Extraction of Sleep Information from Clinical Notes of Alzheimer's Disease Patients Using Natural Language Processing

Haneef Ahamed Mohammad[1*], Sonish Sivarajkumar[2*], Samuel Viggiano[1], David Oniani[1], Shyam Visweswaran, MD, PhD[2,3], Yanshan Wang, PhD[1,2,3†]
[1]Department of Health Information Management, University of Pittsburgh, Pittsburgh, PA; [2]Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA; [3]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA

Abstract

Alzheimer's Disease (AD) is the most common form of dementia in the United States. Sleep is one of the lifestyle-related factors that has been shown critical for optimal cognitive function in old age. . However, there is a lack of research studying the association between sleep and AD incidence. A major bottleneck for conducting such research is that the traditional way to acquire sleep information is time-consuming, inefficient, non-scalable, and limited to patients' subjective experience. In this study, we developed a rule-based NLP algorithm and machine learning models to automate the extraction of sleep-related concepts, including snoring, napping, sleep problem, bad sleep quality, daytime sleepiness, night wakings, and sleep duration, from the clinical notes of patients diagnosed with AD. We trained and validated the proposed models on the clinical notes retrieved from the University of Pittsburgh Medical Center (UPMC). The results show that the rule-based NLP algorithm consistently achieved the best performance for all sleep concepts.

Introduction

Alzheimer's Disease (AD) is the most common form of dementia in the United States (U.S.), which affects at least 5.7 million Americans with a projected increase to 13.8 million by mid-century due to a global aging population[1-3]. In 2015, official death certificates recorded more than 110 thousand deaths from AD, making it the sixth leading cause of death in the U.S. and the fifth leading cause of death in Americans 65 years of age or older[1]. Unlike deaths from stroke and heart disease, which decreased between 2000 and 2015, deaths from AD increased 123%. By the year 2050, 13.8 million Americans are expected to have AD with an associated cost of $1.2 trillion U.S. dollars, not including unpaid caretaker hours. Postponing dementia onset by even one year could result in nine million fewer cases worldwide than predicted by 2050[4] and a reduction in care costs. Therefore, early intervention to reduce the risk of AD will have a better population health impact.

Social and behavioral determinants of health (SDOH) are modifiable factors and offer opportunities for reducing the risk of AD[5]. Sleep is one of the lifestyle-related SDOH factors that has been shown critical for optimal cognitive function in old age[6]. However, the association between sleep and AD incidence is complex, as shown in the literature. On the one hand, previous studies suggest that sleep problems, such as insomnia[7], excessive daytime sleepiness[8, 9], snoring[10], sleep duration[11], poor sleep quality[12], and difficulties maintaining sleep[11], are associated with an increased risk of incident cognitive impairment and could be an early predictor of future AD dementia[13]. On the other hand, some studies find no association between sleep variables (e.g., sleep duration, sleep difficulties, and snoring) and cognitive function[14]. Moreover, a bi-directional relationship is seen between sleep and cognitive function decline in the elderly with underlying AD; in other words, AD also causes circadian and sleep disturbances. Despite a growing interest in studying sleep-AD relationship, longitudinal epidemiological research in a large cohort is still needed to understand the relationship. A major bottleneck for conducting such research is that the traditional way to acquire sleep and AD data through multi-year follow-ups is time-consuming, inefficient, non-scalable, and limited to patients' subjective experience.

Large volumes of electronic health records (EHRs) collected by healthcare organizations offer an opportunity to use a large sample size to investigate intervention outcomes in routine care, such as predictors of response, safety, comparative effectiveness, and health economic evaluations[15]. EHRs have become popular in AD research, such as resource use in AD care[16], comorbidities[17], case capture efficiency[18], and health disparities[19]. However, EHRs remain

[*] Co-first authors
[†] Corresponding author: Yanshan Wang, PhD, FAMIA; yanshan.wang@pitt.edu

underused in collecting sleep information for AD research. A major barrier that hinders the use of sleep information from EHRs for AD research is that most of the sleep information in EHRs is embedded in clinical narratives. Natural language processing (NLP), a technique in computational linguistics that uses computational models for understanding natural language, has been used to extract meaningful information from clinical narratives[20]. However, there existed no NLP algorithms in the literature to extract sleep information from clinical notes, particularly for patients with AD, to the best of our knowledge. In this study, we developed both rule-based NLP algorithm and machine learning-based NLP models to automate the extraction of sleep-related concepts, including snoring, napping sleep problem, poor sleep quality, daytime sleepiness, night wakings, and sleep duration, from the clinical narratives of patients diagnosed with AD. We trained and validated the proposed models on the clinical notes retrieved from the University of Pittsburgh Medical Center (UPMC). The results show that the rule-based NLP algorithm achieved consistently the best performance for all sleep concepts.

