machine learning models that integrate all three dimensions, and few have included social factors (Hobensack et al., 2023b).

Social determinants of health—such as economic stability, education, healthcare access, social context, and the built environment—are estimated to account for over 50% of health outcomes (Truong et al., 2020). Social risk factors represent the adverse aspects of these determinants (e.g., lack of social support vs. presence of social support) that negatively impact health. Prior research suggests that social risk factors are more frequently documented in unstructured, free-text clinical notes than in structured fields such as drop-down assessments (Patra et al., 2021). Emerging research suggests that incorporating social risk factors into machine learning models may help mitigate performance disparities among racially and ethnically minoritized populations by better accounting for the underlying social complexities of health (Obermeyer et al., 2019; Rajkomar et al., 2018). However, no previous study has examined whether social risk factors documented in clinical notes influence machine learning model performance in the HHC setting or compared performance across racial and ethnic subgroups to evaluate fairness.

In prior work, natural language processing (NLP) was used to extract six social risk factors from HHC clinical notes and found that their presence was significantly associated with hospitalization and ED visits (Hobensack et al., 2023a). A follow-up study reported that social risk factors were more frequently documented in the clinical notes of patients from racially and ethnically minoritized groups suggesting potential differences in social complexity, clinician perception, or documentation practices (Hobensack et al., 2025). While these studies suggest a relationship between social risk factors documented in clinical notes and hospitalization and ED visits, they did not evaluate the predictive utility of these factors in machine learning models. Building on this foundational work, the present study investigated two key questions: (a) whether incorporating NLP-extracted social risk factors influenced machine learning model performance in predicting hospitalization and ED visits in the HHC setting, and (b) whether the inclusion of these factors affected model performance across racial and ethnic subgroups.

Methods

Setting and Sample

The study included a cohort of Medicare Fee-For-Service beneficiaries who received services from one not-for-profit HHC agency in the Northeastern United States between January 2015 and December 2017. The dataset included 86,866 HHC episodes of care (i.e., the period during which patients received HHC services) for 65,593 unique patients. The two types of EHR data included were structured (i.e., Outcome and Assessment Information Set [OASIS]) and unstructured free-text clinical notes (e.g., visit and care coordination notes) (Centers for Medicare and Medicaid Services, 2017). The OASIS is a standardized assessment tool mandated by Medicare and Medicaid Services for all HHC patients and is administered upon admission and discharge. It includes over 100 variables assessing patients' demographics, clinical characteristics, and functional status. The study's outcome, hospitalization, or ED visit was also extracted from the OASIS. In total, 2,341,018 clinical notes were included in the analysis.

Machine Learning Pipeline

All analyses were conducted using Anaconda Python (version 2.3.1) (Anaconda, Inc. New York, NY), and the unit of analysis was at the episode level. The machine learning pipeline focused on predicting if a HHC episode would result in a hospitalization or ED visit and involved several key steps:

Step 1: Text Preprocessing

First, the text was preprocessed before applying the machine learning models. This included lemmatization to simplify the words into their root meaning, transforming words like "caring" into "care." Then, stop words (e.g., "the," "and," "is," etc.) and nonalphanumeric symbols (e.g., "-", "!") were removed. These actions reduced the "noise" in the text, enhancing the efficiency of the machine learning models.

Step 2: Term Frequency Inverse Document Frequency

Term frequency -inverse document frequency (TF-IDF) was used to convert frequently occurring words from the clinical notes into features for the machine learning model (Nadkarni et al., 2011). TF-IDF considers both the frequency of a word within a clinical note and its prevalence across the entire corpus of clinical notes, highlighting the relevance of more meaningful words. A higher TF-IDF score indicates greater distinctiveness and informativeness of a term within a given context. The top 5,000 TF-IDF features were included in the machine learning models to enhance model performance (Chen et al., 2018).

Step 3: Development of Machine Learning Models

Based on a prior study, all significant predictors from an adjusted logistic regression model were included as binary variables in the machine learning models (Hobensack et al., 2023a). The NLP-social risk factors extracted from clinical notes included Social Environment, Physical Environment, Education and Literacy, Food Insecurity, Access to Care, and Housing and Economic Circumstances. Features selected for the machine learning model were mapped to the Biopsychosocial Model to ensure a comprehensive and holistic approach. Full details of this study are described elsewhere (Hobensack et al., 2023a).

Step 4: Dataset Creation and Cross-Validation

Two datasets were created: (a) a baseline dataset with characteristics from the OASIS + TF-IDF tokens and (b) a dataset with characteristics from the OASIS + TF-IDF tokens + *NLP-social risk factors*. Fivefold cross-validation was applied at the episode level by dividing the data into five equal parts or "folds." This ensured reliable machine learning performance estimates and reduced the overfitting risk (Tougui et al., 2021).

Step 5: Machine Learning Models

Four machine learning models were applied to both datasets to identify the best-performing machine learning model to predict hospitalization and ED visits among HHC patients. The four models included Logistic Regression, Support Vector Machine, Light Gradient Boosting Machine (LightGBM), and AutoGluon. The definition and rationale for selecting each model is detailed in Supplemental File A.

