Generalizability and portability of natural language processing system to extract individual social risk factors

Tanja Magoc, PhD[a], **Katie S. Allen, BS**[b,d], **Cara McDonnell, MPH**[a], **Jean-Paul Russo, MPH**[a,c], **Jonathan Cummins, BS**[b], **Joshua R Vest, PhD, MPH**[b,d], **Christopher A. Harle, PhD**[b,d]

[a]College of Medicine, University of Florida, Gainesville, FL, USA

[b]Regenstrief Institute, Inc., Indianapolis, IN, USA

[c]Miller School of Medicine, University of Miami, Miami, FL, USA

[d]Richard M. Fairbanks School of Public Health, IUPUI, Indianapolis, IN, USA

Abstract

Objective—The objective of this study is to validate and report on portability and generalizability of a Natural Language Processing (NLP) method to extract individual social factors from clinical notes, which was originally developed at a different institution.

Materials and Methods—A rule-based deterministic state machine NLP model was developed to extract financial insecurity and housing instability using notes from one institution and was applied on all notes written during 6 months at another institution. 10% of positively-classified notes by NLP and the same number of negatively-classified notes were manually annotated. The NLP model was adjusted to accommodate notes at the new site. Accuracy, positive predictive value, sensitivity, and specificity were calculated.

Results—More than 6 million notes were processed at the receiving site by the NLP model, which resulted in about 13,000 and 19,000 classified as positive for financial insecurity and housing instability, respectively. The NLP model showed excellent performance on the validation dataset with all measures over 0.87 for both social factors.

Discussion—Our study illustrated the need to accommodate institution-specific note-writing templates as well as clinical terminology of emergent diseases when applying NLP model for social factors. A state machine is relatively simple to port effectively across institutions. Our study showed superior performance to similar generalizability studies for extracting social factors.

Conclusion—Rule-based NLP model to extract social factors from clinical notes showed strong portability and generalizability across organizationally and geographically distinct institutions.

With only relatively simple modifications, we obtained promising performance from an NLP-based model.

Keywords

Natural language processing; rule-based; social risk factors; portability; generalizability

1 BACKGROUND AND SIGNIFICANCE

As much as 80% of clinical information is stored within electronic health records (EHR) as unstructured text[1]. Information valuable to understanding a holistic view of the patient is contained within these clinical notes. Natural Language Processing (NLP) methods can extract relevant features from clinical notes or be used in cohort definition for a wide-range of conditions[2–17]. However, a longstanding challenge with any NLP model is the portability to new information system environments and generalizability to new populations[18–19]. Notably, documentation practices vary widely across providers, linguistics can vary between different locations, and availability of recommended compute infrastructure and trained informatics staff may differ among institutions[20–22].

These challenges exist in the application of NLP to identify social risk factors, which include the non-medical factors, like an individual's housing situation or socioeconomic status that affect health outcomes, utilization, and costs[23]. Social factors are often multidimensional constructs with numerous related, but non-synonymous terms that have crucial differences. For example, housing instability encompasses a range of issues such as frequent changes of residence, crowding within a household, the loss of housing, and homelessness[24]. Such nuanced differences increase the opportunities for linguistic variation. Similarly, NLP models may leverage the information from local or population-specific social services or programs[25]. Site-specific programs are useful information, but may not be generalizable. Additionally, social risk factors do not benefit from a standardized, controlled terminology to support generalizability across sites as is the case for clinical and medical terms.

Several studies have successfully developed NLP models to identify individual or multiple social risk factors[26]. However, only a very few studies have described the attempt to port and generalize an NLP model to extract a social risk factor that was originally developed at a different institution[27–28]. This study details the steps in porting and generalizing NLP models for two social risk factors (financial insecurity and housing instability) from one institution to another. We focus on portability and generalizability of the original system across distinct compute environments and data sets from different health systems by outlining the steps to install and update the NLP models and by evaluating performance metrics.

2 MATERIALS AND METHODS

In this study, we describe portability and generalizability of NLP models for extracting social risk factors from clinical notes that were developed from health systems in Central Indiana to the University of Florida Health (UF Health). We define portability as the ease of

installing the software in a new environment and to update the models to meet the needs of new data. We define generalizability as the performance of the NLP models when applied on new data.