Background

Several prior studies have used the EHR data for the sleep information, most of which utilized structured EHR data, such as the International Classification of Diseases (ICD) diagnostic codes. For example, Felder et al. used ICD codes to identify sleep disorders and study the association between sleep disorders and preterm birth[21]. Hsiao et al.[22] identified patients with sleep disorders using ICD-9 codes 307.4 and 780.5x and explored the association between sleep disorders and autoimmune diseases. ICD codes have also been used to identify obstructive sleep apnea[23, 24]. However, it has been shown that sleep disorders are poorly coded in structured EHR data. ICD codes for identifying a sleep disorder from inpatient EHR data has only 79.2% sensitivity and 28.4% specificity[25]. A study also found that manual chart review of unstructured EHR data was able to identify 50% more individuals with insomnia and 68% more individuals with sleep problems, compared to using only ICD code[26]. Therefore, unstructured EHR data (e.g., clinical notes) is valuable for identifying sleep information.

Despite the rich sleep information embedded in unstructured EHRs, only a limited number of studies are found in the literature applying NLP and machine learning methods to automatically extract sleep information from unstructured EHRs. Divita et al.[27] applied a keyword matching approach to extract general symptoms that includes sleepiness from VA clinical notes. Similarly, a few studies used NLP to extract disturbed sleep and insomnia from clinical notes to study the association between sleep and mental disorders[28]. For example, Zhou et al.[29] used sleep-related symptoms to identify patients with depression, and Irving et al.[30] considered insomnia and distributed sleep in clinical notes to predict psychosis. Kartoun et al.[31] used text mention of sleep disorders in clinical notes to predict insomnia and found superior performance in identifying insomnia patients compared to ICD codes. Other studies used sleep-related reactions, such as sleeplessness and sleepy, from clinical notes or adverse event reports of adverse drug events[32, 33]. However, none of the studies focused on automated methods of extracting a comprehensive list of sleep variables related to the AD.

Methods

Data Collection

We first defined a cohort of patients diagnosed with AD (ICD-10 codes: G30.0, G30.1, G30.8, and G30.9) between January 1st, 2020 and December 31st, 2020 at UPMC. We collected all their clinical notes that were created between January 1st, 2016 and December 31st, 2020, through the data service provided by the University of Pittsburgh Health Record Research Request (R3). The University of Pittsburgh's Institutional Review Board (IRB) reviewed and approved this study's protocol

Data Preprocessing

A total of 7,266 patients were identified, and approximately 1.1 million de-identified clinical documents were retrieved. Since these clinical documents were auto-populated in the EHR system, we developed a data preprocessing algorithm to clean the clinical text. Specifically, we integrated clinical documents with the same document ID regardless of the line number ID and removed duplicated clinical documents. Although some documents are not duplicated, most contents in those documents are overlapped due to the nature of data population. For example, a physician digitally signing a clinical document will generate a duplicated document with just the addition of "Digitally signed by…" one or two days later. Therefore, we applied a surface lexical similarity approach to identify the highly similar documents. Suppose that V is a set of unique words that occurred in documents D_1 and D_2. D_1 and D_2 can be represented in the same vector space as d_1 and d_2 respectively where each component corresponds to the word in V and the value is the word frequency. Then we calculated the cosine similarity between two document vectors. If the

similarity score is greater than 0.9, we will randomly remove one of the duplicated documents. After the data preprocessing, the total number of clinical documents was 379k.

Another challenge in extracting sleep information from clinical notes is that not every clinical note records relevant information. To identify relevant documents, we used information retrieval (IR) to select the documents. We applied fuzzy search that returns the documents containing keywords related to sleep regardless of the morphological format. The keywords used in our search are shown in Table 1. As a result, 192k (51%) out of 379k documents were returned for further investigation in this project. This set of 192k documents is called adSLEEP corpus hereafter.

Table 1. Sleep-related keywords used to retrieve relevant clinical note documents.

'snore', 'snoring', 'wheeze', 'wheezing', 'sleep', 'sleepiness', 'sleeping', 'sleepless', 'sleeplessness', 'apnea', 'hypopnea', 'osa', 'insomnia', 'nap', 'napping', 'narcolepsy', 'nocturnal', 'somnolence', 'somnolent', 'dizziness', 'hypersomnia', 'rem', 'nrem', 'wake', 'wakefulness', 'waking', 'polysomnography'

Table 2. Definitions and examples of sleep-related concepts.