Step 6: Evaluation of the Machine Learning Models

Performance was evaluated using Accuracy, Precision, Recall, and F-score. Accuracy measures the proportion of predictions that the machine learning model identified correctly. Precision measures the ratio of correctly identified notes to the total number of notes that resulted in a hospitalization or ED visit. Recall measures the model's ability to correctly identify notes that resulted in a hospitalization or

ED visit. F-score is the harmonic mean of Precision and Recall. Values closer to zero indicate poor performance and values closer to one indicate better performance.

Step 7: Application of the Best-Performing Model Among Different Racial and Ethnic Groups

To assess fairness, model performance was compared across racial and ethnic subgroups. The best-performing machine learning model (identified in Step 6) was applied across the following groups: Hispanic, Non-Hispanic Black, Non-Hispanic White, and "Other" (i.e., Asian, Native American, Pacific Islander, and Other). These groups were predefined by categories in the OASIS (Centers for Medicare and Medicaid Services, 2017). Two datasets were created for each group (following the methodology outlined in Step 4) and model performance was assessed using Accuracy, Precision, Recall, and F-score.

Findings

Cohort Characteristics

Among the 86,866 HHC episodes, approximately 14% (n = 12,308) led to hospitalization or ED visits. The average age of the cohort was 78.7 years old (standard deviation = 11.8). A majority of the sample was reported as female (63.9%) and Non-Hispanic White (62.8%). Approximately 18% of the patients were reported as Non-Hispanic Black and approximately 14% were reported as Hispanic. Table 1 displays select characteristics of the total cohort, those who were hospitalized, and those who were not hospitalized. The full list of cohort characteristics is reported in Supplemental File B.

Biopsychosocial Model

In addition to the 5,000 TF-IDF keywords, there were 56 features included in the machine learning models (from OASIS and the clinical notes). Most features mapped to the biological dimension (n = 32; 56.1%); 11 features mapped to the psychological dimension (19.3%); and 9 features mapped to the social dimension (15.8%). The NLP-social risk factors represented 67% (n = 6) of the features in the social dimension. The three additional social factors from the OASIS were related to a patient's insurance status, living arrangements, and race. A fourth dimension, labeled "Other," was introduced to include features related to a patient's hospital utilization, such as length of stay and risk of hospitalization. The full mapping is presented in Table 2.

Evaluation of Machine Learning Models

Across all four models, Accuracy ranged from 0.77 to 0.88 and F-score ranged from 0.41 to 0.59. When incorporating the NLP-social risk factors, there were minimal changes (< 5% variation) in Precision, Recall, and Accuracy and no change in F-score. LightGBM achieved the highest overall performance with an Accuracy of 0.88 and an F-score of 0.59. Incorporating social risk factors led to a 2% increase in Precision (from 0.53 to 0.55) and a 3% decrease in Recall (from 0.68 to 0.65). Detailed results for all models are reported in Table 3.

LightGBM Performance Across Patients From Different Race or Ethnicity Subgroups

Before integrating the NLP-social risk factors, LightGBM consistently performed across all groups. The lowest performance was noted among individuals in the "Other" group (F-score = 0.55), while the highest performance was observed among individuals in the Hispanic group (F-score = 0.60). With the addition of the NLP-social risk factors, the F-score ranged from 0.59 to 0.65. There were minimal changes in F-scores (1%) among Non-Hispanic White, Non-Hispanic Black, and Hispanic individuals when the NLP-social risk factors were included. The exception was individuals in the "Other" group, where the F-score notably increased by 10% (0.55–0.65). Across all groups, Precision and Accuracy increased with the inclusion of the NLP-social risk factors. The most notable change was an 8% increase in Precision (from 0.48 to 0.56) among individuals in the "Other" group. Conversely, the greatest decrease (5%) was observed in Recall (from 0.68 to 0.63) among individuals in the Non-Hispanic Black group. Full results are reported in Figure 1.

Discussion

This study is the first in the HHC setting to integrate social risk factors extracted from clinical notes into predictive models and evaluate fairness through comparing model performance across racial and ethnic subgroups. Guided by the Biopsychosocial Model, the machine learning models incorporated features from the biological, psychological, and social domains, ensuring a comprehensive approach to predicting hospitalization and ED visits. Among the evaluated models, LightGBM demonstrated the highest overall performance with minimal differences in subgroup performance metrics across racial and ethnic groups.

This study provides early evidence supporting the feasibility of incorporating social risk factors documented in clinical notes into machine learning models within the HHC setting. While machine learning has shown promise in reducing adverse events in various care environments (Du et al., 2023; Romero-Brufau et al., 2020; Rossetti et al., 2021), its application in HHC remains limited—especially with respect to the inclusion of social risk factors (Hobensack et al., 2023b). Policymakers and healthcare organizations should incentivize the adoption of machine learning-based tools that explicitly account for social complexities, promoting health equity and enhancing clinical decision-making.