2.1 Settings

Researchers from the Regenstrief Institute and Indiana University (IU) initially developed and tested NLP models using data from two separate health systems that participate in the Indiana Network for Patient Care (INPC)[29]. One was a safety-net system and the other a primarily urban, non-profit system. Together they represent approximately 1.25 million primarily urban (90%) patients, but differ in racial composition and social vulnerability.

The NLP system was ported and generalized to UF Health, which is an academic health center whose enterprise data warehouse contains information on about 2.4 million patients. The study involved data from multiple hospitals located in two demographically different urban areas (Gainesville, FL and Jacksonville, FL) and numerous clinics throughout urban and rural areas in Central Florida.

Health systems in Indiana and Florida serve geographically distinct areas with different local programs and services available to support patients with social needs.

2.2 NLP system

The NLP models were developed and deployed within the Regenstrief Institute's NLP engine nDepth[30]. Figure 1 illustrates nDepth's processing steps. nDepth deploys an independent deterministic finite state machine[31] for each social risk factor[32]. A state machine is used by nDepth to process thousands of notes simultaneously. nDepth stores the notes and processes them through a state machine, which contains a set of states and state transition rules, and a set of relevant keywords and phrases packaged into the state machine's dictionary.

nDepth takes a set of notes in a JavaScript Object Notation (JSON) file as the input and relies on a solr indexing and search engine to find the notes of interest, then processes these notes through the selected state machine. Each state machine starts with several preprocessing states to identify blocks of text that can be analyzed independently from the rest of the note, with sentence delimiters being natural stop points for each block. Each block of text is further analyzed through a series of states with each state deterministically transitioning the text block to one of the possible transition states or the accept state. Each state contains a rule specific to the social risk factor, such as the note contains one of the single words, 2-grams, and 3-grams denoting social risk factors from a dictionary, negation, or formal search spaces[32]. The output of a state machine is a binary result for each note (accept or reject) indicating whether the note meets the definition of the social risk factor.

2.3 Implementation

IU supplied UF with nDepth and the NLP models, which include the state machines and supporting dictionaries for each social risk factor. UF team edited configuration file, ran the installation script provided as a part of nDepth package, and deployed the service locally.

The state machine approach allowed the UF team to undertake simple addition, removal, and edits of states and dictionary terms to accommodate additional cases without the need to change the back-end of the software system or rebuild the executable files.

Figure 2 illustrates steps of porting and generalizing the NLP model to UF. After installing the IU-developed NLP model at UF, the performance was tested on an initial data set using manual annotation as the gold standard. The investigation of incorrectly classified notes in this test run was used to make modifications to the state machine and dictionary to accommodate local differences in notes. This step was followed by testing the performance of the updated NLP on the same set of notes, and finally, the performance was validated on a different set of notes.

2.4 Social risk factors

Two NLP models were designed to extract financial insecurity and housing instability from clinical text across various note types. Financial insecurity is a perceived inability to fulfill current and ongoing financial obligations without fear of not having enough money to cover necessary expenses[33–34]. Housing instability is a housing disruption including frequent moves, difficulty paying rent, eviction, or homelessness[35]. Figure 3 shows some examples how these social risk factors may be represented in clinical notes.

2.5 Clinical notes

The UF team applied the NLP models on all clinical notes written or edited between 3/1/2020–9/1/2020 for all adult UF Health patients. These notes included inpatient and outpatient settings, more than 50 different specialties, and encompassed inpatient progress notes, discharge notes, notes summarizing telephone encounters, letters mailed to the patients, and others.

Most notes contained only narratives written by healthcare providers, including physicians, nurses, and staff in training such as residents. However, some notes also contained sections that are automatically pulled from other fields in the EHR. Figure 4 shows an example of this section. Following the process established at IU and because the goal was to assess NLP's improvement over readily available information, these sections were ignored in the manual annotation process and were excluded by nDepth from further analysis.

2.6 Manual annotation

IU also provided a training manual and a synchronous training session for annotating notes. The training included an explanation of the concepts to be annotated as well as complex examples that IU's annotators had judged difficult to classify.

For each social risk category and each note type, the UF team randomly selected about 10% of notes that NLP classified as positive and the same number of notes, matched by the note type, that were classified as negative. All notes selected for manual annotation were randomly assigned to one of two annotators. The annotators included a pre-med student and a research coordinator, which emphasized the interpretation of notes by non-medical personnel due to the social rather than medical nature of the annotation categories. The

annotation consisted of marking each note as positive or negative with respect to the social risk category being evaluated.

