

Research and Applications

Utilizing patient-nurse verbal communication in building risk identification models: the missing critical data stream in home healthcare

Maryam Zolnoori, PhD^{1,2*}, Sridevi Sridharan, MS, Ali Zolnour, MS³, Sasha Vergez, BS², Margaret V. McDonald, MS², Zoran Kostic, PhD⁴, Kathryn H. Bowles, PhD^{2,5}, Maxim Topaz, PhD^{1,2}

¹School of Nursing, Columbia University, New York, NY 10032, United States, ²Center for Home Care Policy & Research, VNS Health, New York, NY 10017, United States, ³School of Electrical and Computer Engineering, University of Tehran, Tehran 14395-515, Iran, ⁴Electrical Engineering Department, Columbia University, New York, NY 10027, United States, ⁵School of Nursing, University of Pennsylvania, Philadelphia, PA 19104, United States

*Corresponding author: Maryam Zolnoori, PhD, School of Nursing, Columbia University Medical Center, 390 Fort Washington Avenue, New York, NY 10033 (mz2825@cumc.columbia.edu; m.zolnoori@gmail.com)

Abstract

Background: In the United States, over 12 000 home healthcare agencies annually serve 6+ million patients, mostly aged 65+ years with chronic conditions. One in three of these patients end up visiting emergency department (ED) or being hospitalized. Existing risk identification models based on electronic health record (EHR) data have suboptimal performance in detecting these high-risk patients.

Objectives: To measure the added value of integrating audio-recorded home healthcare patient-nurse verbal communication into a risk identification model built on home healthcare EHR data and clinical notes.

Methods: This pilot study was conducted at one of the largest not-for-profit home healthcare agencies in the United States. We audio-recorded 126 patient-nurse encounters for 47 patients, out of which 8 patients experienced ED visits and hospitalization. The risk model was developed and tested iteratively using: (1) structured data from the Outcome and Assessment Information Set, (2) clinical notes, and (3) verbal communication features. We used various natural language processing methods to model the communication between patients and nurses.

Results: Using a Support Vector Machine classifier, trained on the most informative features from OASIS, clinical notes, and verbal communication, we achieved an AUC-ROC = 99.68 and an F1-score = 94.12. By integrating verbal communication into the risk models, the F-1 score improved by 26%. The analysis revealed patients at high risk tended to interact more with risk-associated cues, exhibit more “sadness” and “anxiety,” and have extended periods of silence during conversation.

Conclusion: This innovative study underscores the immense value of incorporating patient-nurse verbal communication in enhancing risk prediction models for hospitalizations and ED visits, suggesting the need for an evolved clinical workflow that integrates routine patient-nurse verbal communication recording into the medical record.

Key words: home healthcare; emergency department visit and hospitalization; audio-recorded patient-nurse verbal communication; natural language processing; machine learning.

Introduction

Recording the conversations between patients and clinicians during regular healthcare visits can capture explicit and subtle information. This information helps identify signs of serious health conditions and communication issues and helps pinpoint risk factors that could lead to negative health outcomes.¹ Despite the importance of this data stream, patients’ spoken language remains an uncaptured data stream. In addition, our previous study in home healthcare showed that about 50% of clinical risk factors discussed during home healthcare visits are not documented in the electronic health record (EHR) system (either in free-text clinical notes or in structured clinical data).² Past studies have shown that it is feasible to audio-record home healthcare visits, and both healthcare professionals and patients, in particular, have expressed positive views about recording conversations between patients and nurses.^{1,2}

Home healthcare is a setting where skilled clinicians (often registered nurses) provide healthcare services to patients in their homes; in the United States, 12 000 home healthcare agencies provide care to more than 6 million patients annually.³ Home healthcare patients are generally older adults aged ≥65 years and often are clinically complex and vulnerable patients with multiple chronic conditions, and they are at risk of negative outcomes such as emergency department (ED) visits and hospitalization.⁴ About 1 in 3 home healthcare patients are hospitalized or visit ED;⁵ timely risk detection and interventions can reduce those risks by up to 40%.⁶ Therefore, developing risk identification models to identify high-risk patients and prioritize them for appropriate intervention is crucial.

Emerging studies have utilized natural language processing (NLP) methods to measure changes in linguistic parameters of

the patients' speech as biomarkers for detecting patients with pathological entities such as mental and neurological disorders or cardiovascular diseases.^{7,8} However, most of the studies were conducted in laboratory settings where patients were instructed to complete some speech production tasks (eg, reading task) in a short time (eg, 1 min). Such collected speech data often did not include important linguistic cues or clinical risk factors associated with negative outcomes. This study is unique in that we audio-recorded patient-nurse verbal communication in home healthcare and developed an analytical pipeline to extract linguistic and communication cues and clinical risk factors from the audio-recorded data for early identification of patients at risk of ED visits and hospitalization.

This study aimed to measure the added value of integrating audio-recorded patient-nurse verbal communication into a risk identification model built on home healthcare EHR data and clinical notes. We hypothesize that adding speech features to the clinical notes and OASIS datasets can improve the performance of the risk identification model for patients at risk of ED visits and hospitalization.

Methods

This research was conducted at one of the largest not-for-profit home healthcare agencies in the United States, and approved by the agency's institutional review board (reference no. E20-003). We recruited 5 registered nurses who provided care for older adults individuals in their homes and agreed to record their home encounters with patients. The study was initially introduced to the patients by the nurses, and if they showed interest, a research assistant followed up to obtain informed consent. To be eligible for this study, patients had to be proficient in English, capable of communicating with nurses without needing a caregiver's assistance, and possess the cognitive capacity to read, comprehend, and independently sign the informed consent form. [Figure 1](#) provides an overview of the methodology of the study.

