




Article

Validation of a Natural Language Processing Algorithm for the Extraction of the Sleep Parameters from the Polysomnography Reports

Mahbubur Rahman^{1,2,3}, Sara Nowakowski^{1,2,4} , Ritwick Agrawal^{2,3} , Aanand Naik^{1,5}, Amir Sharafkhaneh^{2,4} and Javad Razjouyan^{1,2,*} 

- ¹ Houston Veterans Affairs Health Services Research and Development Service, Center for Innovations in Quality, Effectiveness and Safety, Michael E. DeBakey Veteran Affairs Medical Center, Houston, TX 77030, USA
- ² Department of Medicine, Baylor College of Medicine, Houston, TX 77030, USA
- ³ Medical Care Line, Michael E. DeBakey Veteran Affairs Medical Center, Houston, TX 77030, USA
- ⁴ Veterans Affairs South Central Mental Illness Research, Education and Clinical Center, Houston, TX 77030, USA
- ⁵ University of Texas School of Public Health, 1200 Pressler Str., Houston, TX 77030, USA
- * Correspondence: javad.razjouyan@va.gov

Abstract: Background: There is a need to better understand the association between sleep and chronic diseases. In this study we developed a natural language processing (NLP) algorithm to mine polysomnography (PSG) free-text notes from electronic medical records (EMR) and evaluated the performance. Methods: Using the Veterans Health Administration EMR, we identified 46,093 PSG studies using CPT code 95,810 from 1 October 2000–30 September 2019. We randomly selected 200 notes to compare the accuracy of the NLP algorithm in mining sleep parameters including total sleep time (TST), sleep efficiency (SE) and sleep onset latency (SOL), wake after sleep onset (WASO), and apnea-hypopnea index (AHI) compared to visual inspection by raters masked to the NLP output. Results: The NLP performance on the training phase was >0.90 for precision, recall, and F-1 score for TST, SOL, SE, WASO, and AHI. The NLP performance on the test phase was >0.90 for precision, recall, and F-1 score for TST, SOL, SE, WASO, and AHI. Conclusions: This study showed that NLP is an accurate technique to extract sleep parameters from PSG reports in the EMR. Thus, NLP can serve as an effective tool in large health care systems to evaluate and improve patient care.

Keywords: polysomnography; natural language processing; sleep parameters



Citation: Rahman, M.; Nowakowski, S.; Agrawal, R.; Naik, A.; Sharafkhaneh, A.; Razjouyan, J. Validation of a Natural Language Processing Algorithm for the Extraction of the Sleep Parameters from the Polysomnography Reports. *Healthcare* **2022**, *10*, 1837. <https://doi.org/10.3390/healthcare10101837>

Academic Editors: Keun Ho Ryu and Nipon Theera-Umpon

Received: 25 August 2022

Accepted: 16 September 2022

Published: 22 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Polysomnography (PSG) uses multiple parameters to evaluate and diagnose sleep disorders [1]. PSG is considered the gold standard for diagnosing sleep disorders and includes the measurement of several physiological activities, such as brain waves via electroencephalography (EEG), heart rate via electrocardiogram (EKG), eye movements via electrooculogram (EOG), muscle movements via electromyography (EMG), nasal and oral airflow via thermistor/thermocouple and through nasal pressure changes, respiratory effort via abdominal and thoracic bands, and blood oxygen saturation via pulse oximetry [2]. The PSG reports document several components such as patient demographic information, technical detail regarding the physiological records (e.g., number of electrodes), sleep continuity (e.g., total sleep time), sleep stages architecture (e.g., NREM Stage 1), respiratory indices (e.g., apnea hypopnea index [AHI]), and periodic limb movements (e.g., PLMI) [3]. The PSG report's components are typically stored as a free-text note in the electronic medical record (EMR), which is not readable, processable, or computable and is considered unstructured data [4].