100 notes per social risk category were annotated by both annotators to evaluate agreement. If kappa statistics did not show very good agreement, joint discussion with a member of the research team reviewed disagreements and provided clarification. When a specific pattern was found in disagreement, all notes following the same pattern were re-annotated by the original annotator.

2.7 Analysis

For each social risk category, we created two independent datasets for testing and validation. We split all manually annotated notes (see Table 1 for counts) in two sets, so that exactly a half of notes of each note type that were positively classified by the IU-developed NLP model were included in each set. Similarly, each set included a half of all manually annotated notes of each note type that were classified as negative by the IU-developed NLP model. One dataset was used to evaluate the initial performance of the NLP model (test run 1). We computed accuracy, positive predictive value (PPV), sensitivity, and specificity by utilizing manual annotation as the gold standard. We investigated notes that were misclassified by the NLP model to look for patterns that might have caused incorrect classification. After identifying such patterns, we modified the state machine by adding or editing existing states. We then ran the UF-updated NLP model on the same set of notes (test run 2) and calculated the same statistics as previously to determine the improvement in the NLP model. We aimed at PPV of at least 0.85 since the main objective of extracting data from notes is to supplement data available in structured form. Finally, we ran the UF version of the NLP model on the second dataset that was not used for updating the state machine (validation run), and we ran the same statistical analysis.

Finally, to describe the system's portability, we described and evaluated the difficulty in installing the system and editing the state machine.

3 RESULTS

3.1 Manual Annotation

Annotation showed very good agreement (0.84) on housing instability and substantial agreement (0.70) on financial insecurity. For financial insecurity, 40% of disagreements were due to annotator's differences on notes that mentioned a patient could not afford medication. One annotator read the note as consumer price sensitivity and interests in alternative medication choices and therefore did not mark notes as positive for financial insecurity. Through discussion the team clarified that the notes did indicate financial insecurity unless there was a clear indication that an alternative medication was available. After correcting the pattern of disagreement and reannotation, agreement increased to 0.82.

3.2 NLP models performance

We ran IU-developed NLP models on all notes written at UF Health during six months period in 2020. In this step, NLP processed more than six million notes belonging to about

330,000 patients, which resulted in just over 13,000 notes classified as positive for financial insecurity and almost 19,000 for housing instability. Randomly selected 10% of positively classified notes from each note type and the same number of negatively classified notes were manually annotated. Manually annotated notes were split into test and validation datasets, so that each dataset contains the same number of positively and negatively NLP-classified notes of each note type. Table 1 summarizes the number of notes in each step of the process.

The number of notes annotated per note type varies between social risk categories, however, progress notes accounted for more than 50% in both categories. Figure 5 shows the percentage of all note types that contributed at least 1% to the annotation batch.

In test run 1, where the original model from IU was applied to UF Health data, both financial insecurity and housing instability categories had similar accuracy (about 0.86) and PPV (just over 0.70) (see table 2). Investigation of false positives led to similar discoveries for both social risk categories. We identified several EHR-based templates that potentially had a negative impact on the performance of the NLP model. Templates either listed information about site-specific programs or included a set of questions and answers tagged into a note from structured data. Site-specific information, such as availability of reduced cost medication, is often provided to patients in specific clinics and documented in discharge notes regardless of whether the patients have this type of need, therefore not contributing to identifying patients at risk. Templated text that is generated from structured data does not provide additional information to what is already available in structured data, therefore, we ignored templates during both manual annotation and NLP data extraction. To reduce misclassifications by the NLP model, we modified the state machine to ignore templates. In particular, we created a new state that searches for the exact content of the site-specific information or the question in the structured data, and removed these blocks of text from further consideration.

In addition to EHR-based templates, notes used in this study were written during the initial days of COVID-19 pandemic mitigation strategies, and thus often referenced 'shelter in place', which resulted in the housing instability state machines incorrectly classifying these notes as positive. We modified the state machine to distinguish reference to shelter due to COVID-19 from other shelter references by creating a new state that removed the text block from further consideration if it contained a reference to 'shelter in place'.

We also encountered a large number of notes being false positive due to a specific phrase in the dictionary ('uninsured' in financial insecurity and 'living with (other family members)' in housing instability). Thus, we removed these phrases from the corresponding dictionaries since they do not constitute financial insecurity or housing instability on their own. However, removing 'uninsured' from the dictionary introduced numerous false negative classifications because in some cases, the lack of insurance was a direct cause of the inability of a patient to receive proper health care. Thus, we created a new state in the financial insecurity state machine to look for patterns to capture this causal relationship, such as 'due to * lack of insurance', where * represents up to 20 characters to cover for phrases such as his, her, their, etc.