Audio recording of home healthcare patient-nurse verbal communication

The procedure of audio-recording patient-nurse encounters

Through a sequence of pilot studies, we determined the most effective approach for audio-recording patient-nurse conversations. We specifically assessed how well various audio-recording devices worked and how easy they were to use in a home healthcare environment, with 10 patients and 3 nurses actively participating in the audio-recording process; full details were presented in our previous study.¹

Overall, we selected Saramonic Blink658 for its superior usability rating and high-quality audio recording. This device is portable and lightweight, equipped with 2 wireless microphones clipped onto participants' clothing. Captured speech is transmitted to a connected device, such as an iPod, and stored in 2 separate channels (refer to [Appendix SA](#) for an illustration of this device). Participants (patients and nurses) found this recording process acceptable. Patients reported that audio recording did not affect their interaction with the nurses.¹

Accuracy of the automatic speech recognition system for recording transcription and speaker diarization

We used Amazon Web Service (AWS)-General Transcribe (GT) to automatically transcribe the audio-recorded encounters, as it showed the lowest word error rate of 26% compared to other speech recognition systems like AWS Medical Transcribe and Wave2Vec.⁹ AWS-GT provides a transcription of each spoken word and the corresponding start and end times of the word, along with the speaker's identity (for instance, speaker #1 and speaker #2 in a conversation involving 2 people), a feature known as speaker diarization. By manually reviewing the accuracy of speaker diarization for a sample of audio-recorded data in our previous study,¹ we computed the overall high accuracy of 96%.

Differentiating between patient and nurse language: speaker type identification

During patient-nurse encounters in home healthcare, speakers take turns, with each uninterrupted block of speech referred to as an utterance. By aggregating the spoken words of each speaker (speaker #1 and speaker #2), as transcribed by AWS-GT, we formed utterances. Then, to discern the roles of these speakers—in our case, the patient and the nurse, we applied a speaker type identification algorithm developed in our prior study.¹⁰ This algorithm, capable of automatically identifying whether speaker #1 or #2 is the patient or the nurse, achieved an F-1 score of 0.96. To verify the speaker type (patient or nurse) assigned to each utterance, a manual review was carried out by a member of our research team.

Datasets, data preprocessing, and feature generation

Data stored in the home healthcare EHR include the OASIS assessment and clinical notes. We integrated these datasets with the audio-recorded encounters to construct our analytic dataset.

OASIS dataset

This dataset covers about 100 binary, categorical, and continuous data, such as "risk of fall" (Yes vs No), "the patient's ability to perform daily living activities" (subcategories: independent, assistance required, completely dependent), and "age." OASIS is administered on admission and upon completion of a home healthcare episode, typically lasting about 4 weeks. Authors MZ and SS assessed the OASIS dataset's variables. Redundant variables were omitted, and sparse subcategories within categorical variables were combined.

Clinical notes dataset

This dataset consists of visit and care coordination notes written by nurses, offering a narrative evaluation of the patient's clinical condition and outlining communication and administrative activities related to the patient's care. All clinical notes were preprocessed by lowercasing and removing extra punctuation. We did not remove stop words as they provide context to the nurses' intent, and therefore they are important for modeling the conceptual relationship between words, specifically using transformer pre-trained language models. Stop words are commonly used words (eg, "and," "the," "is") that are often filtered out in text processing due to their frequency. Transformer-based models, including BERT¹¹ (Bidirectional Encoder Representations from Transformers) and their

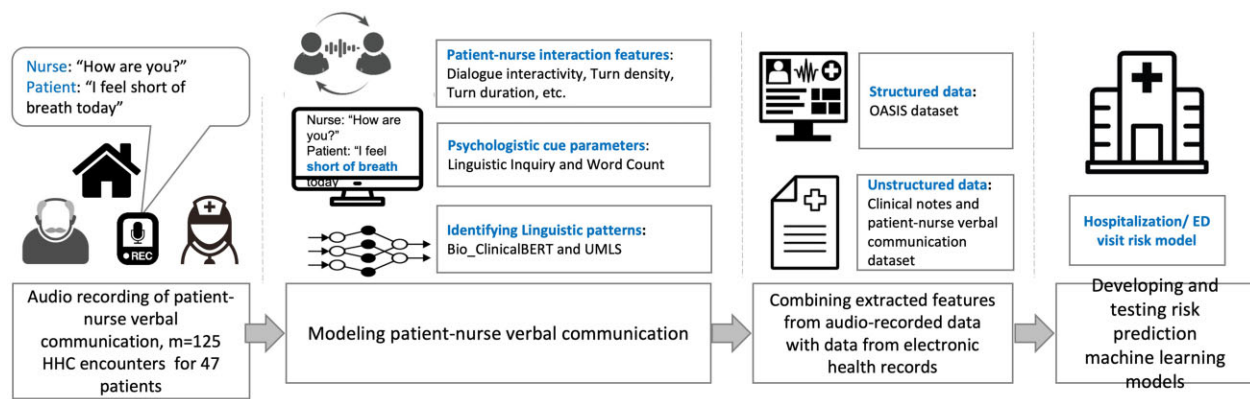


Figure 1. Overview of the methodology of the study.

variants (eg, Bio_ClinicalBERT¹²) are deep learning models tailored for sequence data, especially text. These models have transformed the field of NLP because of their proficiency in modeling the contextual relationship among words in the context. They have been trained on vast datasets to recognize linguistic patterns and can then transfer this knowledge to smaller, specific datasets, improving nuanced interpretations and precise analyses of the text.

- **Feature generation using Bio_ClinicalBERT:** In this study, we used Bio_ClinicalBERT,¹² an extension version of the BERT model to process the conceptual relationships between medical concepts in clinical notes. This model processes the relationships and translates them into feature vectors or embeddings, which represent these relationships in a format suitable for machine learning (ML) algorithms. Bio_ClinicalBERT has been trained on a large medical corpus from PubMed, PMC (PubMed Central), and MIMIC III clinical notes. Previous studies have shown that it can outperform other extensions of the BERT models for health-related NLP tasks, such as identifying patients at risk of negative outcomes.^{13,14}
- **Feature generation using Unified Medical Language System (UMLS):** Along with Bio_ClinicalBERT, we used UMLS to extract references to medical concepts and activities (eg, headache or Neurological assessment test) from clinical notes. UMLS is a compendium of more than 100 standard terminologies in the biomedical sciences, which maps more than 1.5M medical concepts across these terminologies.^{15,16} Synonyms concepts were linked to unique identifiers (CUIs). Each CUI is designated a particular name and semantic type, like “Headache,” which falls under the semantic type “Sign/symptom.” We used the QuickUMLS tool¹⁷ to identify the representation of medical terms and activities within clinical notes and their associated semantic types. The presence, frequency, or absence of specific UMLS concepts and their semantic types can serve as distinct features to inform development of ML models in predicting ED visits.