To convert unstructured data into structured data that is readable, processable, and computable data, researchers use various methods [5–7]. These methods include a traditional approach via manual reviewing of notes and converting them into the structured data. The manual method is time-consuming, not feasible for a large quantity of notes, and prone to human error [8]. More recently, the use of novel machine learning techniques, such as natural language processing (e.g., NLP) algorithms, has been proposed to solve the shortcomings of manual data extractions. NLP has been successfully implemented within the Veterans Health Administration (VHA) EMR to extract data from clinical notes for ejection fraction and heart failure [9–11]. There are several other clinical studies using NLP algorithm on the VHA EMR database [12–15]. Few studies have reported the use of NLP to convert unstructured data from PSG reports into structured data [16,17]. Investigators used regular expression matching techniques to extract total sleep time (TST) with an accuracy of 80%; however, they did not provide detail on validation information in the published study. In addition, they did not extract the part of text to analysis the data and it limited the performance of their algorithm.

The aim of the present study was to develop an NLP algorithm based on prior work [18] to extract sleep continuity parameters including total sleep time (TST), sleep efficiency (SE), sleep onset latency (SOL), wake after sleep onset (WASO), and respiratory index (AHI) from PSG reports in the VHA EMR and to test the accuracy of the NLP algorithm compared to annotators/raters masked to NLP output.

2. Methods

In this study, we used the Corporate Data Warehouse [19]. The study protocol was approved by the Research & Development Committee of the Michael E. DeBakey VA Medical Center and Baylor College of Medicine Institutional Review Board (IRB# H-35366).

2.1. Cohort

This is a retrospective study utilizing the VHA EMR from 1 October 1999, through 30 September 2020. We included patients who had any International Classification of Disease, 9th edition (ICD-9) or 10th edition (ICD-10) sleep disease, Supplementary Table S1. The VA EMR is also known as the VHA's Corporate Data Warehouse (CDW). It is a relational database that collects veterans data from all VHA facilities from October 1999 to the present [20].

2.2. Database

The cohort was selected from 4,237,444 patients. We limited the results to the patients who had the following Current Procedural Terminology (CPT) procedure codes: 95810 and 95811. We included patients' notes that were associated with visits with the CPT codes had "%poly%" OR "%PSG%" in the title of notes; had "%PSG%" OR "%polysomno%" in the body of notes; and did not have "%titrat%" OR "%split%" in the body of text, which represented split-night studies.

For the remaining notes, we developed a document quality score, which we refer to as a document quality index (DQI). The DQI is a score ranging from zero to seven. Zero referred to no sleep parameter phrases used in the note, and seven referred to full documentation of the sleep parameters. The DQI was calculated by summing seven components, and each component related to a sleep parameter. We assigned zero to each component if the sleep parameter of interest was not documented. The components were as follows: (1) TST, "%TST%" or "%total sleep time%"; (2) SoL, "%SoL%" or "%onset latency%"; (3) SE, "%SE%" or "%sleep efficiency%"; (4) WASO, "%WASO%" or "%wake after sleep onset%"; (5) rapid eye movement (REM), "%REM%"; (6) sleep stage 1 (N1), "%N1%"; (7) AHI, "%AHI%" or "%apnea%index%". After consulting with the team of sleep board-certified physicians (AS and RA) providing them with samples of notes with the scores greater than 3, the medical team recommended using a cut of 4 or greater. We had 46,093 notes from 42,991 patients.

2.3. Sampling Strategy for the NLP Development and Reference Standard

We applied the exact power calculation method to determine the minimum required sample size for the validation of the NLP algorithm using the recommend sampling strategy [21]. With assumption of effect size of 0.3, alpha of 0.05, power of 80, and degree of freedom of 5, the minimum number of notes for each variable was 143. However, the annotators reviewed 200 randomly selected notes, with 160 notes for the training performance and 40 notes for the validation performance of the NLP algorithm. Two board-certified sleep medicine specialists reviewed and labeled the sampled generated data for validation of the NLP algorithm (AS and RA). In case of discrepancy, the final decision was judicated liberally by adding another SN as final voter. The two raters were blinded from each other. During the judication process of discrepancy, the raters informed about their labels. Additionally, we used the Cohen's κ , Intraclass correlation coefficient (ICC) [22,23] to understand the level of reviewers' agreement. The ICC ranges from zero (no agreement) to one (perfect agreement). The ICC was interpreted as: poor (<0.5), moderate (0.5–0.7), good (0.75–0.9), and excellent (≥ 0.9).