Concept Category	Definition	Example
Snoring (Yes or No)	Snoring or snoring synonyms	"snoring" "sleep apnea"
Napping (Yes or No)	Napping during daytime	"napping" "doze"
Sleep problem (Yes or No)	Sleep problem Specific sleep disorder/ condition/disease mentioned in the note	"sleep disorder", "insomnia", "hypersomnia"
Bad sleep quality (Yes or No)	Any mention related to bad sleep quality in the note	"sleeplessness" "couldn't sleep during night" "staying up all night"
Daytime sleepiness (Yes or No)	Sleepiness during daytime	"sleep a lot through out the day" "excessive daytime sleepiness"
Night wakings (Yes or No)	Night time wakings	"frequent night waking" "waking up in the middle" "waking up 3-5 times"
Sleep duration (Short <=6h, Medium 6-8h, or Long >=8h)	Duration of night time sleep	"sleeps 4-5 hours" --> Short "sleep more than 12 hours" --> Long

Gold Standard Data Annotation

We randomly sampled 320 clinical note documents from the adSLEEP corpus for manual annotation to create the gold standard dataset. Two health informatics students annotated this sampled dataset. The annotator was directed to annotate mentions of seven sleep-related categories in each clinical note, including snoring, napping, sleep problem, sleep quality, daytime sleepiness, night wakings, sleep duration. These sleep-related concept categories are defined based on previous sleep studies for patients with AD[6, 12, 34]. Table 2 shows the detailed definitions for each of the classes. Then a judge aggregated the concept annotations in a clinical note to a document-level label. Initially, a batch

of 20 documents was given to the annotators to refine the annotation guidelines and to discuss the discrepancies to reach consensus on the concept definition. We repeated this process on another batch of 20 documents until an inter-annotator agreement (IAA) above 0.60 in terms of Cohen's Kappa was achieved. These 40 documents were used to measure a final IAA. Then we annotated the remaining 200 clinical notes using the updated annotation guidelines.

Rule-based NLP Algorithm

We developed a rule-based NLP algorithm named *nlp4sleep* for sleep information extraction using MedTagger[35], a clinical NLP tool based on the Unstructured Information Management Architecture (UIMA) framework[36]. The MedTagger software is publicly available at GitHub[3].

We first used top-down and bottom-up approaches to identify the keywords in the rules for each sleep concept extraction. For the top-down approach, we searched the synonyms for each concept in the medical terminologies and ontologies, including Unified Medical Language System (UMLS) Metathesaurus. For the bottom-up approach, we used word embeddings trained from the clinical corpus to find the top 3 most similar terms. Then we used 70% (200 documents) of the gold standard dataset as training data to develop regular expression rules for the NLP algorithm. Since MedTagger includes the negation detection and hypothetical mention detection, we didn't specify negation rules unless we saw undetected negations in the training data. Table 3 lists the regular expression rules used in the *nlp4sleep* algorithm to extract sleep concepts. The NLP system extracted sleep concepts from each clinical document and assigned a document-level classification for each concept. If there were multiple mentions one of a concept in a document, we applied majority voting strategy to obtain the final document label. The NLP algorithm is publicly available through the Open Health Natural Language Processing (OHNLP) consortium at GitHub (https://github.com/OHNLP/nlp4sleep).

Table 3. Regular expression rules used in the NLP algorithm for the extraction of sleep concepts from clinical notes.

Concept Category	Keywords and Regular Expressions
Snoring	snor(es|ing|e)?; snorings; sleep apnea; osa; obstructive sleep apnea
Napping	nap(s|ping)?
Sleep problem	insomnia; sleeplessness; sleep (disorders?|problems?); hypersomnia; parasomnia; osa; obstructive sleep apnea; sleep apnea; hypersomnolence
Bad sleep quality	staying up; (trouble|irritable|tense) (\S+\s+){0,5}(sleep(ing)?|asleep); sleep(s|ing)? poorly; sleep is poor; restless sleep; ca(n't|nnot) sleep; sleep issues?; sleep(ing)? (\S+\s+){0,5}(problems?|problematic); sleeps? a lot; difficulty (\S+\s+){0,5}(asleep|sleep(ing)?); sleep disturbance; disturbance in sleep; sleep quality: (fair|bad); not sleeping; no sleep; sleep difficulty; nocturnal agitation; up (during|at) night; nocturnal; often awake
Daytime sleepiness	(excessive)?daytime sleep(iness|inesses)?; (excessive)?daytime somnolence; sleep(s|ing|iness)? at times; sleep(s|iness)? in (\S+\s+){0,2}day(time)?; sleep(s|iness)? during (\S+\s+){0,2}day(time)?; sleep all day
Night wakings	night (wakings|awakenings), wak(e|es|ing up) (\S+\s+){0,5}night; awake(ning|n)? (from|during|at) night(mares)?
Sleep duration	Short: sleep(s|ing)? (less than|up to) (1|2|3|4|5|6) hours Medium: sleep(s|ing)? (\S+\s+){0,5}(6|7|8)-(6|7|8) hours Long: sleep(s|ing)? (\S+\s+){0,5}more than (8|9|10|11|12|13|14|15|16|17|18|19|20) hours; sleep(s|ing)? (\S+\s+){0,5}(8|9|10|11|12)-(8|10|11|12|13|14|15|16|17|18|19|20) hours