Audio-recorded patient-nurse verbal communication dataset

During our study, we audio-recorded a total of 126 encounters with 47 participating patients. Of these patients, 18 (38%) had a single recording, 15 (32%) had 2 recordings,

and 14 (30%) had 3 or more recordings. Overall, the 126 encounters consist of 8099 utterances, with 4016 and 4083 utterance for patients and nurses, respectively.

- **Modeling the patient-nurse interaction:** All 126 audio-recorded encounters were transcribed using AWS-GT. Each transcription consisted of patients’ and nurses’ utterances, alternating throughout the encounter, with each utterance associated with a corresponding start and end time. We leveraged this data to model patient-nurse interaction, using readily quantifiable social interaction parameters,^{18–20} including the total number of patient turns speaking, dialogue interactivity, turn density, turn duration, and relative timing of turns. This information can characterize patterns of the patient’s interaction that may indicate subtle signals of communication deficits in elderly patients. These deficits could be linked to negative outcomes.^{18–20}
- **Modeling the psycholinguistic cues in patient and nurse language using Linguistic Inquiry and Word Count (LIWC) 2015:** LIWC 2015 is a manually curated lexical-based NLP tool developed by experts in the psychology of language. It contains a large selection of commonly used words and terms in daily communication, which are organized into 11 top-level categories, including function words, affective processes, social processes, cognitive processes, perceptual processes, biological processes, drives, relativity, informal language, personal concerns, and time orientation. Past research has demonstrated LIWC’s effectiveness in identifying psycholinguistic cues associated with negative outcomes,^{21,22} such as ED visits and hospitalization.
- **Feature generation using Bio_ClinicalBERT and UMLS:** Similar to the analysis of clinical notes, we used Bio_ClinicalBERT to model the conceptual relationship between words in patient-nurse verbal communication. This process yielded word embedding features that characterize the unique linguistic patterns used by patients and nurses. Word embedding features is a type of numerical representation of text where words or phrases from the vocabulary are mapped to vectors of real numbers. The hypothesis underlying our approach is that the language dynamics in patient-nurse interactions could potentially vary based on the level of patient risk, and therefore, Bio_ClinicalBERT can capture these nuanced variations, assisting in

identifying high-risk patients. Also, we used UMLS to identify the medical concepts conveyed in patient-nurse verbal communication.

Outcome variable

The outcome of interest for this study is the incidence of ED visits and hospitalizations recorded in the home care EHR within a 60-day period post-admission into home healthcare. Of 47 patients, 8 experienced an ED visit or hospitalization within this timeframe.

Building ML models

- **ML algorithms:** To build the risk identification model for patients at risk of ED visit and hospitalization, we used different ML classifiers, including Logistic Regression, ensemble decision trees, specifically: Random Forest, Extra Trees, Adaptive Boosting, Gradient Boosting Machine, XGB boost, and support vector machine (SVM). With the small sample size of this study, we decided not to use deep neural network models to avoid potential overfitting. In ML, a classifier is a supervised algorithm that leverages labeled training data to learn patterns and relationships within that data. Once trained, classifiers can be used to categorize or label new, unseen data into specific classes or categories.
- **Feature selection:** The ML algorithms were trained on the most informative features selected using the Joint Mutual Information Maximization (JMIM)²³ method. JMIM selects a subset of features by maximizing the joint mutual information between the selected features and the outcome class while minimizing redundancy among the selected features. JMIM has been shown to have a high generalization ability, especially on small samples with many generated features.^{23,24} See [Appendix SB](#) for more information about the JMIM.
- **Model evaluation:** Leave-One-Out Cross Validation²⁵ (LOOCV) with standard performance metrics of area under the curve-receiver operating characteristic (AUC-ROC) and F1-score (the harmonic mean of sensitivity and precision) used to evaluate and report the performance of ML classifiers. LOOCV is a cross-validation method often utilized with smaller datasets. It iteratively uses one sample for validation and the rest for training, repeating this for each data point. This method gives an unbiased estimation of model generalizability by averaging the error rate across all iterations.
- We iteratively constructed ML models, employing the OASIS dataset as the baseline, complemented by the combination of clinical notes and audio-recorded patient-nurse encounters. We hypothesize that the integration of audio-recorded encounters with OASIS and clinical notes can significantly improve the performance of ML models in predicting patients at risk of ED visits and hospitalization.

Approach (1): building ML models using all data in the sample

Unlike the OASIS dataset, which is completed a finite number of times, the number of clinical notes generated for each patient depends on the number of scheduled homecare visits. Furthermore, for each patient in our sample, the number of

audio-recordings varied from one to more than 3 encounters. As part of our approach to building the ML models, we compiled a single document containing all existing clinical notes for each patient and another separate document compiling all their transcribed patient-nurse encounters. This strategy enabled the inclusion of all available patient clinical data into the ML model, which may enhance the model's ability to identify patients at risk of negative outcomes.

Approach (2): building ML models using data from the most recent audio-recorded encounter

Additionally, to ensure data consistency across the patient sample, our secondary approach focused on considering only the most recent audio-recorded encounter and its corresponding clinical notes for each patient. Overall, we hypothesize that incorporating a smaller quantity of clinical notes and audio-recorded encounters into our models, such as considering the data from the most recent encounter for each individual patient, could potentially diminish the ML algorithms' performance in predicting patients at risk of ED visits and hospitalization.