2.4. NLP Algorithm Development for the Extraction of the Sleep Parameters

We developed an NLP algorithm for each sleep parameter of interest. Therefore, the following NLP algorithm for extracting the sleep continuity parameters was repeated for each sleep parameter, Figure 1. The NLP algorithm receives the PSG reports as input texts and generates the sleep parameter associated with the respected quantity along with the patient's identification and visit date to a structured output format. We divided the NLP algorithm into five steps (Figure 1). The algorithm was developed with the Python programming language [24]. We used the Natural Language Toolkit (NLTK), which is a suite of libraries and program to process human language data [25].

1. Reading: Read the PSG reports and store in the memory.
2. Tokenizing: Tokenize the PSG reports.
3. Locating: Identify the location of the sleep parameter of interest.
4. Mining: Identify the related quantity associated with the sleep parameter of interest.
5. Storing: Store the sleep parameter and associated quantity to a csv file.

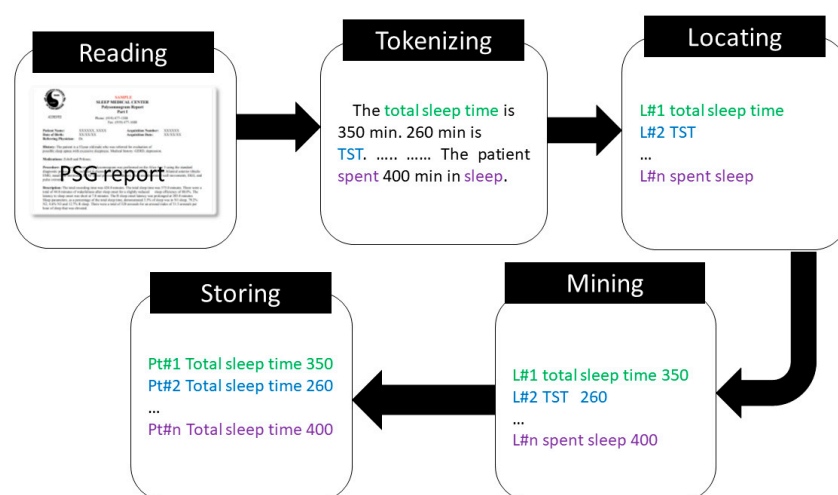


Figure 1. An NLP flow diagram to extract sleep parameters from the PSG reports.

Reading: We imported the PSG reports using the NLTK text reading function. The imported text data were stored for the next step.

Tokenizing: We used the NLTK tool to parse the PSG note into sentences and lines, respectively. Such parsing helped to locate the sleep parameters for the extraction and association with the quantity respectively (Figure 1).

Locating: We developed a set of regular expressions to extract the sleep parameters from the corresponding lines received from the NLTK toolkit. These regular expression sets were developed based on the annotated notes by our sleep specialist team. For instance, the TST has been expressed entirely or its abbreviated form, i.e., TST, or a different format completely in some PSG reports (Supplementary Table S2). However, we developed regular expressions to capture common possible formats of the sleep parameters documented in the PSG reports (Supplementary Table S2). The following string patterns were used for each concept: TST: 'total sleep time', 'slept for', 'monitored for', 'spent for', 'Total Sleep Time (TST)'; SOL: 'sleep onset latency', 'sleep onset', 'latency for', 'Sleep Onset Latency (SOL)'; SE: 'sleep efficiency', 'sleep efficiency for', 'Sleep Efficiency (SE)'; WASO: 'wake after', 'wake time after', 'total wake', 'WASO', 'Wake After Sleep Onset (WASO)'; AHI: 'apnea hypopnea index', 'apnea/hypopnea index', 'AHI for', 'Apnea Hypopnea Index (AHI)'.