[3] https://github.com/OHNLP/MedTagger

Machine Learning Models

We compared the rule-based NLP algorithm with machine learning models. Five major machine learning-based clinical text classification models, namely Gradient Boosting Machine (GBM), Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest (RF), were trained and tested for each sleep concept. Since real-world clinical text data may be inadequate and inconsistent for machine learning-based NLP models to comprehend, we adopted multiple preprocessing steps before feeding the text data to the machine learning models. After converting into lower case, the text was tokenized to break the sentences into tokens, including words, phrases, symbols, or other meaningful elements. Stop words and non-numeric were removed from the token lists. The tokens were then lemmatized to the base forms to reduce the complexity of the text. Then we converted the entire document to a numeric vector using Term Frequency-Inverse Document Frequency (TF-IDF) vectorization. The experiments were conducted as a binary text classification task, with positive and negative predictions for each concept category. The same training and test datasets were used as in the rule-based NLP algorithm, and the performance on the testdataset was reported.

Evaluation

We used 30% (120 documents) of the gold standard dataset as the testing data to validate the rule-based NLP algorithm and machine learning models. All models were evaluated using sensitivity, specificity, positive predictive value (PPV), and F1 score. Since our dataset is imbalanced, we report the weighted-averaged F1 score. The definitions of the evaluation metrics are shown below:

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Nagative}$$

$$Specificity = \frac{True\ Nagative}{True\ Nagative + False\ Positive}$$

$$PPV = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$F1\ score = \frac{2 \cdot True\ Positive}{2 \cdot True\ Positive + False\ Positive + False\ Nagative}$$

Results

Table 4 shows the demographics of the AD cohort of 7,266 patients. Patients were primarily white (91%), female (64%), and not Hispanic or Latino (95%), with a mean age of 85 years. The demographics of this cohort are similar to the demographics of the population in western Pennsylvania.

Table 5 lists the number of documents for each sleep concept in the annotated training and test datasets. As shown in the table, the frequency of these sleep concepts is low in the gold standard dataset. Though the clinical documents were identified by using a list of relevant keywords, most documents do not contain any sleep-related concepts. The reason might be that some keywords may not be only related to sleep; for example, wheezing might be related to respiratory diseases.

The performance of the rule-based NLP algorithm and machine learning models is listed in Table 6. The result for the sleep duration concept is ignored since there are no annotations in the test dataset. As shown in Table XX, we can clearly see that the rule-based NLP algorithm outperform the machine learning models for all six sleep-related concepts. For identifying snoring, the rule-based NLP algorithm achieved the best performance of 0.97 and 0.89 in terms of F1 score and PPV, respectively, and a specificity of 0.97 that is close to the best specificity of 0.98 achieved by GBM. RF had the best sensitivity of 1.00 but a low PPV and F1 score. LR, KNN, and SVM had zero F1 scores since they failed to identify the positive cases. This is also the reason why we observed many zero F1 scores as the machine learning models could not learn patterns effectively from the training dataset with only a few positive cases. For the napping concept, the rule-based NLP algorithm achieved the best

Table 4. Demographics of the AD cohort.

Demographics	Total (n=7,266)
Age, years	85 (79,91)
Sex	
Female	4,649 (64%)
Male	2,617 (36%)
Race	
White	6,628 (91%)
Black	482 (6.6%)
Asian	67 (1.1%)
Others	67 (1.1%)
Not Specified	22 (0.3%)
Ethnicity	
Hispanic or Latino	33 (0.5%)
Not Hispanic or Latino	6,900 (95%)
Not Specified	333 (4.5%)

performance in terms of four metrics while all the machine learning models failed to identify the positive cases. We can observe the same results for the night wakings concept. The rule-based NLP algorithm achieved the best performance consistently for sleep problem, bad sleep quality, and daytime sleepiness. SVM had a comparable performance for sleep problem and daytime sleepiness in terms of sensitivity and specificity and couldn't accurately identify positive cases, which resulted in low PPVs and F1 scores. These results are consistent with previous studies[37, 38] that machine learning models might not be effective in clinical text classification when the size of the annotated training dataset is small, and the concepts of interest are sparse and infrequent in teh dcouments.