It is noteworthy that due to the time lapse between a patient's home healthcare admission and obtaining their consent for this study, none of the recorded encounters pertained to the admission visit. Consequently, neither the audio recordings nor the clinical notes from the most recent encounters addressed the admission visit. Yet, when aggregating all available patient data (for the first approach), the clinical notes from the admission visit were incorporated.

Results

[Table 1](#) describes the attributes of study participants with a nearly equal male-female ratio. Approximately 50% of the participants identify as Black and 30% as White. Less than half lived alone, while over 70% required assistance with daily activities, especially those with ED visits or hospitalizations. In health characteristics, 46% were at fall risk, 14% had vision impairment, 14% had cancer, 7% had diabetes, and 3 patients had cardiovascular disease. This profile underscores diverse health needs within the study group. None of the patients with ED visits or hospitalizations had private insurance.

[Table 2](#) presents descriptive statistics on the duration of audio-recorded patient-nurse encounters and spoken words by patients and nurses at the encounter level. On average, an encounter lasted 19 min, with a quarter under 12 min. Encounters contained an average of 63 utterances, with the median also being 56. Nurses typically spoke more words than patients during these encounters, a trend seen at both encounter and utterance levels. The median number of spoken words during an encounter was 589 for patients and 842 for nurses, underscoring nurses' dominant role in this verbal communication.

The results of the top-performing ML model on the OASIS dataset are presented in [Table 3](#) as the baseline model along with additional models adding the combination of clinical notes and audio-recorded patient-nurse encounters. The evaluation includes data from the most recent encounter and the available data for each patient (clinical notes and audio data) in the sample.

Based on only the OASIS dataset, the initial risk identification model yielded a modest F1-score of 48.01. However, by

Table 1. Clinical characteristics of the study's cohort.

	Sample N = 47	ED/hospitalization=yes N = 8	ED/hospitalization=no N = 39
Age			
Gender			
Female	24(51%)	4(50%)	20(48%)
Male	23(49%)	4(50%)	19(48%)
Race			
Asian	4(8%)	0(-)	4(10%)
Black	26(56%)	4(50%)	22(56%)
Hispanic	3(6%)	1(12%)	2(5%)
White	14(30%)	3(37%)	11(28%)
Insurance			
Patient has private insurance	10(21%)	0(-)	10(25%)
Patient has managed Medicare	16(34%)	5(62%)	11(28%)
Patient has Medicaid	2(4%)	0(-)	2(5%)
Patient has FFS Medicare	15(31%)	2(25%)	13(33%)
Living alone: yes	21(44%)	4(50%)	17(43%)
Status of most problematic surgical wound that is observable			
Newly epithelialized	27(58%)	6(75%)	21(53%)
Fully granulating	5(10%)	0(-)	5(12%)
Early/partial granulation	2(4%)	2(25%)	0(-)
Not healing	13(27%)	0(-)	13(33%)
Frequency of pain			
Sometimes but do not interfere with daily activities	13(22%)	3(37%)	10(25%)
Daily but not consistently	33(71%)	5(62%)	28(71%)
Always	1(2%)	0(-)	1(2%)
Ambulation/locomotion			
Able to walk independently	2(4%)	0(-)	2(5%)
With the use of a handed device	5(11%)	0(-)	5(15%)
Able to walk only with the supervision	35(74%)	6(75%)	27(69%)
Chairfast/bedfast	5(11%)	2(25%)	3(10%)
Other clinical conditions			
Presence of renal disease: yes	5(10%)	2(25%)	3(7%)
Risk of fall: yes	22(46%)	5(62%)	17(43%)
Impaired vision: yes	7(14%)	2(25%)	5(12%)
Patients have cancer: yes	7(14%)	1(12%)	6(15%)
The patient has a surgical wound: yes	23(48%)	3(37%)	20(51%)
Patient received instruction on special precautions for all high-risk medications: yes	43(91%)	7(87%)	36(92%)
Diabetes with chronic complications: yes	8(17%)	4(50%)	4(10%)
peripheral vascular disease: yes	3(6%)	1(12%)	2(5%)
Frequency of disruptive behavior symptoms			
Never	45(95%)	8(100%)	37(94%)
Once or less than once in a month	2(5%)	0(-)	2(5%)
Activities of daily living (ADL)for instrumental activities of daily living (IADL)			
Completely independent	1(2%)	0(-)	1(2%)
A level of assistance required	11(23%)	0(-)	11(28%)
Completely dependent	35(75%)	8(100%)	27(69%)
Management of oral medications			
Able to independently take the correct oral medication(s)	4(9%)	0(-)	4(10%)
Able to take medication(s) at the correct times with assistance from caregivers	10(21%)	2(25%)	8(20%)
Unable to take medication unless administered by another person	33(70%)	6(75%)	27(69%)
Feeding or eating ability			
Able to independently feed self.	9(19%)	0(-)	9(76%)
Be assisted or supervised throughout the meal/snack	38(81%)	8(100%)	30(76%)
Unable to take in nutrients orally	0(-)	0(-)	0(-)

incorporating clinical notes features from the most recent encounter, the model's F1-score improved significantly to 73.68, marking a 53.47% improvement. Furthermore, when we integrated features from all available patient clinical notes, the F1-score rose to 75.01, reflecting a 56.24% improvement.

The risk identification model's performance significantly improved when patients' verbal communication features were

added. Specifically, there was an F1-score increase to 85.72 with the most recent encounter data and 87.5 when all audio-recorded encounters were incorporated. This is a 19% performance boost compared to the model built only on clinical notes and OASIS datasets.

Adding speech features from the nurses' verbal communication further enhanced the model's F1-score. This became

Table 2. Statistics on duration, utterance frequency, and spoken word counts in patient-nurse verbal communication.