Mining: The nearest neighbor based quantification step extracts the associated quantity for each of the extracted sleep parameters [26]. In some PSG reports, the quantities precede the sleep parameter and vice versa (Supplementary Table S2). The nearest neighbor algorithm selects the nearest quantity as the associated quantity for the sleep parameter of interest, while selecting the second nearest for some cases (Supplementary Table S2). The successive operations of both locating and mining steps associate the sleep parameter of interest with its corresponding quantity [27]. All the information retrieved in this step is forwarded to the next step.

Storing: In this step, the sleep parameter, the associated quantity, patient identifiers, and visit date were stored to a csv file, aka structured data.

2.5. Performance of the NLP Algorithm

The performance of the NLP algorithm has been measured by the accuracy, precision, recall, and F-1 score [28], as this is a standard for performance analysis of the information retrieval algorithm [29]. The accuracy defines the fraction of the relevant information retrieved from all the documents (accuracy = true positive + true negative/total number of notes). The precision defines the fraction of retrieved documents that are in fact relevant (precision = true positive/(true positive + false positive)). The recall defines the fraction of relevant documents that are retrieved by the algorithm (recall = true positive/(true positive + false negative)). The F-1 score measures the combined performance of the recall and precision (F-1 score = $2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$) [30].

3. Results

In this study, we collected 407,730 PSG notes from 4,237,444 patients who used the VHA for medical sleep advises (88.3% male, 26.3% age ≥ 65 years, 57.7% obese, BMI ≥ 30 Kg/m²), Supplementary Table S3 and Supplementary Figure S1. Of 407,730 notes, 46,093 (11.3%) met the DQI criteria ≥ 4 , Figure 2.

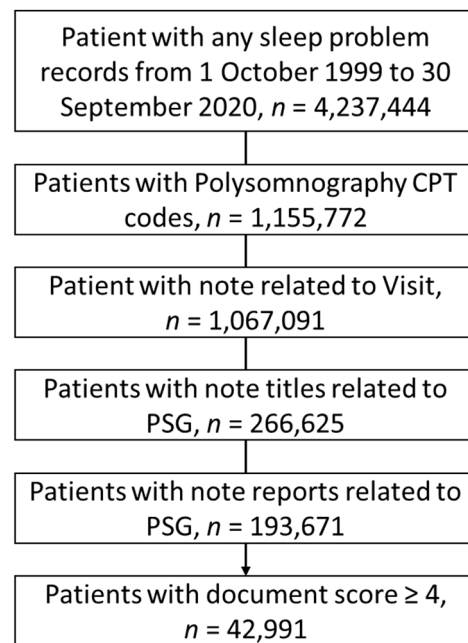


Figure 2. The strobe diagram for collecting patients' polysomnography reports with CPT codes of '95810', and '95811' who visited veteran health administration (VHA) from 1 October 1999 to 30 September 2020.

3.1. Reliability Analysis

The reliability between annotators ranged from 0.51 to 0.89, Table 1. We observed the highest ICC value in SOL (0.89) and the lowest value in WASO (0.51). The ICC values for SE, WASO, and AHI were considered moderate, while the ICC values for TST and SOL were good.

Table 1. The agreement between two raters measured by inter class correlation (95% confidence intervals) on sleep notes.

Sleep Parameter *	N (%)	ICC (95% CI)
TST	195 (97.5)	0.83 (0.78, 0.87)
SE	196 (98.0)	0.59 (0.49, 0.68)
SOL	194 (97.0)	0.89 (0.86, 0.92)
WASO	184 (92.0)	0.51 (0.40, 0.61)
AHI	192 (96.0)	0.62 (0.52, 0.70)

* TST = total sleep time, SE = sleep efficiency, SOL = sleep onset latency, WASO = wake after sleep onset. ICC = inter class correlation.

3.2. Performance Analysis

We provided the performance analysis for the training and validation datasets in Table 2. In the training phase, the NLP algorithm showed accuracy of 0.91 across all sleep parameters. The highest performance was observed in the SE and WASO, and the lowest performance was observed in the TST (accuracy, 0.91). In the validation phase, SE had the highest performance, i.e., 1.00. The accuracy and recall level of SOL was the lowest (0.90).

Table 2. Performance analysis of the NLP algorithm.