The impact of social risk factors in improving machine learning model performance is mixed (Li et al., 2022; Obuobi et al., 2021; Vest & Ben-Assuli, 2019; Zhao et al., 2021). While some studies suggest that individual-level social risk factors documented in clinical notes enhance the prediction of health outcomes (e.g., service referrals, risk of 30-day readmission) (Chen et al., 2020), other studies report no significant differences in model performance (Hammond et al., 2020; Seligman et al., 2018). A study by Zhang et al. (2020) found that social risk factors increased model performance among patients with different insurance types and ages. Another study by Segar et al. (2022) reported that among Black patients with heart failure, including social risk factors (e.g., zip-code) led to improved risk prediction for hospital mortality. This inconsistency underscores the complexity and contextual nature of social determinants, indicating that further investigation is necessary to fully understand their influence across diverse patient populations and healthcare contexts.

It is crucial to recognize that racial and ethnic categories are heterogeneous and that social risk factors may influence individuals differently within these groups (National Academies of Sciences, Engineering, and Medicine et al., 2017). Grounded in the Intersectionality Framework—which posits that overlapping social and political identities shape unique experiences of privilege and discrimination (Crenshaw, 1989), future research should examine how these intersecting identities interact with social risk factors. This approach may enhance understanding of how such factors influence the predictive accuracy of machine learning models in healthcare and, more critically, how they contribute to the perpetuation of health disparities (Ronquillo et al., 2022; Ying Yang et al., 2023).

This study examined the impact of six social risk factors in machine learning models as binary variables, meaning it did not evaluate the cumulative effect of these factors, previously referred to in the literature as social risk factor burden. Prior research has reported

Table 1
Cohort Characteristics

Variables	Total (*n* = 86,866) (%)	Hospitalized (*n* = 12,308) (%)	Not Hospitalized (*n* = 74,558) (%)
Demographics			
Age (years) (Mean [SD])	78.7 (11.8)	78.9 (11.9)	78.5 (11.7)
Length of stay (days)* (Mean [SD])	57.7 (80.1)	76.8 (110.6)	38.6 (49.6)
Sex			
Female (reference)	55,543 (63.9)	7,567 (61.5)	47,976 (64.3)
Male	31,323 (36.1)	4,741(38.5)	26,582 (35.7)
Race			
Non-Hispanic White (reference)	54,563 (62.8)	7,019 (57.0)	47,544 (63.8)
Non-Hispanic Black	15,207 (17.5)	2,650 (21.5)	12,557 (16.8)
Hispanic	11,710 (13.5)	2,039 (16.6)	9,671 (13.1)
Other (i.e., Asian, Native American, and Pacific Islander)	5,296 (6.1)	600 (4.9)	4,696 (6.3)
Insurance			
Dual	5,693 (6.6)	1,145 (9.3)	4,548 (6.1)
Medicare	81,099 (93.4)	11,154 (90.6)	69,945 (93.8)
Medicaid	21 (0.0)	4 (0.0)	17 (0.0)
Other (i.e., private)	52 (0.1)	4 (0.0)	48 (0.1)
Past medical history			
Risk of hospitalization			
Multiple prior hospitalizations	21,655 (24.9)	4,904 (39.8)	16,751 (22.5)
Taking more than five medications	69,041 (79.5)	10,313 (83.3)	58,728 (78.8)
Decline in mental, emotional, or behavioral status	12,839 (14.8)	2,178 (17.7)	10,661 (14.3)
Overall status			
Stable	5,837 (6.7)	721 (5.9)	5116 (6.9)
Temporary high risk	67,824 (78.1)	9,015 (73.2)	58,809 (78.9)
Remain high risk	12,700 (14.6)	2,447 (19.9)	10,253 (13.8)
Serious conditions could lead to death within a year	505 (0.6)	125 (1.0)	380 (0.5)
Social history			
Risk factors			
Smoking	5,906 (6.8)	930 (7.6)	4,976 (6.7)
Obesity	11,754 (13.5)	1,953 (15.9)	9,801 (13.1)
Alcohol dependency	989 (1.1)	172 (1.4)	817 (1.1)
Drug dependency	465 (0.5)	99 (0.8)	366 (0.5)
Living arrangements			
Lives in congregate care	2,738 (3.2)	461 (3.7)	2,277 (3.1)
Functional status			
ADLs (Mean [SD])			
Severity	16.6 (7.3)	17.62 (7.8)	15.50 (6.8)
Needed	8.1 (1.4)	8.26 (1.3)	8.03 (1.5)
Clinical assessment			
Integumentary			
Risk for pressure ulcer	35,728 (41.1)	6,340 (51.5)	29,388 (39.4)
Presence of an unhealed pressure ulcer	6,493 (7.5)	1,770 (14.4)	4,723 (6.3)
Presence of a surgical wound	23,076 (26.6)	2,179 (17.7)	20,897 (28.0)
Presence of skin lesion or open wound	17,979 (20.7)	3,500 (28.4)	14,479 (19.4)
Respiratory status			
Shortness of breath			
With minimal exertion	33,754 (4.3)	975 (7.9)	2,779 (3.7)
When walking more than 20 feet or climbing stairs	32,630 (37.6)	5,282 (42.9)	27,348 (36.7)
Elimination			
Urinary tract infection	5,967 (6.9)	1,211 (9.8)	4,756 (6.4)
Presence of urinary incontinence			
Incontinent	41,563 (47.8)	6,690 (54.4)	34,873 (46.8)
Catheter	3,080 (3.5)	885 (7.2)	2,195 (2.9)
Neuro, emotional, and behavioral status			
Cognitive functioning			
Alert and oriented	53,958 (62.1)	6,804 (55.3)	47,154 (63.2)
Requires prompting	21,616 (24.9)	3,456 (28.1)	18,160 (24.4)
Requires assistance	7,131 (8.2)	1,229 (10.0)	5,902 (7.9)
Not alert and oriented	2,948 (3.4)	542 (4.4)	2,406 (3.2)
Totally dependent	1,213 (1.4)	277 (2.3)	936 (1.3)
Social risk factors identified from clinical notes			
NLP-social risk factors			
Social environment	13,525 (15.6)	2,881 (23.4)	10,644 (14.3)
Physical environment	13,815 (15.9)	2,943 (23.9)	10,872 (14.6)
Education and literacy	4,528 (5.2)	1,052 (8.5)	3,476 (4.7)
Food insecurity	2,801 (3.2)	628 (5.1)	2,173 (2.9)
Access to care	5,146 (5.9)	1,210 (9.8)	3,936 (5.3)
Housing and economic circumstances	7,018 (8.1)	1,499 (12.2)	5,519 (7.4)