	Average (standard deviation)	25% quartile	50% quartile	75% quartile
Duration of audio-recorded patient-nurse encounters (in min)	19 (9)	12	17	23
Count of utterances per encounter	63 (49)	31	56	80
Count of utterances for each patient per encounter	30 (24)	15	27	30
Count of utterances for each nurse per encounter	31 (24)	15	28	40
Count of spoken words (tokens) in the sample	830 (608)	406	686	1090
Count of spoken words (tokens) by patients during an encounter	690 (501)	303	589	960
Count of spoken words (tokens) by nurses during an encounter	969 (672)	461	842	1218

Table 3. Top-performing ML model on the OASIS dataset, as the baseline, along with the combination of clinical notes and audio-recorded patient-nurse verbal communication.

Feature generation methods	Best performing algorithm	AUC-ROC	F1-score
Sample: N = 47			
Baseline dataset			
OASIS dataset	XG-boost	67.63	48.01
Combination of OASIS and clinical notes and audio-recorded encounter for the most recent encounter			
OASIS+features extracted from clinical notes	SVM-RBF	79.17	73.68
OASIS+features extracted from clinical notes+features extracted from the patient's speech during an encounter	SVM-RBF	94.55	85.72
OASIS+features extracted from clinical notes+features extracted from the patient's speech during an encounter+the nurse's speech during an encounter	SVM-RBF	96.15	87.5
Combination of OASIS and clinical notes and audio-recorded encounters for all available encounters			
OASIS+features extracted from clinical notes	SVM-RBF	86.54	75.01
OASIS+features extracted from clinical notes + features extracted from the patient's speech during an encounter	XGB	96.79	87.5
OASIS+features extracted from clinical notes+features extracted from the patient's speech during an encounter+the nurse's speech during an encounter	SVM-RBF	99.68	94.12

especially significant when all audio-recorded interactions were incorporated, raising the F1-score to 94.12, a notable 26% performance increase compared to the model built only on clinical notes and OASIS datasets.

These findings highlight the significance of patient-nurse communication in risk prediction models for home healthcare patients, emphasizing that this missing critical dataset provides important health insights not currently captured by structured EHR and clinical notes alone.

These findings also show that enriching the model with additional data, such as clinical notes and audio data from multiple encounters for a subgroup of patients in the sample, can notably improve the risk identification model's performance. As we found, models built solely on data from the most recent encounter demonstrated a lower performance (F-score = 87.5) compared to models constructed using all available audio and clinical notes in the sample (F-score = 94.12). This suggests that ML models are capable of effectively managing variations in data availability among patients, thereby implying that enriching data, even if only for a subgroup of patients in the sample, boosts their predictive performance.

Figure 2A and B are based on JMIM analysis of LIWC features for patient and nurse language and highlight key communication features associated with increased risk of ED visits and hospitalization. High-risk patients tend to interact more with risk-associated cues, ask more questions (indicated by keywords: "question marks" and "interrogative"), and exhibit more "sadness" and "anxiety," possibly signaling health deterioration. Discussions about their familial and

social circumstances were also more frequent for patients at risk. Conversely, nurses' communication with such patients typically leans toward instructional. This is evident in their choice of phrases, employing comparative terms such as "better than," "less effective," or "similar to," and quantifiers like "many," "few," "several," or "a substantial amount." The nurses often displayed more empathy, indicated by "positive emotions" and "affective language" keywords. Also, the term "money" often emerges, likely to refer to treatment expenses or address the financial concerns of high-risk patients. Figure 2C, informed by the JMIM analysis of turn-taking features, indicates lengthier encounters for patients with an increased likelihood of ED visits and hospitalizations. The communication between these patients and nurses generally extends over a longer duration and displays a greater coefficient of variation (CV) in their speaking times. The CV is a statistical measure that describes the relative variability of data in relation to its mean. The importance of this measure, especially in context of patient-nurse communication, is that it can highlight inconsistencies in communication durations which can be a clinically important predictor in identifying patients at risk of ED visits and hospitalization. High-risk patients also undergo extended periods of silence, potentially due to their health conditions hindering efficient communication. Informed by JMIM's UMLS semantic type analysis, Figure 2D identifies 5 main features associated with increased ED visits and hospitalization risk. These features focus on therapeutic/preventive measures, body substances (eg, blood, hormones), manufactured objects (eg, medical devices), lab