Sleep Parameter	Accuracy	Precision	Recall	F-1 Score
Training				
TST	0.91	0.98	0.93	0.95
SOL	0.91	1.0	0.91	0.96
SE	0.98	1.0	1.0	0.99
WASO	0.98	1.0	0.98	0.99
AHI	0.96	0.99	0.96	0.98
Validation				
TST	0.95	1.0	0.95	0.97
SOL	0.90	1.0	0.90	0.95
SE	1.0	1.0	1.0	1.0
WASO	0.95	1.0	0.95	0.97
AHI	0.95	1.0	0.95	0.97

4. Discussion

4.1. Methods and Results of the NLP Algorithm

In this study, we developed and validated an NLP algorithm to extract sleep parameters from PSG reports stored in the EMR as free-text notes. Additionally, we developed a quality metric to assess whether the PSG report was sufficient quality to be included in data extraction by consulting with a sleep medicine-certified group of physicians (A.S. and R.A.). This is one of the first published studies that used NLP to extract sleep continuity parameters and a respiratory index (AHI) and transform them from unstructured data to structured data.

In the process of preparing the notes for the NLP algorithm, we excluded notes that reported split nights and performed titration studies. In the split-night PSG, two nights' recordings are combined into a single overnight observation. The split-night PSG does not provide adequate sleep architecture data for a complete night [31]. We developed an inclusion metric, DQI, that ensured the quality of documentation of the PSG reports. The DQI was developed in-house based on the clinical experiences of our certified sleep medicine physicians (A.S. and R.A.). It provides an additional inclusion layer to ensure quality of the PSG report prior to processing and extracting PSG sleep parameter data. Next, we randomly selected 200 PSG reports to compare accuracy of the NLP algorithm to extract correct sleep parameter data compared to visual inspection by multiple masked raters (JR and MR). We reported moderate-to-good ICC among raters, which originated from variation of expressions in documenting sleep parameter (Supplementary Table S2). Because of variation in documentation, the performance of the NLP also varies. For example, the SE has a regular pattern of number and percentage association that helps to achieve the highest performance (SE sentences in Supplementary Table S2). However, the performance degrades for the anomalies, with either the sleep parameter statement expressed as a different expression or subsequent expression of another sleep parameter (TST, SOL in Table 2). Clinical interpretation and research of PSG reports could benefit by creating a standardized template for sleep reports. Compared to a previous study with average performance of 80% [17], the overall performance of the algorithm to extract sleep parameter is greater than 90%. The previous study developed based on a regular expression approach, while we used a more sophisticated algorithm to extract the sleep parameters [17].

4.2. Strength and Weakness

There are several limitations of this study. First, we limited the notes to those with proper documentation of PSG reports; and we excluded notes that related to split-night study or home sleep tests. We did this to reduce the variability that may introduce noise when developing and testing the NLP algorithm to extract sleep parameters. However,

by doing so we acknowledge we limited the number of notes reviewed. A separate NLP algorithm is warranted to mine the home sleep test notes and extract home sleep test sleep-related respiratory parameters such as AHI and blood oxygen saturation level (SaO₂) versus the sleep continuity variables typically found in PSG reports. PSG reports were selected from the VA EMR. Generalizability of the NLP algorithm to extract PSG sleep parameters from other facilities remains to be tested. Finally, future work should develop and test NLP algorithms to extract sleep architecture (Stage N1, REM) and other indices (RDI, SpO₂, PLMI).

5. Conclusions

We developed and validated an NLP algorithm for systematic extraction of the sleep continuity parameters and a respiratory index (AHI) from the PSG reports. The algorithm has >90% performance in all the performance metrics in both training and validation stages. In recent years there has been a rapid emergence of artificial intelligence (AI) in the field of sleep medicine. AI refers to the capability of computer systems to perform tasks conventionally thought to require human intelligence, such as speech recognition, decision making, and visual recognition of patterns and objects. Additionally, a new advanced machine learning based NLP algorithm such as Bidirectional Encoder Representations from Transformers (BERT) has been introduced developed by Google [32]. Another study is warranted to develop an NLP based on BERT to extract sleep parameters.