Table 5. Number of clinical documents for each sleep concept in the annotated training and test datasets.

Concept Category	# Documents in Training (Yes/No)	# Documents in Test (Yes/No)
Snoring	30 / 170	14 / 106
Napping	4 / 196	1 / 119
Sleep problem	35 / 165	20 / 100
Bad sleep quality	20 / 180	8 / 112
Daytime sleepiness	9 / 191	1 / 119
Night wakings	2 / 198	2 / 118
Sleep duration	0 (Short) / 1 (Medium) / 2 (Long)	0 (Short) / 0 (Medium) / 0 (Long)

Table 6. Performance of the rule-based NLP algorithm and machine learning models.

Sensitivity **Specificity** **PPV** **F1**	Snoring	Napping	Sleep Problem	Bad Sleep Quality	Daytime Sleepiness	Night Wakings
Rule-based NLP	0.94 0.97 **0.89** **0.97**	**0.5** **0.99** **0.5** **0.98**	0.85 **0.93** **0.8** **0.91**	**0.62** 0.51 **0.6** **0.91**	**1.00** **1.00** **1.00** **1.00**	**1.00** **0.99** **0.75** **0.99**
GBM	0.88 **0.98** 0.88 0.88	- 0.99 0 0	0.79 0.95 0.75 0.77	0 **0.93** 0 0	0 0.97 0 0	0 0.99 0 0
LR	0 0.86 0 0	- 0.99 0 0	- 0.83 0 0	- **0.93** 0 0	- 0.97 0 0	- 0.99 0 0
KNN	0 0.86 0 0	- 0.99 0 0	- 0.83 0 0	- **0.93** 0 0	- 0.97 0 0	- 0.99 0 0
SVM	0 0.86 0 0	0 0.99 0 0	**1.00** 0.85 0.1 0.18	- **0.93** 0 0	1.00 0.97 0.25 0.4	- 0.99 0 0
RF	**1.00** 0.87 0.06 0.11	- 0.99 0 0	- 0.83 0 0	- **0.93** 0 0	- 0.97 0 0	- 0.99 0 0

Error Analysis of the rule-based NLP algorithm

We conducted an error analysis of the documents misclassified by the rule-based NLP algorithm and analyzed the causes of the false positives and false negatives for each sleep concept. Some false positives were due to our annotators failing to annotate the information. For example, in the text "Histories Past Medical History Combined Chronic Systolic/Diastolic CHF COPD (industrial exposure) CAD s/p stents PMH Right Ocular Stroke - chronic visual defect BPH Type 2 DM OSA on BiPAP", the rule-based NLP algorithm identified OSA as snoring and sleep problem. However, the concept had not been annotated. Many semi-structured clinical text auto-populated in the EHR system is difficult for the annotator to read and annotate. In another false positive, the sentence "He had episodes yesterday in which he became confused after waking up from a nap" indicates that the patient had a nap that may not be related to sleep pattern. In another false positive case for sleep problem "Depression screen done 7/2017, PHQ9 score 16 points for sleep problem which seems better now", the NLP algorithm couldn't identify that this sentence was not about a positive sleep problem. In a false positive case for bad sleep quality, the sentence "Take Melatonin 5 mg at bedtime every night for 3- 4 weeks for difficulty falling asleep" was a suggestion for the patient.

Some false negatives are due to errors in negation detection. For example, in the sentence "The patient's daughter states that she has not been complaining of her back pain or of her leg cramps we discussed the fact that she is doing less and does nap during the day", the algorithm incorrectly identified this mention as negated since it failed to identify two sentences. In another example for sleep problem "Change in social contacts/activities? No Patient Active Problem List Diagnosis Primary open angle glaucoma Urge incontinence Backache, unspecified Pneumonia, organism unspecified Insomnia", the NLP algorithm incorrectly split the sentence into "Change in social contacts/activities?" and "No Patient Active Problem List Diagnosis Primary open angle glaucoma Urge incontinence Backache, unspecified Pneumonia, organism unspecified Insomnia" and wrongly identified the negation. The semi-structured clinical text also confused the NLP algorithm in detecting sentences and negations.