After modifying the state machines and corresponding dictionaries, we re-ran the NLP models on the same set of notes (test run 2), which resulted in improved accuracy, PPV, and sensitivity for both social risk categories. Lastly, for the validation run, we ran the modified NLP model on a different set of notes. This produced excellent performance with all measures being over 0.83 for both financial insecurity and housing instability (see table 2).

4 DISCUSSION

In this study, we demonstrated portability and generalizability of a rule-based NLP system to extract social risk factors from clinical notes. Based on the performance of the implemented NLP, we found that free-text clinical notes were mostly consistent in describing social needs at geographically distinct institutions, and thus the models required only minimal adjustments when porting. The system installation required few steps and a state machine is relatively simple to port across institutions by adding or editing states. These modifications do not require an understanding of the back-end engines or rebuilding of executable files. Together, these characteristics make it relatively simple to port this type of NLP model for social risk factors across institutions.

Critically, minimal modifications of the state machines and dictionaries resulted in excellent performance of the NLP model on new data, therefore supporting generalizability of this approach across different institutions. For one, performance in this study compares favorably to other recent work on NLP for social risk factors. For example, Rouillard and colleagues[28] developed an NLP model that included housing needs. Like our results, their models were more specific than sensitive. However, their definition of housing needs included affective perceptions like safety that did not appear in our housing instability concept. Similarly to our approach, Reeves and colleagues[27] adopted and retrained a previously developed rule-based model to extract eight social and behavioral factors in a new patient population. They obtained similar performance metrics as in this study, but again were extracting different concepts.

Our experience highlighted several needs and challenges. First, this study exemplified the need for a robust framework[36] for training and annotation and for an iterative process in manually annotating clinical notes for social risk factors. Even though the annotators went through the same training, initial agreement was insufficient for subsequent model testing. Thus, calculating agreement and investigating differences early in the time-consuming manual annotation process may result in more efficient generalizability of NLP models. Second, our study illustrated the need to accommodate institution-specific note-writing templates. EHR templates such as predefined questions are increasingly common as option to help alleviate providers' documentation burdens[37]. NLP projects will have to dedicate more time to identifying and accounting for the impact of templated text and develop process to update models in response to new or changing templates. Lastly, both clinical concepts and social risk factors share common terminology, e.g. "shelter" in case of COVID-19. Whether new development or the application of an existing model, researchers would be advised to compare social risk factor dictionaries against current medical terminologies to avoid false positives.

5   CONCLUSION

Rule-based NLP models to extract financial insecurity and housing instability information from clinical notes showed strong portability and generalizability across organizationally and geographically distinct institutions. With only relatively simple modifications, we obtained promising performance from an NLP-based approach to identifying social risk factors.

ACKNOWLEDGEMENT

REFERENCES

1. Pak HS. Unstructured data in healthcare. Available from https://artificial-intelligence.healthcaretechoutlook.com/cxoinsights/unstructured-data-in-healthcare-nid-506.html [Accessed 11-10-2022]

2. Yim WW, Yetisgen M, Harris WP, et al. Natural Language Processing in Oncology: A Review. JAMA Oncol. 2016;2(6):797. [PubMed: 27124593]

3. Reading Turchioe M, Volodarskiy A, Pathak J, et al. Systematic review of current natural language processing methods and applications in cardiology. Heart. 2022;108(12):909–16. [PubMed: 34711662]

4. Le Glaz A, Haralambous Y, Kim-Dufor DH, et al. Machine Learning and Natural Language Processing in Mental Health: Systematic Review. J Med Internet Res. 2021;23(5):e15708. [PubMed: 33944788]

5. Savova GK, et al. Mayo clinical text analysis and knowledge extraction system (cTAKES): architecture, component evaluation and applications. Journal of Medical Informatics Association 2010;17:507–513.

6. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. Journal of American Medical Informatics Association. 2010;17:229–236.

7. Soysal E, et al. CLAMP–a toolkit for efficiently building customized clinical natural language processing pipelines. Journal of the American Medical Informatics Association. 2017;25:331–336.