Note. ADL, activities of daily living; NLP, natural language processing; SD, standard deviation.
* Since this is an episode-level analysis, ranges over 60 days are considered outliers.

Table 2
Features Mapped to Biopsychosocial Model

Biopsychosocial Dimension	Source of Data	Features Included in the Machine Learning Model
Biological (*n* = 32)	OASIS	Age, sex, comorbidities (acute myocardial infarction, arthritis, cancer, cardiac dysrhythmias, dementia, diabetes, heart failure, pulmonary disease, peripheral vascular disease, renal disease, and skin ulcer), conditions within the past 14 days (urinary incontinence, indwelling catheter, intractable pain, impaired decision-making, disruptive behavior, and memory loss), overall status, sensory status (vision, ability to hear, and frequency of pain), integumentary (risk for pressure ulcer, presence of an unhealed pressure ulcer, presence of a surgical wound, and presence of skin lesion or open wound), respiratory status (shortness of breath), elimination (urinary tract infection, presence of urinary incontinence), and activities of daily living (severity, needed)
Psychological (*n* = 11)	OASIS	Neuro, emotional, and behavioral status (cognitive functioning, anxious), cognitive symptoms (memory deficit, impaired decision-making, verbal disruption, physical aggression, disruptive, and delusional), and risk factors (smoking, obesity, alcohol dependency, and drug dependency)
Social (*n* = 9)	OASIS	Insurance status, living arrangements (living in congregate care), and race
	Clinical notes	Social environment, physical environment, education and literacy, food insecurity, access to care, and housing and economic circumstances
Other (*n* = 4)	OASIS	Length of stay, risk of hospitalization (multiple prior hospitalizations, taking more than five medications, and decline in mental, emotional, or behavioral status)

Note. OASIS, outcome and assessment information set.
The total number of features was 56. Information in parenthesis represents the multiple features within each category included in the machine learning model.

Table 3
Evaluation of Machine Learning Models

	Logistic Regression		SVM		LightGBM		AutoGluon	
	Baseline	+NLP-Social Risk Factors	Baseline	+NLP-Social Risk Factors	Baseline	+NLP-Social Risk Factors	Baseline	+NLP-Social Risk Factors
Precision	0.32	0.32	0.36	0.37	**0.53**	**0.55**	0.45	0.44
Recall	0.57	0.58	0.52	0.50	**0.68**	**0.63**	0.60	0.62
Accuracy	0.77	0.77	0.80	0.81	**0.88**	**0.88**	0.84	0.83
F-score	0.41	0.41	0.43	0.43	**0.59**	**0.59**	0.51	0.51

Note. LightGBM, light gradient boosting machine; NLP, natural language processing; SVM, support vector machines.
All models also included TF-IDF tokens.
LightGBM had the best performance across the other algorithms and thus is highlighted.

that the cumulative impact of social risk factors increases the odds of adverse outcomes (Reshetnyak et al., 2020; Wray et al., 2022). Future studies should explore the benefits of incorporating features that capture the combined presence of multiple social risk factors to better understand their overall impact.

This study adds to the growing literature demonstrating the capability of the machine learning model, LightGBM, in clinical prediction (Daoud, 2019; Wang & Wang, 2020). Key strengths of LightGBM include its high performance, predictive accuracy, stability, and efficiency (Banerjee et al., 2019). Although LightGBM is