Discussion

Detailed descriptions of SDOH are usually captured in unstructured clinical text; however, the SDOH information may be sparsely documented due to the lack of clinical practice guidelines for documenting such information. Our study shows that sleep information is infrequently recorded in clinical notes for patients with AD. For example, in the gold standard dataset, only 14% of clinical documents recorded snoring concepts, 1.6% napping, 17.2% sleep problem, 8.8% bad sleep quality, 3.1% daytime sleepiness, 1.3% night wakings, and 1% sleep duration. This observation raises the question of whether such under-documented SDOH information in clinical notes will be useful for downstream statistical analysis in a cohort study or an epidemiology study.

Another challenge we encountered during the project was the definition of sleep-related concepts. We initially considered eight concepts, including sleep disorder, sleep problem symptoms, snoring, napping, sleep quality, daytime sleepiness, night wakings, and sleep duration, with detailed granularity according to the relevant sleep research in the literature. For example, there were four categories associated with snoring or daytime sleepiness: negated, positive, sometimes, and all the time. There were three categories for night wakings: 0, 1-2, and >2. However, during the annotation process, we found that the granular categories for each concept were rarely used and there were significant overlaps between sleep problem symptoms and other concepts. For example, phrases like "staying up all night" meet the description of insomnia, but the patient was never diagnosed with insomnia. Likewise, snoring and sleep apnea shared concept-likeness but are not always annotated similarly. For example, the concept of snoring is annotated as snoring but not sleep apnea. However, the concept of sleep apnea is annotated as a sleep problem and snoring because the concept meets both definitions. Thus, we simplified the concept definition and the categories for each concept.

Since SDOH comprises more conditions related to socioeconomic status, living environment, housing, education, food, community, it might be more challenging to define these SDOH concepts. It is also questionable whether such information is adequately documented in EHRs and whether such information from EHRs would be useful for research. Thus, a feasibility study of assessing the availability of SDOH in EHRs for a certain cohort of patients might be necessary before algorithm development. In addition, it is also a tedious and time-consuming process to manually annotate a gold standard dataset. A potential beginning point of building automated systems to extract SDOH from EHRs might be a community effort to build an SDOH ontology and terminology.

Additionally, sleep information is infrequently documented in the clinical notes and keywords are shared with concepts. Although we used IR to select the documents with keywords related to sleep, keywords including wheeze, wheezing, and apnea appear often but are unrelated to the patient's sleep. For example, physicians commonly check a patient's respiratory health and record the presence of wheezing in the clinical notes. Wheezing, a shared concept

between respiratory health and sleep, was found problematic when retrieving sleep-specific documents. Although sleep information is infrequently documented in the clinical notes, the proposed rule-based NLP algorithm still achieved promising results. In comparison, the machine learning and deep learning based approaches didn't achieve good results, which is due to the small size of sleep information in the training data. This study focused on the clinical notes of patients with AD, but could be extended to general sleep information extraction for other diseases.

Limitations

There are several limitations in this study. First, the ICD codes used to define AD may not be optimal. However, a more comprehensive way to define AD is out of scope of this study. Second, the initial search keywords used to retrieve sleep-related clinical notes may not be complete and could miss some documents. However, this could be a common problem for SDOH information extraction due to the sparse and infrequent documentation in clinical notes. Third, the gold standard dataset is relatively small due to the time-consuming and expensive manual annotation. Last, we didn't consider sleep information in other EHR data types, such as diagnosis codes, survey data, questionnaire data, sleep studies such as polysomnography, and sleep tests such as multiple sleep latency test (MSLT), which should be considered for further cohort studies or epidemiology studies.

Future Work

In future work, we plan to explore more sophisticated methods in retrieving relevant documents to the considered medical concepts with high precision. This might be a key challenge in collecting a corpus for studying an SDOH concept. In addition, we will also investigate novel machine learning methods that require less or no data for training.

Acknowledgements

This project was partially supported by the University of Pittsburgh Momentum Seed Funds and the National Institutes of Health through Grant Number UL1TR001857 funds. The funders had no role in the design of the study, and collection, analysis, and interpretation of data and in preparation of the manuscript. The views presented in this report are not necessarily representative of the funder's views and belong solely to the authors.