8. Yang X, et al. A natural language processing tool to extract quantitative smoking status from clinical narratives. IEEE International Conference on Healthcare Informatics 2020.

9. Kent DM, Leung LY, Zhou Y, Luetmer H, Kallmes DF, Nelson J, Fu S, Zheng C, Liu H, Chen W. Association of Silent Cerebrovascular Disease Identified Using Natural Language Processing and Future Ischemic Stroke. Neurology 2021;97:e1313–e1321. [PubMed: 34376505]

10. Wang G, Jung K, Winnenburg R, Shah NH. A method for systematic discovery of adverse drug events from clinical notes. Journal of American Medical Information Association 2015;22:1196–1204.

11. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review. Journal of Medical Internet Research Medical Informatics 2019;7(2).

12. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. Journal of the American Medical Informatics Association 2019;26: 364–379. [PubMed: 30726935]

13. Sarmiento RF, DErnoncourt F. Improving patient cohort identification using natural language processing. In: Secondary analysis of electronic health records. Springer 2016:405–417.

14. Liao KP, Cai T, Savova GK, et al. development of phenotype algorithms using electronic medical records and incorporating natural language processing. BMJ 2015;350.

15. Kreimeyer K, Foster M, Pandey A, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. Journal of Biomedical Informatics 2017;74:14–29.

16. Sheikhalishahi S, Miotto R, Dudley JT, et al. Natural language processing of clinical notes on chronic diseases: systematic review. JMIR Med Inform 2019;7(2).

17. Koleck TA, Dreisbach C, Bourne PE, et al. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. Journal of the American Medical Informatics Association 2019;26(4):364–379. [PubMed: 30726935]

18. Carrell DS, Schoen RE, Leffler DA, et al. Challenges in adapting existing clinical natural language processing systems to multiple, diverse health care settings. J Am Med Inform Assoc. 2017;24(5):986–991. [PubMed: 28419261]

19. Wang Y, Wang L, Rastegar-Mojarad M, et al. Clinical information extraction applications: a literature review. J Biomed Inform 2018;77:34–49. [PubMed: 29162496]

20. Cohen GR, Friedman CP, Ryan AM, et al. Variation in physicians' electronic health record documentation and potential patient harm from that variation. J Gen Intern Med 2019;34(11):2355–2367. [PubMed: 31183688]

21. Carrell DS, Schoen RE, Leffler DA, et al. Challenges in adapting existing clinical natural language processing systems to multiple, diverse health care settings. J Am Med Inform Assoc. 2017;24(5):986–991. [PubMed: 28419261]

22. Meystre SM, Friedlin FJ, South BR et al. Automatic de-identification of textual documents in the electronic health record: a review of recent research. BMC Med Res Methodol 2010;10(70).

23. Hatef E, Ma X, Rouhizadeh M, Singh G, et al. Assessing the impact of social needs and social determinants of health on health care utilization: using patient- and community-level data. Popul Health Manag. 2021;24(2):222–30. [PubMed: 32598228]

24. Institute of Medicine. Capturing Social and Behavioral Domains in Electronic Health Records: Phase 2. Washington, DC: The National Academies Press 2014.

25. Chapman AB, Jones A, Kelley AT, et al. ReHouSED: a novel measurement of veteran housing stability using natural language processing. J Biomed Inform 2021;122.

26. Patra BG, Sharma MM, Vekaria V, et al. Extracting social determinants of health from electronic health records using natural language processing: a systematic review. J Am Med Inform Assoc 2021;28(12):2716–2727. [PubMed: 34613399]

27. Reeves RM, Christensen L, Brown JR, et al. Adaptation of an NLP system to a new healthcare environment to identify social determinants of health. J Biomed Inform 2021;120:103851. [PubMed: 34174396]

28. Rouillard CJ, Nasser MA, Hu H, et al. Evaluation of a natural language processing approach to identify social determinants of health in electronic health records in a diverse community cohort. Med Care 2022;1;60(3):248–255.

29. McDonald CJ, Overhage JM, Barnes M, et al. The Indiana Network For Patient Care: A Working Local Health Information Infrastructure. Health Affairs. 2005;24(5):1214–1220. [PubMed: 16162565]

30. Regenstrief Institute. What is nDepth? Available from: https://www.regenstrief.org/real-world-solutions/ndepth/what-is-ndepth [Accessed 11-20-2022]

31. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc. 2011;18(5):544–551. [PubMed: 21846786]

32. Allen KS, Hood D, Cummins J, et al. Natural Language Processing-driven State Machines to Extract Social Factors from Unstructured Clinical Documentation. JAMIA Open.