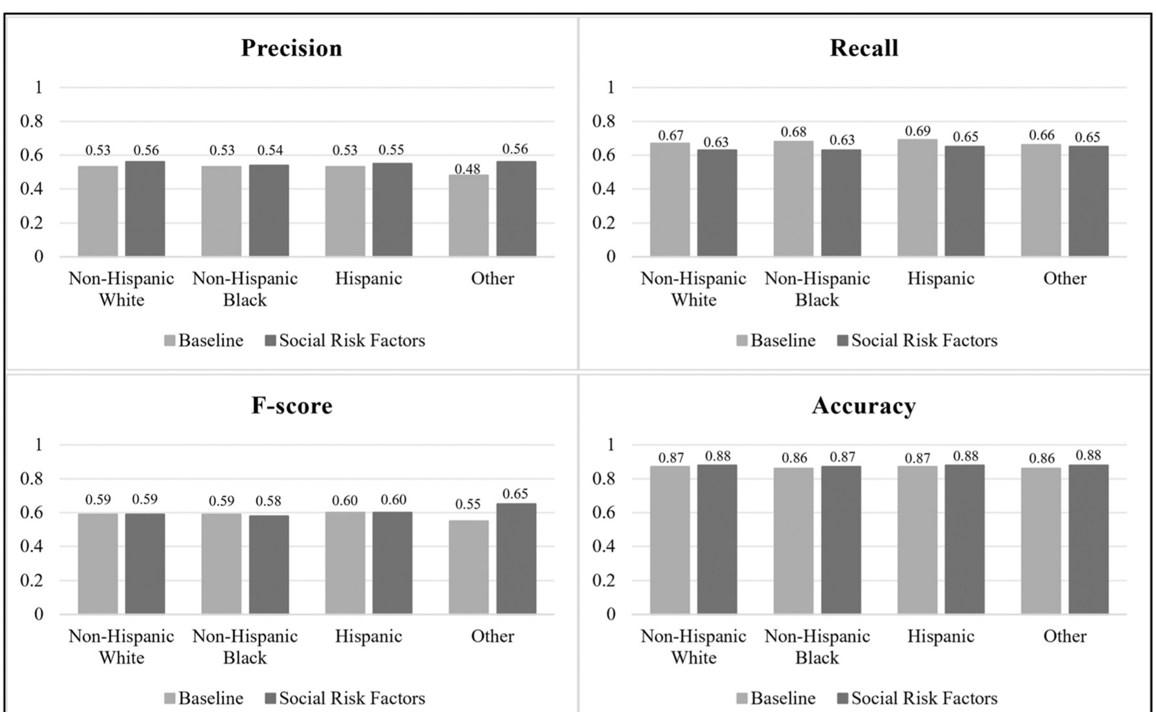

Figure 1. Machine learning performance across different racial and ethnic groups. *Note.* All models included TF-IDF tokens. The highest value for each metric is one indicating best performance. TF-IDF, term frequency -inverse document frequency.

generally considered less interpretable among tree-based models, recent research has introduced new methods to clarify how each feature contributes to model performance. These methods include Local Interpretable Model-Agnostic Explanations and Shapley Additive Explanations (Khanna et al., 2023; Omobolaji Alabi et al., 2022). A prior study reported that the explainability of a machine learning model is vital to fostering clinician trust and utilization in practice (Schwartz et al., 2022). Future work should explore supplemental methodologies to increase LightGBM's explainability to proactively support clinical integration.

The minimal racial and ethnic subgroup differences observed in model performance may reflect several factors. First, because the dataset came from a single agency in the Northeastern United States, consistent documentation norms and clinical practices across patient subgroups may have reduced variability, thereby potentially minimizing observable disparities (Rajkomar et al., 2018). Additionally, aggregating smaller racial and ethnic populations into an "Other" category might obscure nuanced performance differences (Movva et al., 2023). Limited subgroup sample sizes could further restrict the detection of meaningful disparities (Movva et al., 2023). Lastly, the inclusion of structured clinical data and NLP-derived social risk factors likely contributed to consistent performance by enhancing the model's predictive generalizability across diverse patient subgroups. Future research should investigate how dataset diversity, subgroup representation, and feature selection influence model fairness.

Several limitations should be noted. The use of EHR data from 2015 to 2017 and a single HHC agency in the Northeastern United States limits generalizability. Small sample sizes required combining Asian, Native American, Pacific Islander, and other racial or ethnic groups into an "Other" category, reducing the ability to explore subgroup nuances between social risk factors and hospitalization or ED visits. Individual contributions of the 5,000 TF-IDF keywords were not analyzed, nor were individual contributions or collinearity among specific NLP-extracted social risk factors evaluated. The NLP extraction method may have introduced inaccuracies due to inconsistent or biased clinician documentation practices. Future research should validate NLP-derived social risk factors against structured assessments, apply advanced NLP techniques, and comprehensively analyze social risk factors' feature importance and collinearity. Additionally, this study assessed fairness by comparing subgroup performance metrics; subsequent studies should explore other fairness metrics, mitigation strategies, and how variations in clinical workflow and documentation practices influence machine learning model performance across demographic groups.

Conclusion

This study is the first in the HHC setting to integrate social risk factors extracted from clinical notes into machine learning models and assess their impact on fairness across racial and ethnic groups. Grounded in the Biopsychosocial Model, the analysis incorporated a comprehensive set of biological, psychological, and social features—offering one of the most holistic approaches to date for predicting hospitalization and ED visits in HHC. The observed variability in model performance across subgroups underscores the complex and context-dependent role of social risk factors. These findings highlight the need for further research into the intersectionality of patient identities and the cumulative burden of social risks, both of which are critical for advancing more accurate, equitable, and clinically meaningful predictive tools in home-based care.

Funding

This work was supported by the National Institute of Nursing Research (NINR) (grant T32NR007969 [M.H.], T32NR009356 [J.S.]), the Jonas Scholarship (M.H.), and the Agency for Healthcare Research and Quality (grant R01 HS027742).