References

1. Association As. 2018 Alzheimer's disease facts and figures. Alzheimer's & Dementia. 2018;14(3):367-429.
2. Association As. 2019 Alzheimer's disease facts and figures. Alzheimer's & dementia. 2019;15(3):321-87.
3. Jia J, Wei C, Chen S, Li F, Tang Y, Qin W, Zhao L, Jin H, Xu H, Wang F. The cost of Alzheimer's disease in China and re‐estimation of costs worldwide. Alzheimer's & Dementia. 2018;14(4):483-91.
4. Feigin VL, Abajobir AA, Abate KH, Abd-Allah F, Abdulle AM, Abera SF, Abyu GY, Ahmed MB, Aichour AN, Aichour I. Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet Neurology. 2017;16(11):877-97.
5. Rosenberg A, Ngandu T, Rusanen M, Antikainen R, Bäckman L, Havulinna S, Hänninen T, Laatikainen T, Lehtisalo J, Levälahti E. Multidomain lifestyle intervention benefits a large elderly population at risk for cognitive decline and dementia regardless of baseline characteristics: The FINGER trial. Alzheimer's & Dementia. 2018;14(3):263-70.
6. Keage HAD, Banks S, Yang KL, Morgan K, Brayne C, Matthews FE. What sleep characteristics predict cognitive decline in the elderly? Sleep medicine. 2012;13(7):886-92.
7. Cricco M, Simonsick EM, Foley DJ. The impact of insomnia on cognitive functioning in older adults. Journal of the American Geriatrics Society. 2001;49(9):1185-9.
8. Foley D, Monjan A, Masaki K, Ross W, Havlik R, White L, Launer L. Daytime sleepiness is associated with 3‐year incident dementia and cognitive decline in older Japanese‐American men. Journal of the American Geriatrics Society. 2001;49(12):1628-32.
9. Elwood PC, Bayer AJ, Fish M, Pickering J, Mitchell C, Gallacher JE. Sleep disturbance and daytime sleepiness predict vascular dementia. Journal of Epidemiology & Community Health. 2011;65(9):820-4.
10. Quesnot A, Alperovitch A. Snoring and risk of cognitive decline: A 4‐year follow‐up study in 1389 older individuals. Journal of the American Geriatrics Society. 1999;47(9):1159-60.
11. Tworoger SS, Lee S, Schernhammer ES, Grodstein F. The association of self-reported sleep duration, difficulty sleeping, and snoring with cognitive function in older women. Alz Dis Assoc Dis. 2006;20(1):41-8.
12. Potvin O, Lorrain D, Forget H, Dubé M, Grenier S, Préville M, Hudon C. Sleep quality and 1-year incident cognitive impairment in community-dwelling older adults. Sleep. 2012;35(4):491-9.