Sleep medicine is well positioned to benefit from advances that use big data to create artificially intelligent computer programs. Leveraging longitudinal data accumulated within EMR, sleep medicine is primed to benefit from AI. As a greater number of sleep studies occur each year within and outside the VHA, utilizing AI and advanced machine learning becomes more appealing to automate data extraction from the EMR [33]. AI techniques will allow clinicians and investigators to examine the wealth of rich, longitudinal real-world data that is housed within the EMR, currently in the form of free-text notes. By using AI techniques, investigators can leverage “big data” to offer new insights into sleep physiology, improve the accuracy of diagnosis of sleep disorders, predict response and adherence to treatment, and use sleep parameters as predictors of future physical and mental health, leading to treatment optimization and personalization.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/healthcare10101837/s1>, Table S1: list of International Classification of Diseases Clinical Modification (ICD) ninth and tenth edition and Current Procedural Terminology (CPT) code; Table S2: Examples of TST, SE, AHI, WASO, SOL sentences found in the PSG reports; Table S3: Patients' characteristics and demographics; Figure S1: The distribution of sleep report quality aka document quality index (DQI). The DQI is a score ranged from zero to seven. Zero referred to no sleep parameters phrases used in the note and seven referred to full documentation of the sleep parameters. The DQI is calculated by summing seven components and each component related to a sleep parameter as follows: 1) total sleep time, sleep onset latency, sleep efficiency (SE), wake after sleep onset (WASO), Rapid eye movement (REM), sleep stage 1 (N1), and Apnea Hypopnea Index (AHI).

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

Funding: This research was funded partly by Department of Veteran Affairs, Veterans Health Administration, Office of Research and Development; the Center for Innovations in Quality, Effectiveness and Safety (CIN 13-413), Sara NOWAKOWSKI, National Institutes of Health (NIH) grant, grant number R01NR018342, Javad RAZJOUYAN, National Heart, Lung, and Blood Institute (NHLBI) K25 funding (#: 1K25HL152006-01), and Investigator Initiate by Amir Sharafkhaneh and Javad Razjouyan by ZOLL Respicardia, Inc.

Institutional Review Board Statement: The study protocol was approved by the Research & Development Committee of the Michael E. DeBakey VA Medical Center and Baylor College of Medicine Institutional Review Board (IRB# H-35366).

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset is archived at the Corporate Data Warehouse behind the VA firewall and any official request required VA approval.

Acknowledgments: The opinions expressed are those of the authors and not necessarily those of the Department of Veterans Affairs, the US government, or Baylor College of Medicine.