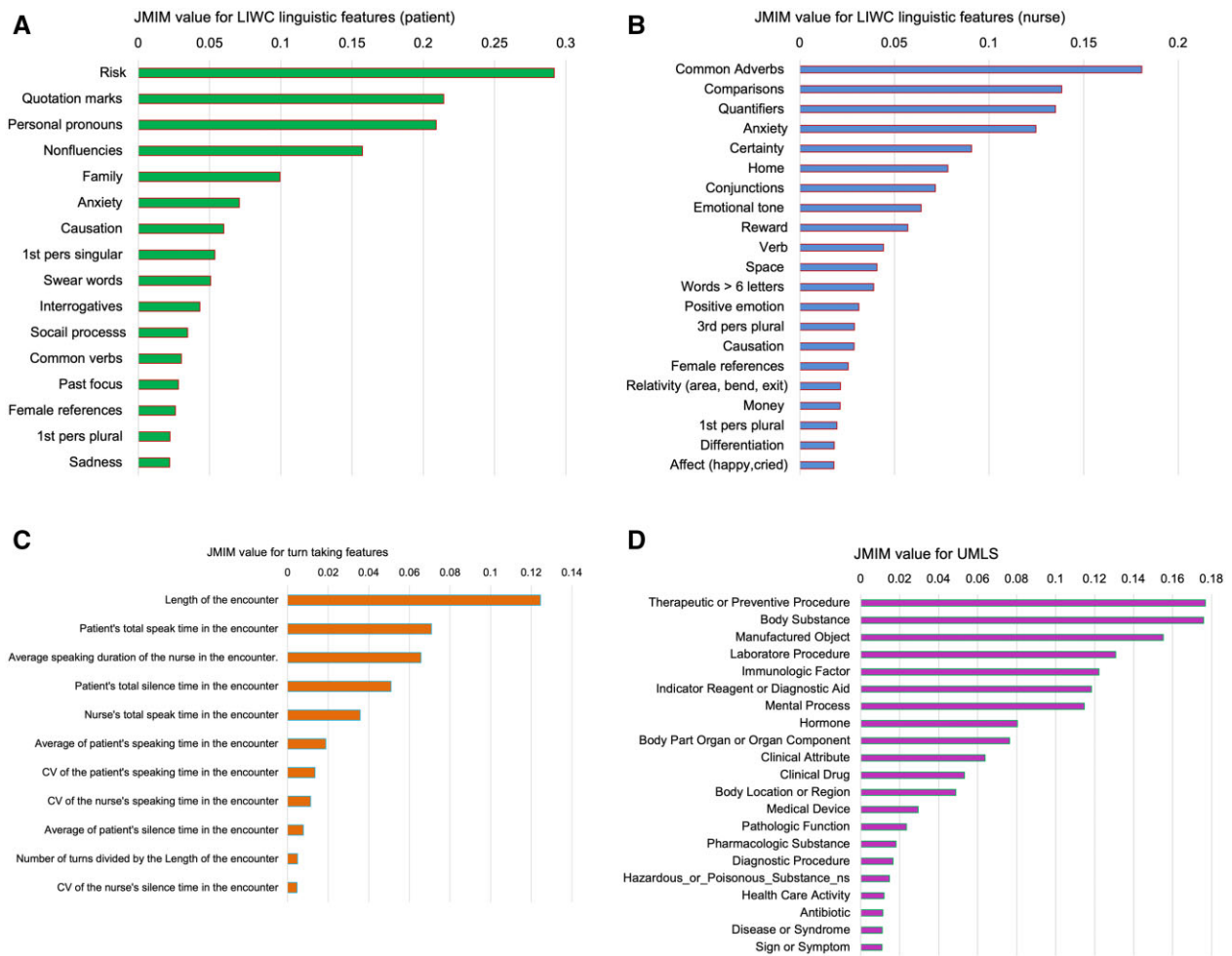


Figure 2. (A) The most informative features identified using JMIM analysis of LIWC features associated with increased risk of ED visits and hospitalization in patient language. (B) The most informative features identified using JMIM analysis of LIWC features associated with increased risk of ED visits and hospitalization in nurse language. (C) The most informative features identified using JMIM analysis of turn-taking features associated with increased risk of ED visits and hospitalization. (D) The most informative features identified using JMIM analysis of UMLS semantic type features associated with increased risk of ED visits and hospitalization.

procedures, and immunologic factors. As these features form essential aspects of patient care, they might naturally dominate communication between nurses and high-risk patients. Communications also involved guidance on medication and discussions about mental health issues, indicative of the potentially escalated risk of adverse outcomes for high-risk patients.

Discussion

As demonstrated in this study, incorporating patient speech with an initial model built on EHR data and clinical notes significantly improved the model's performance by 17%. After integrating the nurse's speech into the model, we achieved a further performance enhancement, improving the model's F1-score by 26%. This underscores the importance of integrating patient and nurse language into risk prediction models, indicating the need for a more inclusive approach to modeling patient-nurse verbal communication. This study is the first to investigate the added values of verbal patient-nurse communication in home healthcare to enhance a risk model for predicting ED visits and hospitalizations. While free-text clinical notes and structured EHR data help identify patients at risk,

they may miss critical information shared during patient-nurse conversations. Critical information like a patient's financial concerns or emotional instability during communication might not be recorded in the EHR, limiting the performance of risk prediction models.²

The language used by patients and nurses in healthcare settings plays a critical role in understanding and improving patient outcomes. This is particularly true for patients at higher risk of negative outcomes, who often require a more focused and preventive approach to their healthcare. Our study showed that nurses' communication with this group of patients often centers around key components of their care plan, lab tests, medication, usage of medical equipment, and discussing issues like mental health concerns. The automatic identification of major clinical themes of communication, particularly for nurses, can be effectively conducted using standardized medical terminologies such as UMLS. Conversely, the informal nature of patient language requires alternative analytic methods, like the LIWC lexical dictionary, which decodes everyday terms to comprehend the underlying psychology of the patient's language. LIWC can detect social, cognitive, and emotional cues and elements relating to family and financial concerns within a conversation. Furthermore,

LIWC can identify language disfluencies, which could suggest underlying cognitive, psychological, or physiological distress. As demonstrated in our study, these identified cues may be linked to an increased risk of negative health outcomes. Attention to the dynamics of verbal communication, such as turn-taking patterns and periods of silence, can further enrich our understanding of patients at high risk. As our research indicated, patients at a higher risk of ED visits and hospitalization often exhibit longer pauses and more prolonged periods of silence.

Transformer models, like BERT and its extension versions like Bio_ClinicalBERT, are important methods for analyzing clinical notes due to their advanced understanding of language semantics. The bidirectional architecture of the models allows for the understanding of word context by interpreting words in both directions, a feature particularly vital in healthcare, where the context of medical terminology often shapes meaning. For example, clinical note that reads, “The patient denies cold symptoms but has a history of cold sores,” the first instance of “cold” refers to a common illness with symptoms like a runny nose or cough, whereas the second instance refers to outbreaks caused by the herpes virus. By processing words bidirectionally, the model can better distinguish between these 2 contexts, enhancing its accuracy in interpreting and representing the meaning of medical terminology. By leveraging the power of transfer learning, Bio_ClinicalBERT, pre-trained on an extensive collection of clinical text, is equipped to recognize and generalize linguistic patterns. This is particularly beneficial for capturing the nuances of complex medical language for high-risk patients and when the model needs to be built using smaller datasets while maintaining high generalizability. As indicated in our results, integrating Bio_ClinicalBERT’s word-embedding and UMLS features boosted the risk identification model’s performance by 56%, compared to the initial model built solely on the OASIS dataset. We also used Bio_ClinicalBERT to model the conceptual relationship between words in utterances of patient-nurse verbal communication. Previous studies showed that BERT and its extended versions could model the nuance in semantic and syntactic levels of language organization of patients’ speech, such as disfluency, to identify those at risk of negative outcomes.^{26,27}