13. Burke SL, Cadet T, Alcide A, O'Driscoll J, Maramaldi P. Psychosocial risk factors and Alzheimer's disease: the associative effect of depression, sleep disturbance, and anxiety. Aging & mental health. 2018;22(12):1577-84.
14. Jelicic M, Bosma H, Ponds RW, Van Boxtel MP, Houx PJ, Jolles J. Subjective sleep problems in later life as predictors of cognitive decline. Report from the Maastricht Ageing Study (MAAS). Int J Geriatr Psych. 2002;17(1):73-7.
15. Blumenthal D. Launching hitech. New England Journal of Medicine. 2010;362(5):382-5.
16. Perera G, Pedersen L, Ansel D, Alexander M, Arrighi HM, Avillach P, Foskett N, Gini R, Gordon MF, Gungabissoon U. Dementia prevalence and incidence in a federation of European Electronic Health Record databases: the European Medical Informatics Framework resource. Alzheimer's & Dementia. 2018;14(2):130-9.
17. Chen L, Reed C, Happich M, Nyhuis A, Lenox-Smith A. Health care resource utilisation in primary care prior to and after a diagnosis of Alzheimer's disease: a retrospective, matched case–control study in the United Kingdom. BMC geriatrics. 2014;14(1):1-9.
18. Poblador-Plou B, Calderón-Larrañaga A, Marta-Moreno J, Hancco-Saavedra J, Sicras-Mainar A, Soljak M, Prados-Torres A. Comorbidity of dementia: a cross-sectional study of primary care older patients. BMC psychiatry. 2014;14(1):1-8.
19. Mayeda ER, Glymour MM, Quesenberry CP, Whitmer RA. Inequalities in dementia incidence between six racial and ethnic groups over 14 years. Alzheimer's & Dementia. 2016;12(3):216-24.
20. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, Liu S, Zeng Y, Mehrabi S, Sohn S. Clinical information extraction applications: a literature review. Journal of biomedical informatics. 2018;77:34-49.
21. Felder JN, Baer RJ, Rand L, Jelliffe-Pawlowski LL, Prather AA. Sleep disorder diagnosis during pregnancy and risk of preterm birth. Obstetrics & Gynecology. 2017;130(3):573-81.
22. Hsiao Y-H, Chen Y-T, Tseng C-M, Wu L-A, Lin W-C, Su VY-F, Perng D-W, Chang S-C, Chen Y-M, Chen T-J. Sleep disorders and increased risk of autoimmune diseases in individuals without sleep apnea. Sleep. 2015;38(4):581-6.
23. Ramesh J, Keeran N, Sagahyroon A, Aloul F, editors. Towards Validating the Effectiveness of Obstructive Sleep Apnea Classification from Electronic Health Records Using Machine Learning. Healthcare; 2021: Multidisciplinary Digital Publishing Institute.
24. Larsen AJ, Rindal DB, Hatch JP, Kane S, Asche SE, Carvalho C, Rugh J. Evidence supports no relationship between obstructive sleep apnea and premolar extraction: an electronic health records review. Journal of Clinical Sleep Medicine. 2015;11(12):1443-8.
25. Jolley RJ, Liang Z, Peng M, Pendharkar SR, Tsai W, Chen G, Eastwood CA, Quan H, Ronksley PE. Identifying cases of sleep disorders through international classification of diseases (ICD) codes in administrative data. International Journal of Population Data Science. 2018;3(1).
26. Singer E, Malow B, Niarchou M, Davis L, Johnston R, Sutcliffe J, Hucks D. Characterizing Sleep Disorders in an Autism-Specific Collection of Electronic Health Records. medRxiv. 2021.
27. Divita G, Luo G, Tran L-TT, Workman TE, Gundlapalli AV, Samore MH. General symptom extraction from VA electronic medical notes. MEDINFO 2017: Precision Healthcare through Informatics: IOS Press; 2017. p. 356-60.
28. Jackson RG, Patel R, Jayatilleke N, Kolliakou A, Ball M, Gorrell G, Roberts A, Dobson RJ, Stewart R. Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. BMJ open. 2017;7(1):e012012.
29. Zhou L, Baughman AW, Lei VJ, Lai KH, Navathe AS, Chang F, Sordo M, Topaz M, Zhong F, Murrali M. Identifying patients with depression using free-text clinical documents. MEDINFO 2015: eHealth-enabled Health: IOS Press; 2015. p. 629-33.
30. Irving J, Patel R, Oliver D, Colling C, Pritchard M, Broadbent M, Baldwin H, Stahl D, Stewart R, Fusar-Poli P. Using natural language processing on electronic health records to enhance detection and prediction of psychosis risk. Schizophrenia bulletin. 2021;47(2):405-14.
31. Kartoun U, Aggarwal R, Beam AL, Pai JK, Chatterjee AK, Fitzgerald TP, Kohane IS, Shaw SY. Development of an algorithm to identify patients with physician-documented insomnia. Sci Rep-Uk. 2018;8(1):1-9.
32. Tang H, Solti I, Kirkendall E, Zhai H, Lingren T, Meller J, Ni Y. Leveraging food and drug administration adverse event reports for the automated monitoring of electronic health records in a pediatric hospital. Biomedical informatics insights. 2017;9:1178222617713018.
33. Wang X, Hripcsak G, Markatou M, Friedman C. Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study. Journal of the American Medical Informatics Association. 2009;16(3):328-37.

34. Devore EE, Grodstein F, Schernhammer ES. Sleep duration in relation to cognitive function among older adults: a systematic review of observational studies. Neuroepidemiology. 2016;46(1):57-78.
35. Liu H, Bielinski SJ, Sohn S, Murphy S, Wagholikar KB, Jonnalagadda SR, Ravikumar K, Wu ST, Kullo IJ, Chute CG. An information extraction framework for cohort identification using electronic health records. AMIA Summits on Translational Science Proceedings. 2013;2013:149.
36. Ferrucci D, Lally A. UIMA: an architectural approach to unstructured information processing in the corporate research environment. Natural Language Engineering. 2004;10(3-4):327-48.
37. Wang Y, Sohn S, Liu S, Shen F, Wang L, Atkinson EJ, Amin S, Liu H. A clinical text classification paradigm using weak supervision and deep representation. Bmc Med Inform Decis. 2019;19(1):1-13.
38. Cusick M, Adekkanattu P, Campion Jr TR, Sholle ET, Myers A, Banerjee S, Alexopoulos G, Wang Y, Pathak J. Using weak supervision and deep learning to classify clinical notes for identification of current suicidal ideation. Journal of psychiatric research. 2021;136:95-102.