33. Sinclair RR, Cheung JH. Money Matters: Recommendations for Financial Stress Research in Occupational Health Psychology. Stress Health. 2016;32(3):181–193. [PubMed: 27400815]

34. Consumer Financial Protection Bureau. Financial Well-Being Scale: Scale development technical report [Internet]. Available from: https://www.consumerfinance.gov/data-research/research-reports/financial-well-being-technical-report [Accesses 02-01-2023]

35. Burgard SA, Seefeldt KS, Zelner S. Housing instability and health: findings from the Michigan Recession and Recovery Study. Soc Sci Med. 2012;75(12):2215–2224. [PubMed: 22981839]

36. Gupta AK, Kasthurirathne SN, Xu H, et al. A framework for a consistent and reproducible evaluation of manual review for patient matching algorithms. J Am Med Inform Assoc 2022;29(12):2105–2109. [PubMed: 36305781]

37. Avendano JP, Gallagher DO, Hawes JD et al. Interfacing with the Electronic Health Record (HER): a comparative review of modes of documentation. Cureus 2022:14(6).

HIGHLIGHTS

- Rule-based NLP model for extraction of social risk factors is easily portable and generalizable

- There is need to accommodate institution-specific note-writing templates when generalizing an NLP model

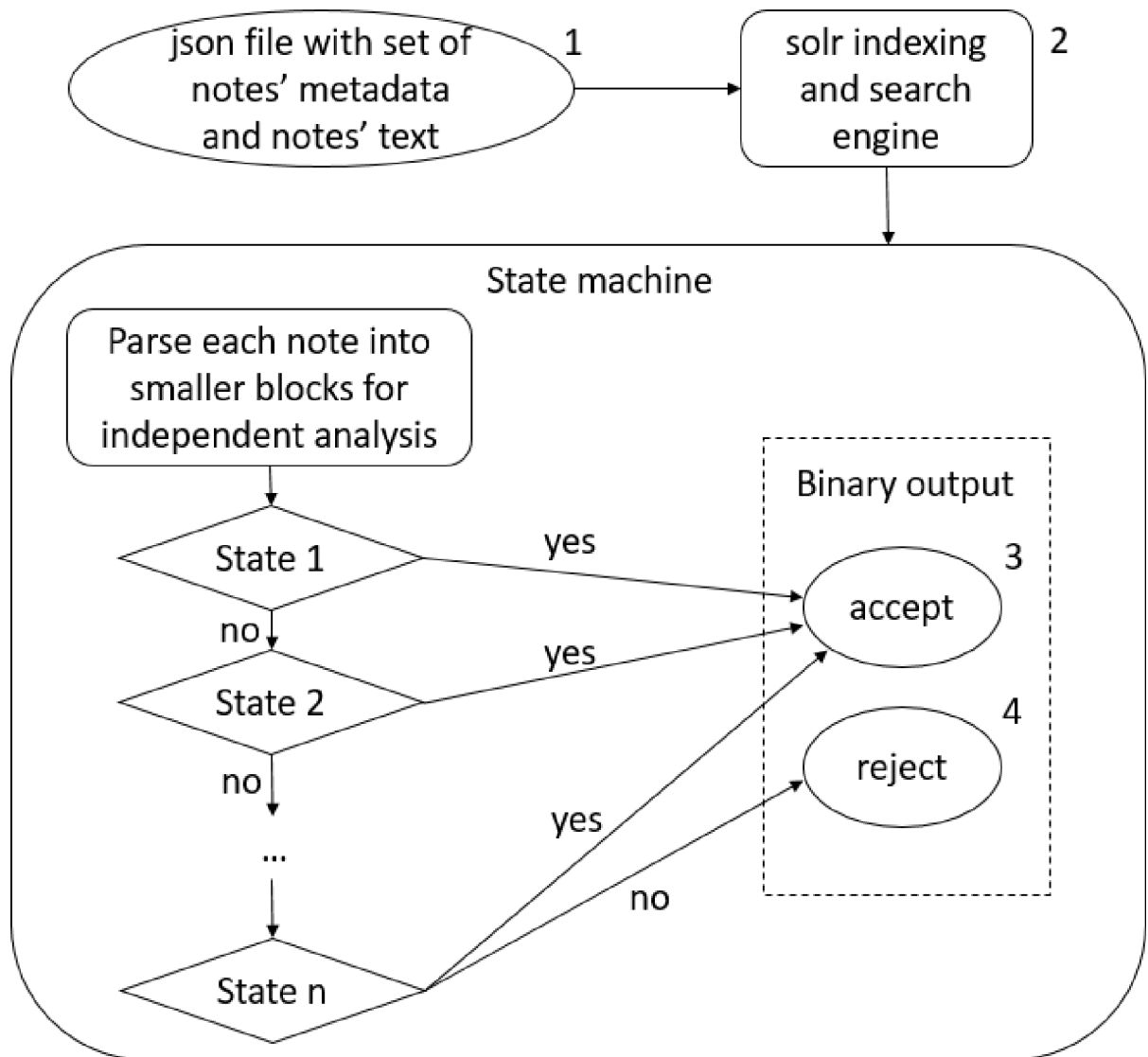- Rigorous manual annotation training is necessary for proper evaluation of NLP performance