CRediT Statement

Mollie Hobensack: Writing- Review and Editing, Writing- Original draft preparation, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. *Anahita Davoudi:* Writing- Review and Editing, Formal analysis. *Jiyoun Song:* Writing- Review and Editing, Methodology. *Kenrick Cato:* Writing- Review and Editing, Supervision. *Kathryn H. Bowles:* Writing- Review and Editing, Supervision. *Maxim Topaz:* Writing- Review and Editing, Supervision, Methodology, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used ChatGPT for editorial support. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Acknowledgments

This study also benefited from the generous support of my dissertation committee. A special thanks to Drs. Gregory Alexander and Meghan Turchioe for their feedback on this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.outlook.2025.102431.

References

Admi, H., Shadmi, E., Baruch, H., & Zisberg, A. (2015). From research to reality: Minimizing the effects of hospitalization on older adults. *Rambam Maimonides Medical Journal, 6*(2), Article e0017. https://doi.org/10.5041/RMMJ.10201

Arieli, M., Kizony, R., Gil, E., & Agmon, M. (2023). Participation in daily activities after acute illness hospitalization among high-functioning older adults: A qualitative study. *Journal of Clinical Nursing, 32*(13-14), 3456–3468. https://doi.org/10.1111/JOCN.16418

Baciu, A., Negussie, Y., Geller, A., & Weinstein, J. N. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Population Health and Public Health Practice, Committee on Community-Based Solutions to Promote Health Equity in the United States. (2017). The root causes of health inequity. In J. N. Weinstein, A. Geller, Y. Negussie, & A. Baciu (Eds.). *Communities in action: Pathways to health equity* (pp. 1–558). National Academies of Sciences, Engineering, and Medicine 10.17226/24624.

Banerjee, M., Reynolds, E., Andersson, H. B., & Nallamothu, B. K. (2019). Tree-based analysis: A practical approach to create clinical decision making tools. *Circulation Cardiovascular Quality and Outcomes, 12*(5), Article e004879. https://doi.org/10.1161/CIRCOUTCOMES.118.004879

Barton, M., Hamza, M., & Guevel, B. (2023). Racial equity in healthcare machine learning: Illustrating bias in models with minimal bias mitigation. *Cureus, 15*(2), https://doi.org/10.7759/CUREUS.35037

Borell-Carrió, F., Suchman, A. L., & Epstein, R. M. (2004). The biopsychosocial model 25 years later: Principles, practice, and scientific inquiry. *Annals of Family Medicine, 2*(6), 576. https://doi.org/10.1370/AFM.245

Busby, J., Purdy, S., & Hollingworth, W. (2015). A systematic review of the magnitude and cause of geographic variation in unplanned hospital admission rates and length of stay for ambulatory care sensitive conditions. *BMC Health Services Research, 15*(1), 1–15. https://doi.org/10.1186/S12913-015-0964-3

Centers for Medicare and Medicaid Services. (2003). *Medicare and home health care*. US Department of Health and Human Services. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/Downloads/HHQIHHBenefits.pdf.

Centers for Medicare and Medicaid Services. (2017). *Outcome and Assessment Information Set (OASIS)-C2 Guidance Manual*. Centers for MEdicare and Medicaid Services.

Chan, L., Simmons, C., Tillem, S., Conley, M., Brazil, I. A., & Baskin-Sommers, A. (2023). Classifying conduct disorder using a biopsychosocial model and machine learning method. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 8*(6), 599–608. https://doi.org/10.1016/J.BPSC.2022.02.004

Chen, M., Tan, X., & Padman, R. (2020). Social determinants of health in electronic health records and their impact on analysis and risk prediction: A systematic review. *Journal of the American Medical Informatics Association, 27*(11), 1764–1773. https://doi.org/10.1093/jamia/ocaa143

Chen, P. H., Zafar, H., Galperin-Aizenberg, M., & Cook, T. (2018). Integrating natural language processing and machine learning algorithms to categorize oncologic response in radiology reports. *Journal of Digital Imaging, 31*(2), 178. https://doi.org/10.1007/S10278-017-0027-X

CMS. (2025). *Home health quality measures*. Centers for Medicare and Medicaid Services. Retrieved October 1, 2021, from https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/Home-Health-Quality-Measures.

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum, 1*(8)http://chicagounbound.uchicago.edu/uclfhttp://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8.

Daoud, E. A. (2019). Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset. *International Journal of Computer and Information Engineering, 13*(1).

Du, Y., McNestry, C., Wei, L., Antoniadi, A. M., McAuliffe, F. M., & Mooney, C. (2023). Machine learning-based clinical decision support systems for pregnancy care: A systematic review. *International Journal of Medical Informatics, 173*, Article 105040. https://doi.org/10.1016/J.IJMEDINF.2023.105040

Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science, 196*(4286), 129–136. https://doi.org/10.1126/SCIENCE.847460

Hammond, G., Johnston, K., Huang, K., & Joynt Maddox, K. E. (2020). Social determinants of health improve predictive accuracy of clinical risk models for cardiovascular hospitalization, annual cost, and death. *Circulation: Cardiovascular Quality and Outcomes, 13*(6), Article e006752. https://doi.org/10.1161/CIRCOUTCOMES.120.006752

Hobensack, M., Scharp, D., Song, J., & Topaz, M. (2025). Documentation of social determinants of health across individuals from different racial and ethnic groups in home healthcare. Journal of nursing scholarship: an official publication of Sigma Theta Tau International Honor Society of Nursing, 57(1), 39–46. https://doi.org/10.1111/jnu.12980.