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

References

- Gerstenslager, B.; Slowik, J.M. *Sleep Study*; StatPearls Publishing: Treasure Island, FL, USA, 2021.
- Mayo Clinic. Polysomnography (Sleep Study). 2022. Available online: <https://www.mayoclinic.org/tests-procedures/polysomnography/about/pac-20394877> (accessed on 10 January 2022).
- Shrivastava, D.; Jung, S.; Saadat, M.; Sirohi, R.; Crewson, K. How to interpret the results of a sleep study. *J. Community Hosp. Intern. Med. Perspect.* **2014**, *4*, 24983. [[CrossRef](#)] [[PubMed](#)]
- Bajeh, A.O.; Abikoye, O.A.; Mojeed, H.A.; Saliku, S.A.; Oladipo, I.D.; Abdulraheem, M.; Awotunde, J.B.; Sangaiyah, A.K.; Adewole, K.S. Application of computational intelligence models in IoMT big data for heart disease diagnosis in personalized health care. In *Intelligent IoT Systems in Personalized Health Care*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 177–206.
- Luo, L.; Li, L.; Hu, J.; Wang, X.; Hou, B.; Zhang, T.; Zhao, L. A hybrid solution for extracting structured medical information from unstructured data in medical records via a double-reading/entry system. *BMC Med. Inform. Decis.* **2016**, *16*, 114. [[CrossRef](#)] [[PubMed](#)]
- Elbattah, M.; Arnaud, E.; Gignon, M.; Dequen, G. The Role of Text Analytics in Healthcare: A Review of Recent Developments and Applications. In Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC), Vienna, Austria, 11–13 February 2021; pp. 825–832.
- Su, Y.-H.; Chao, C.-P.; Hung, L.-C.; Sung, S.-F.; Lee, P.-J. A Natural Language Processing Approach to Automated Highlighting of New Information in Clinical Notes. *Appl. Sci.* **2020**, *10*, 2824. [[CrossRef](#)]
- Murtaugh, M.A.; Gibson, B.S.; Redd, D.; Zeng-Treitler, Q. Regular expression-based learning to extract bodyweight values from clinical notes. *J. Biomed. Inform.* **2015**, *54*, 186–190. [[CrossRef](#)] [[PubMed](#)]
- Garvin, J.H.; DuVall, S.L.; South, B.R.; Bray, B.E.; Bolton, D.; Heavirland, J.; Pickard, S.; Heidenreich, P.; Shen, S.; Weir, C.; et al. Automated extraction of ejection fraction for quality measurement using regular expressions in Unstructured Information Management Architecture (UIMA) for heart failure. *J. Am. Med. Inform. Assoc.* **2012**, *19*, 859–866. [[CrossRef](#)]
- Garvin, J.H.; Kim, Y.; Temple Gobbel, G.; Matheny, M.E.; Redd, A.; Bray, B.E.; Heidenreich, P.; Bolton, D.; Heavirland, J.; Kelly, N.; et al. Automating quality measures for heart failure using natural language processing: A descriptive study in the department of veterans affairs. *JMIR Med. Inform.* **2018**, *6*, e9150. [[CrossRef](#)] [[PubMed](#)]
- Veena, G.; Hemanth, R.; Hareesh, J. Relation extraction in clinical text using NLP based regular expressions. In Proceedings of the 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT), Kannur, India, 5–6 July 2019.
- Sada, Y.; Hou, J.; Richardson, P.; El-Serag, H.; Davila, J. Validation of case finding algorithms for hepatocellular cancer from administrative data and electronic health records using natural language processing. *Med. Care* **2016**, *54*, e9. [[CrossRef](#)] [[PubMed](#)]
- Reeves, R.M.; Christensen, L.; Brown, J.R.; Conway, M.; Levis, M.; Gobbel, G.T.; Shah, R.U.; Goodrich, C.; Ricket, I.; Minter, F.; et al. Adaptation of an NLP system to a new healthcare environment to identify social determinants of health. *J. Biomed. Inform.* **2021**, *120*, 103851. [[CrossRef](#)]
- Ehrenfeld, J.M.; Gottlieb, K.G.; Beach, L.B.; Monahan, S.E.; Fabbri, D. Development of a natural language processing algorithm to identify and evaluate transgender patients in electronic health record systems. *Ethn. Dis.* **2019**, *29* (Suppl. 2), 441. [[CrossRef](#)]
- Gundlapalli, A.V.; South, B.R.; Phansalkar, S.; Kinney, A.Y.; Shen, S.; Delisle, S.; Perl, T.; Samore, M.H. Application of natural language processing to VA electronic health records to identify phenotypic characteristics for clinical and research purposes. *Summit Transl. Bioinform.* **2008**, *2008*, 36–40.
- Nowakowski, S.; Razjouyan, J.; Sharafkhaneh, A.; Kunik, M.; Naik, A. Polysomnographic Sleep Is Associated with Time to Develop Dementia: A Study Using 19-Year VA National EHR Data. *Innov. Aging* **2020**, *4* (Suppl. 1), 469. [[CrossRef](#)]
- Nowakowski, S.; Razjouyan, J.; Naik, A.; Agrawal, R.; Velamuri, K.; Singh, S.; Sharafkhaneh, A. 1180 The Use of Natural Language Processing to Extract Data from Psg Sleep Study Reports Using National Vha Electronic Medical Record Data. *Sleep* **2020**, *43*, A450–A451. [[CrossRef](#)]
- Khurshid, S.; Reeder, C.; Harrington, L.X.; Singh, P.; Sarma, G.; Friedman, S.F.; Di Achille, P.; Diamant, N.; Cunningham, J.W.; Turner, A.C.; et al. Cohort design and natural language processing to reduce bias in electronic health records research. *NPJ Digit. Med.* **2022**, *5*, 47. [[CrossRef](#)] [[PubMed](#)]