Figure 1.
nDepth notes processing steps.

Notes: 1. nDepth engine takes a set of notes in JavaScript Object Notation (JSON) format and processes them in parallel. 2. Solr is an open-source Apache software that indexes the input and enables quick search for required phrases or patterns. 3. The 'accept' state indicates a note as 'positive' for a specific social factor; otherwise, 4. The 'reject' state indicates a note as 'negative'.
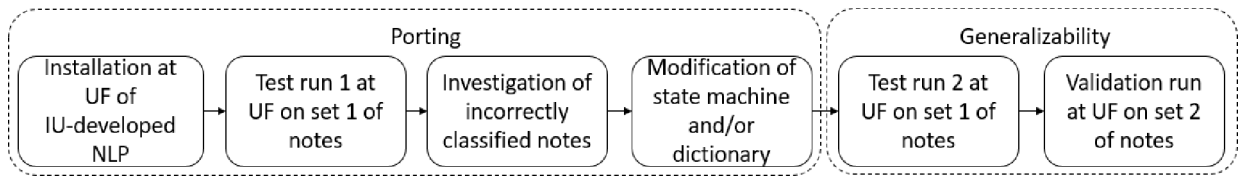
Figure 2.
Porting and generalizing to UF.

sedates her and she cannot keep using it to function. She admits to being started on escitalopram approximately 3 years ago, took it for now more than a month. She had lost insurance and could not afford the medication and so had to stop completely. She does admit after the loss of her mother, she began drinking heavily. Would drink hard liquor, finish a fifth and just pass out. This became a daily routine until

Pt reported a fear that she will not have financial means to find a place of her own post-discharge.

Type 2 diabetes mellitus (CMS-HCC: 19) [E11.9] Note Last Updated: [**DATE**] +CKD secondary to AIN from PPI- on lisinopril. +ASA. +Consultation placed for eye exam, but Pt has not yet gone (financial concern). - Pt was instructed to schedule appointment for annual diabetic eye exam on [**DATE**] (previous consultation placed but appt delayed from [**DATE**] because of financial concerns). ? Stage

gentleman, who was lethargic during interview. Prone to slurring of his words into one long run on sentence. Patient would also intermittently stop and state ""I am in soo much pain. I need something. I am in pain"". Patient stated he is currently homeless, not close to his family and has been '

response to therapy [**NAME**], PA-C 11:04 AM [**DATE**]--pt able to tolerated breakfast, ambulate and A&Ox3. Pt has spoken to her dad but isn't sure that she can stay with him. Pt will be d/c with resources on homeless shelters, clinic and uninsured clinics.

asked for & received multiple sandwiches & Gatorades. Insight/judgment @ baseline. Admits, ultimately, that the reason he went to & then asked to come to the hospital was because all the shelters were full. He is perseverative on him not having anywhere else to go and needing a place to stay. He then threatens suicide only in the context of being discharged. Patient appears to have a well-established lifestyle of homelessness and survival via agency

Figure 3.
Examples of text that indicates financial insecurity (the first three examples) and housing instability (the last three examples).

status: Single Spouse name: Not on file ? Number of children: Not on file ? Years of education: Not on file ? Highest education level: 10th grade Occupational History ? Not on file Social Needs ? Financial resource strain: Hard ? Food insecurity Worry: Sometimes true Inability: Sometimes true ? Transportation needs Medical: Yes Non-medical: No Tobacco Use ? Smoking status: Current Some Day Smoker