Hobensack, M., Song, J., Oh, S., Evans, L., Davoudi, A., Bowles, K. H., McDonald, M., Barron, Y., Sridharan, S., Wallace, A., & Topaz, M. (2023a). Social risk factors are associated with risk for hospitalization in home healthcare: A natural language processing study. *The Journal of Post-Acute and Long-Term Care Medicine*.

Hobensack, M., Song, J., Scharp, D., Bowles, K. H., & Topaz, M. (2023b). Machine learning applied to electronic health record data in home healthcare: A scoping review. *International Journal of Medical Informatics, 170*, Article 104978. https://doi.org/10.1016/J.IJMEDINF.2022.104978

Howard, J., Kent, T., Stuck, A. R., Crowley, C., & Zeng, F. (2019). Improved cost and utilization among medicare beneficiaries dispositioned from the ED to receive home health care compared with inpatient hospitalization. *The American Journal of Accountable Care, 7*(1), 10–16.

Huang, J., Galal, G., Etemadi, M., & Vaidyanathan, M. (2022). Evaluation and mitigation of racial bias in clinical machine learning models: Scoping review. *JMIR Medical Informatics, 10*(5), https://doi.org/10.2196/36388

Huang, Y., Talwar, A., Chatterjee, S., & Aparasu, R. R. (2021). Application of machine learning in predicting hospital readmissions: A scoping review of the literature. *BMC Medical Research Methodology, 21*(1), 1–14. https://doi.org/10.1186/S12874-021-01284-Z/TABLES/3

Khanna, V. V., Chadaga, K., Sampathila, N., Prabhu, S., & Rajagopala Chadaga, P. (2023). A machine learning and explainable artificial intelligence triage-prediction system for COVID-19. *Decision Analytics Journal, 7*, Article 100246. https://doi.org/10.1016/J.DAJOUR.2023.100246

Le, N. C., Rahman, T., Kapralik, J. L., Ibrahim, Q., Lear, S. A., & Van Spall, H. G. C. (2022). The hospital at home model vs routine hospitalization for acute heart failure: A survey of patients' preferences. *CJC Open, 4*(3), 263–270. https://doi.org/10.1016/J.CJCO.2021.10.005

Lehman, B. J., David, D. M., & Gruber, J. A. (2017). Rethinking the biopsychosocial model of health: Understanding health as a dynamic system. *Social and Personality Psychology Compass, 11*(8), Article e12328. https://doi.org/10.1111/SPC3.12328

Li, Y., Wang, H., & Luo, Y. (2022). Improving fairness in the prediction of heart failure length of stay and mortality by integrating social determinants of health. *Circulation: Heart Failure, 15*(11), Article E009473. https://doi.org/10.1161/CIRCHEARTFAILURE.122.009473

Movva, R., Tech, C., Shanmugam, D., Hou, K., Pathak, P., Guttag, J., Garg, N., & Pierson, E. (2023). Coarse race data conceals disparities in clinical risk score performance. *Proceedings of Machine Learning Research, 219*, 1–34.

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association, 18*(5), 544–551. https://doi.org/10.1136/amiajnl-2011-000464

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453. https://doi.org/10.1126/SCIENCE.AAX2342/SUPPL_FILE/AAX2342_OBERMEYER_SM.PDF

Obuobi, S., Chua, R. F. M., Besser, S. A., Tabit, C. E., S, O., RFM, C., SA, B., & CE, T. (2021). Social determinants of health and hospital readmissions: Can the HOSPITAL risk score be improved by the inclusion of social factors. *BMC Health Services Research, 21*(1), 5. https://doi.org/10.1186/S12913-020-05989-7

Omobolaji Alabi, R., Almangush, A., Elmusrati, M., Leivo, I., & Mäkitie, A. A. (2022). An interpretable machine learning prognostic system for risk stratification in oropharyngeal cancer. *International Journal of Medical Informatics, 168*, Article 104896. https://doi.org/10.1016/J.IJMEDINF.2022.104896

Patra, B. G., Sharma, M. M., Vekaria, V., Adekkanattu, P., Patterson, O. V., Glicksberg, B., Lepow, L. A., Ryu, E., Biernacka, J. M., Furmanchuk, A., George, T. J., Hogan, W., Wu, Y., Yang, X., Bian, J., Weissman, M., Wickramaratne, P., Mann, J. J., Olfson, M., ... Pathak, J. (2021). Extracting social determinants of health from electronic health records using natural language processing: a systematic review. *Journal of the American Medical Informatics Association, 2021*, 1–12. https://doi.org/10.1093/JAMIA/OCAB170

Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine, 169*(12), 866–872. https://doi.org/10.7326/M18-1990

Reshetnyak, E., Ntamatungiro, M., Pinheiro, L. C., Howard, V. J., Carson, A. P., Martin, K. D., & Safford, M. M. (2020). Impact of multiple social determinants of health on incident stroke. *Stroke, 51*(8), 2445–2453. https://doi.org/10.1161/STROKEAHA.120.028530