19. The Department of Veterans Affairs. VHA Corporate Data Warehouse (CDW). 2022. Available online: https://www.hsrd.research.va.gov/for_researchers/cdw.cfm (accessed on 12 January 2022).
20. Razjouyan, J.; Helmer, D.A.; Li, A.; Naik, A.D.; Amos, C.I.; Bandi, V.; Sharafkhaneh, A. Differences in COVID-19-related testing and healthcare utilization by race and ethnicity in the veterans health administration. *J. Racial Ethn. Health Disparities* **2022**, *9*, 519–526. [[CrossRef](#)] [[PubMed](#)]
21. Kang, H. Sample size determination and power analysis using the G*Power software. *J. Educ. Eval. Health Prof.* **2021**, *18*, 1149215. [[CrossRef](#)] [[PubMed](#)]
22. Ko, J.; Lim, H.K. Reliability Study of the Items of the Alberta Infant Motor Scale (AIMS) Using Kappa Analysis. *Int. J. Environ. Res. Public Health* **2022**, *19*, 1767. [[CrossRef](#)] [[PubMed](#)]
23. Chen, G.; Taylor, P.A.; Haller, S.P.; Kircanski, K.; Stoddard, J.; Pine, D.S.; Leibenluft, E.; Brotman, M.A.; Cox, R.W. Intra-class correlation: Improved modeling approaches and applications for neuroimaging. *Hum. Brain Mapp.* **2018**, *39*, 1187–1206. [[CrossRef](#)]
24. Python Software Foundation. 2022. Available online: <https://www.python.org/psf-landing/> (accessed on 16 January 2022).
25. NLTK. Natural Language Processing Toolkit. 2022. Available online: <https://www.nltk.org/> (accessed on 15 January 2022).
26. Li, B.; Chen, Y.W.; Chen, Y.Q. The nearest neighbor algorithm of local probability centers. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2008**, *38*, 141–154. [[CrossRef](#)]
27. Akgün, K.M.; Sigel, K.; Cheung, K.-H.; Kidwai-Khan, F.; Bryant, A.K.; Brandt, C.; Justice, A.; Crothers, K. Extracting lung function measurements to enhance phenotyping of chronic obstructive pulmonary disease (COPD) in an electronic health record using automated tools. *PLoS ONE* **2020**, *15*, e0227730. [[CrossRef](#)]
28. Velupillai, S.; Dalianis, H.; Hassel, M.; Nilsson, G.H. Developing a standard for de-identifying electronic patient records written in Swedish: Precision, recall and F-measure in a manual and computerized annotation trial. *Int. J. Med. Inform.* **2009**, *78*, e19–e26. [[CrossRef](#)]
29. Hripcsak, G.; Rothschild, A.S. Agreement, the F-Measure, and Reliability in Information Retrieval. *J. Am. Med. Inform. Assoc.* **2005**, *12*, 296–298. [[CrossRef](#)] [[PubMed](#)]
30. Huang, Y.J.; Powers, R.; Montelione, G.T. Protein NMR recall, precision, and F-measure scores (RPF scores): Structure quality assessment measures based on information retrieval statistics. *J. Am. Chem. Soc.* **2005**, *127*, 1665–1674. [[CrossRef](#)] [[PubMed](#)]
31. Loewen, A.H.; Korngut, L.; Rimmer, K.; Damji, O.; Turin, T.C.; Hanly, P.J. Limitations of split-night polysomnography for the diagnosis of nocturnal hypoventilation and titration of non-invasive positive pressure ventilation in amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler. Front. Degener.* **2014**, *15*, 494–498. [[CrossRef](#)] [[PubMed](#)]
32. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
33. Haghayegh, S.; Khoshnevis, S.; Smolensky, M.H.; Diller, K.R.; Castriotta, R.J. Accuracy of wristband Fitbit models in assessing sleep: Systematic review and meta-analysis. *J. Med. Internet Res.* **2019**, *21*, e16273. [[CrossRef](#)]