Analysis of patient-clinician verbal communication has been a focal point in various studies investigating patient-clinician communication. For instance, Drew et al.¹⁹ conducted a study analyzing patient-practitioner communication to identify practitioners’ communication patterns in addressing patient concerns. In another study, Mejdahl et al.²⁸ analyzed patient-clinician communication in epilepsy outpatient clinics to explore the impact of patients’ self-reported data on the outcome. These studies, along with several other studies published in the field of conversational analysis,^{29–31} highlight the necessity of content analysis in audio-recorded patient-clinician verbal communication. In all these studies, the content of verbal communication was manually annotated to identify the themes of conversation, mainly using coding systems such as Roter’s Interaction Analysis System (RIAS).³² The coding systems need the involvement of annotators to annotate cues within conversations, focusing on elements like displays of concern, instances of disfluency, and information exchanged. While the manual annotation can shed light on the primary themes of a conversation, it necessitates a significant investment of time and effort due to its labor-intensive and time-consuming nature. Our study demonstrates NLP’s

significant potential for automatically modeling patient-nurse verbal communication for building risk identification models. In our upcoming research, we intend to investigate how the coding systems such as RIAS can improve the performance of risk identification models.

Audio recording of patient-nurse verbal communication is not currently part of the clinical workflows. We conducted a series of pilot studies to identify convenient procedures for audio-recording patient-nurse verbal communication.¹ The findings of the studies showed that both patients and nurses were comfortable with the audio recording procedure, and patients indicated that they recording had the potential for personal use (eg, reviewing the clinician’s instruction).¹ The involvement of healthcare stakeholders, especially clinicians and managers, plays a crucial role in effectively implementing the speech processing system. Their active participation is essential for determining the integration of audio recording into clinical workflows and establishing the necessary processing methods for continued use in patient care management.^{33,34}

Our study indicates that integrating audio-based risk prediction models into home healthcare workflows could expedite the identification of at-risk patients. Such models have the potential to underpin clinical decision support systems, thereby enabling healthcare providers to swiftly detect deteriorating patient conditions and take timely interventions. However, we acknowledge the financial and logistical hurdles associated with the adoption of speech processing systems, particularly for home healthcare agencies constrained by narrow profit margins. The implementation of these audio-recording systems would entail an upfront investment in hardware and software and recurring expenses for data storage and management. Moreover, the technical expertise required for system deployment and ongoing maintenance may be beyond the in-house capabilities of many agencies. To address these challenges, we recommend conducting in-depth feasibility studies, supplemented by pilot implementations, to evaluate both the practicality and the potential impact on patient outcomes. Collaborative efforts with technology vendors or the utilization of government incentives may offer viable solutions to mitigate these financial and technical obstacles.

Limitations

The study has several limitations. First, we relied on a relatively modest sample size of 126 audio-recorded patient-nurse conversations for 47 patients, all collected from one of the US’s largest non-profit home healthcare organizations. Second, while using a pre-trained Clinical BERT model provided the advantage of transfer learning and using a leave-one-out validation technique offered an unbiased view of our model, the single-source nature of our data could potentially limit the overall generalizability of our findings. Furthermore, we did not explore the effectiveness of other feature-generation methods and ML classifiers. In the future, it will be beneficial to expand the research to involve data from multiple sites, investigate the performance of other feature-generation methods and ML classifiers such as BiLSTM with an attention layer, and further assess the system’s usability in the home healthcare setting.

Conclusion

In conclusion, this innovative study underscores the immense value of incorporating patient-nurse verbal communication in

enhancing risk prediction models for hospitalizations and ED visits. We significantly improved the model's performance by integrating such communication with existing EHR data and clinical notes. The study paves the way for a more effective approach to risk identification, suggesting a need for an evolved clinical workflow that integrates routine patient-nurse verbal communication recording.

Author Contributions

M.Z.: Contribution to the conception, study design, data acquisition, data analysis, drafting the manuscript. S.V.: Data acquisition and data annotation. S.S.: Data analysis and interpretation. A.Z.: Data analysis and interpretation. M.V.M.: Data acquisition and reviewing the manuscript critically for important intellectual content. K.H.B.: Data acquisition and reviewing the manuscript critically for important intellectual content. M.T.: Contributions to the conception, data acquisition, and reviewing the manuscript critically for important intellectual content.

Supplementary material

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

Funding

This study was supported by K99AG076808 and R01AG081928 from National Institute on Aging; Amazon in collaboration with Columbia University Center of AI Technology; NVS Health Doyle Fund for pilot studies; and Columbia University School of Nursing Pilot Award.

Conflicts of interest

None declared.

Data availability

Due to the limitations imposed by the IRB protocol of VNS Health, no data is available for public access.