Romagnoli, K. M., Handler, S. M., & Hochheiser, H. (2013). Home care: More than just a visiting nurse. *BMJ Quality & Safety, 22*(12), 972. https://doi.org/10.1136/BMJQS-2013-002339

Romero-Brufau, S., Wyatt, K. D., Boyum, P., Mickelson, M., Moore, M., & Cognetta-Rieke, C. (2020). Implementation of Artificial Intelligence-Based clinical decision support to reduce hospital readmissions at a regional hospital. *Applied Clinical Informatics, 11*(4), 570–577. https://doi.org/10.1055/s-0040-1715827

Ronquillo, C. E., Mitchell, J., Alhuwail, D., Peltonen, L. M., Topaz, M., & Block, L. J. (2022). The untapped potential of nursing and allied health data for improved representation of social determinants of health and intersectionality in artificial intelligence applications: A rapid review. *Yearbook of Medical Informatics, 31*(1), 94–99. https://doi.org/10.1055/S-0042-1742504/BIB

Rossetti, S. C., Dykes, P. C., Knaplund, C., Kang, M. J., Schnock, K., Garcia, J. P., Fu, L. H., Chang, F., Thai, T., Fred, M., Korach, T. Z., Zhou, L., Klann, J. G., Albers, D., Schwartz, J., Lowenthal, G., Jia, H., Liu, F., & Cato, K. (2021). The Communicating Narrative Concerns Entered by Registered Nurses (CONCERN) clinical decision support early warning system: Protocol for a cluster randomized pragmatic clinical trial. *JMIR Research Protocols, 10*(12), https://doi.org/10.2196/30238

Schwartz, J. M., George, M., Rossetti, S. C., Dykes, P. C., Minshall, S. R., Lucas, E., & Cato, K. D. (2022). Factors influencing clinician trust in predictive clinical decision support systems for in-hospital deterioration: Qualitative descriptive study. *JMIR Human Factors, 9*(2), https://doi.org/10.2196/33960

Segar, M. W., Hall, J. L., Jhund, P. S., Powell-Wiley, T. M., Morris, A. A., Kao, D., Fonarow, G. C., Hernandez, R., Ibrahim, N. E., Rutan, C., Navar, A. M., Stevens, L. M., & Pandey, A. (2022). Machine learning–based models incorporating social determinants of health vs traditional models for predicting in-hospital mortality in patients with heart failure. *JAMA Cardiology, 7*(8), 844–854. https://doi.org/10.1001/JAMACARDIO.2022.1900

Seligman, B., Tuljapurkar, S., & Rehkopf, D. (2018). Machine learning approaches to the social determinants of health in the health and retirement study. *SSM - Population Health, 4*, 95–99. https://doi.org/10.1016/J.SSMPH.2017.11.008

Tougui, I., Jilbab, A., & Mhamdi, J. El (2021). Impact of the choice of cross-validation techniques on the results of machine learning-based diagnostic applications. *Healthcare Informatics Research, 27*(3), 189. https://doi.org/10.4258/HIR.2021.27.3.189

Truong, H. P., Luke, A. A., Hammond, G., Wadhera, R. K., Reidhead, M., & Joynt Maddox, K. E. (2020). Utilization of social determinants of health ICD-10 Z-codes among hospitalized patients in the United States, 2016-2017. *Medical Care, 58*(12), 1037. https://doi.org/10.1097/MLR.0000000000001418

Vest, J. R., & Ben-Assuli, O. (2019). Prediction of emergency department revisits using area-level social determinants of health measures and health information exchange information. *International Journal of Medical Informatics, 129*, 205–210. https://doi.org/10.1016/J.IJMEDINF.2019.06.013

Wang, Y., & Wang, T. (2020). Application of improved LightGBM model in blood glucose prediction. *Applied Sciences, 10*(9), 3227. https://doi.org/10.3390/APP10093227

Wray, C. M., Tang, J., López, L., Hoggatt, K., & Keyhani, S. (2022). Association of social determinants of health and their cumulative impact on hospitalization among a national sample of community-dwelling US adults. *Journal of General Internal Medicine, 37*(8), 1935–1942. https://doi.org/10.1007/S11606-021-07067-Y/FIGURES/1

Yang, M.Y., Kwak, G.H., Pollard, T., Celi, L.A., & Ghassemi, M. (2023, August). Evaluating the impact of social determinants on health prediction in the intensive care unit. In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (pp. 333-350).

Zhang, Y., Zhang, Y., Sholle, E., Abedian, S., Sharko, M., Turchioe, M. R., Wu, Y., & Ancker, J. S. (2020). Assessing the impact of social determinants of health on predictive models for potentially avoidable 30-day readmission or death. *PLoS ONE, 15*(6), https://doi.org/10.1371/JOURNAL.PONE.0235064

Zhao, Y., Wood, E. P., Mirin, N., Cook, S. H., & Chunara, R. (2021). Social determinants in machine learning cardiovascular disease prediction models: A systematic review. *American Journal of Preventive Medicine, 61*(4), 596–605. https://doi.org/10.1016/J.AMEPRE.2021.04.016