References

- Zolnoori M, Vergez S, Kostic Z, et al. Audio recording patient-nurse verbal communications in home health care settings: pilot feasibility and usability study. *JMIR Hum Factors*. 2022;9(2):e35325.
- Song J, Zolnoori M, Scharp D, et al. Do nurses document all discussions of patient problems and nursing interventions in the electronic health record? A pilot study in home healthcare. *JAMIA Open* 2022;5(2):ooac034.
- Romagnoli KM, Handler SM, Hochheiser H. Home care: more than just a visiting nurse. *BMJ Qual Saf*. 2013;22(12):972-974.
- Shang J, Russell D, Dowding D, et al. A predictive risk model for infection-related hospitalization among home healthcare patients. *J Healthc Qual*. 2020;42(3):136-147.
- Zolnoori M, McDonald MV, Barrón Y, et al. Improving patient prioritization during hospital-homecare transition: protocol for a mixed methods study of a clinical decision support tool implementation. *JMIR Res Protoc*. 2021;10(1):e20184.
- Topaz M, Trifilio M, Maloney D, Bar-Bachar O, Bowles KH. Improving patient prioritization during hospital-homecare transition: a pilot study of a clinical decision support tool. *Res Nurs Health*. 2018;41(5):440-447.
- Petti U, Baker S, Korhonen A. A systematic literature review of automatic Alzheimer's disease detection from speech and language. *J Am Med Inform Assoc*. 2020;27(11):1784-1797.
- Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: a systematic review. *Laryngoscope Investig Otolaryngol*. 2020;5(1):96-116.
- Schneider S, Baevski A, Collobert R, Auli M. wav2vec: unsupervised pre-training for speech recognition. arXiv Prepr. arXiv1904.05862. 2019, preprint: not peer reviewed.
- Zolnoori M, Vergez S, Sridharan S, et al. Is the patient speaking or the nurse? Automatic speaker type identification in patient-nurse audio recordings. *J Am Med Inform Assoc*. 2023;30(10):1673-1683.
- Devlin J, Chang M-W, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. arXiv Prepr. arXiv1810.04805. 2018, preprint: not peer reviewed.
- Hugging Face, @. ClinicalBERT – Bio+clinical BERT model. *Hugging Face*. 2023. Accessed March, 2023. https://huggingface.co/emilyalsentzer/Bio_ClinicalBERT.
- Ahsan H, Ohnuki E, Mitra A, You H. MIMIC-SBDH: a dataset for social and behavioral determinants of health. In: *Machine Learning for Healthcare Conference*. PMLR; 2021:391-413.
- Yao Z, Tsai J, Liu W, et al. Automated identification of eviction status from electronic health record notes. *J Am Med Inform Assoc*. 2023;30(8):1429-1437.
- Boyd AD, Dunn Lopez K, Lugaresi C, et al. Physician nurse care: a new use of UMLS to measure professional contribution: are we talking about the same patient a new graph matching algorithm? *Int J Med Inform*. 2018;113:63-71.
- Lange LL. Representation of everyday clinical nursing language in UMLS and SNOMED. In: *Proceedings of the AMIA Annual Fall Symposium*. American Medical Informatics Association; 1996:140.
- Soldaini L, Goharian N. QuickUMLS: a fast, unsupervised approach for medical concept extraction. In: *MedIR Workshop*. SIGIR; 2016:1-4.
- Roter DL, Larson SM, Beach MC, Cooper LA. Interactive and evaluative correlates of dialogue sequence: a simulation study applying the RIAS to turn taking structures. *Patient Educ Couns*. 2008;71(1):26-33.
- Drew P, Chatwin J, Collins S. Conversation analysis: a method for research into interactions between patients and health-care professionals. *Health Expect*. 2001;4(1):58-70.
- Dodge HH, Mattek N, Gregor M, et al. Social markers of mild cognitive impairment: proportion of word counts in free conversational speech. *Curr Alzheimer Res*. 2015;12(6):513-519.
- Bahgat M, Wilson S, Magdy W. LIWC-UD: classifying online slang terms into LIWC categories. In: *14th ACM Web Science Conference*. Association for Computing Machinery, New York, NY, United States; 2022:422-432.
- Belz FF, Adair KC, Proulx J, Frankel AS, Sexton JB. The language of healthcare worker emotional exhaustion: a linguistic analysis of longitudinal survey. *Front Psychiatry*. 2022;13:1044378.
- Bennasar M, Hicks Y, Setchi R. Feature selection using joint mutual information maximisation. *Expert Syst Appl*. 2015;42(22):8520-8532.
- Varatharajah Y, Ramanan VK, Iyer R, Vemuri P; Alzheimer's Disease Neuroimaging Initiative. Predicting short-term MCI-to-AD progression using imaging, CSF, genetic factors, cognitive resilience, and demographics. *Sci Rep*. 2019;9(1):2235.
- Vazquez-Zapian GJ, Mata-Miranda MM, Garibay-Gonzalez F, Sanchez-Brito M. Artificial intelligence model validation before its application in clinical diagnosis assistance. *World J Gastroenterol*. 2022;28(5):602-604.
- Rocholl JC, Zayats V, Walker DD, Murad NB, Schneider A, Liebling DJ. Disfluency detection with unlabeled data and small BERT models. arXiv Prepr. arXiv2104.10769. 2021, preprint: not peer reviewed.
- Tian Y, Nieradzick T, Jalali S, Shiu D. How does BERT process disfluency? In: *Proceedings of the 22nd Annual Meeting of the Special*

- Interest Group on Discourse and Dialogue*. Association for Computational Linguistics; 2021:208-217.
28. Mejdahl CT, Schougaard LMV, Hjollund NH, Riiskjær E, Lomborg K. Patient-reported outcome measures in the interaction between patient and clinician – a multi-perspective qualitative study. *J Patient Rep Outcomes*. 2020;4:1-10.
 29. Halpin SN, Konomos M, Roulson K. Using applied conversation analysis in patient education. *Glob Qual Nurs Res*. 2021;8:23333936211012990.
 30. Pino M, Doehring A, Parry R. Practitioners' dilemmas and strategies in decision-making conversations where patients and companions take divergent positions on a healthcare measure: an observational study using conversation analysis. *Health Commun*. 2021;36(14):2010-2021.
 31. Jones A. Nurses talking to patients: exploring conversation analysis as a means of researching nurse-patient communication. *Int J Nurs Stud*. 2003;40(6):609-618.
 32. Stortenbeker I. Coding linguistic elements in clinical interactions: a step-by-step guide for analyzing communication form. *BMC Med Res Methodol*. 2022;22:191.
 33. Ball SL, Weiner SJ, Schwartz A, et al. Implementation of a patient-collected audio recording audit & feedback quality improvement program to prevent contextual error: stakeholder perspective. *BMC Health Serv Res*. 2021;21:1-11.
 34. Smith SM, Stelmar J, Lee G, Carroll PR, Garcia MM. Use of voice recordings in the consultation of patients seeking genital gender-affirming surgery: an opportunity for broader application throughout surgery? *J Surg Res*. 2022;5:618-625.