Figure 4.
An example of template text that is pulled into a note from structured fields in the EHR. Even though this text shows that the person has financial strain, it is ignored by manual annotators and the NLP model since this data already exists in structured format, and the purpose of the NLP model is to supplement readily available information.
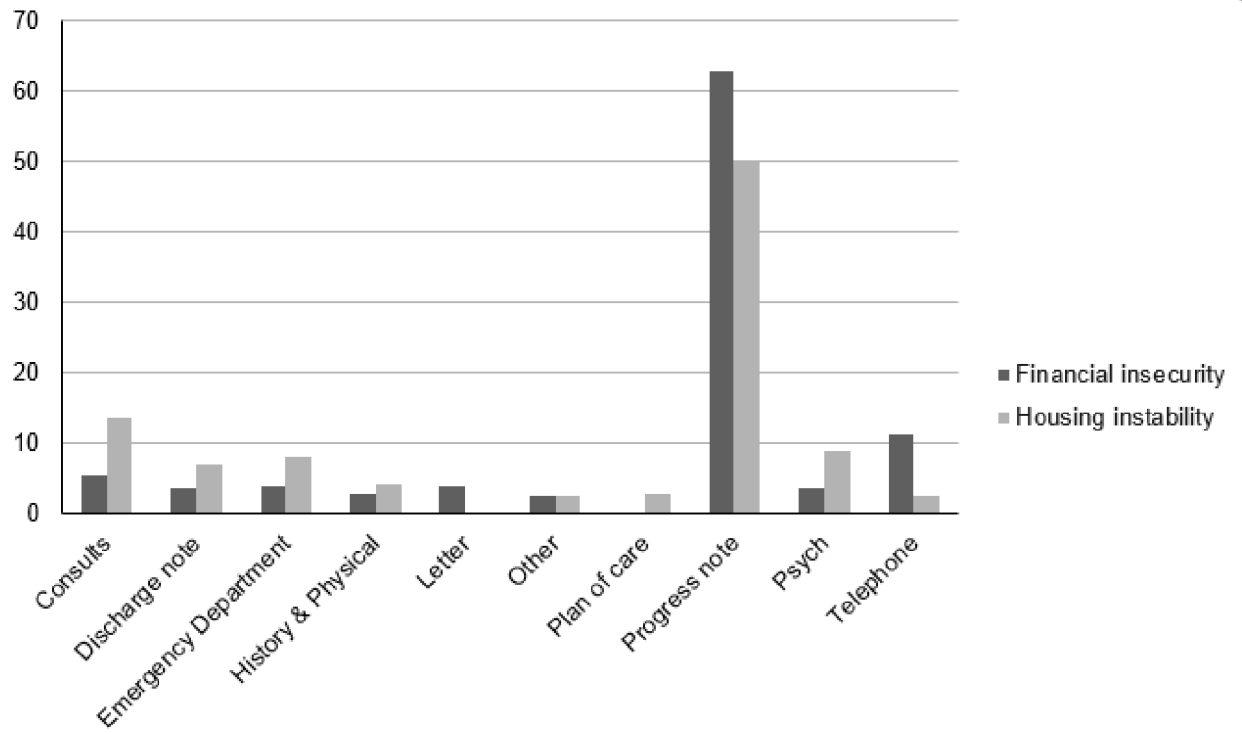
Figure 5.
The percent of top note types classified as positive by NLP model for each social factor category.

Table 1.

The number of notes processed initially by NLP, classified as positive by NLP, manually annotated, and included in test and validation datasets.

	Financial insecurity	Housing instability
# notes processed by NLP	6,250,127	6,250,127
# notes classified as positive by NLP	13,227	18,714
# notes classified as positive by NLP that were manually annotated	1,371	1,891
# manually annotated notes	2,742	3,782
# notes used in test runs	1,371	1,891
# notes used in validation run	1,371	1,891

Table 2.

Model performance for financial insecurity, housing instability, and both combined. Test run 1 was the initial run at UF Health; Test run 2 was after the state machine and dictionary were modified; Validation run was on a set of notes not used in developing the state machines.

	Financial insecurity			Housing instability			Total
	Test run 1	Test run 2	Validation run	Test run 1	Test run 2	Validation run	Validation run
Accuracy	0.86	0.92	0.91	0.87	0.94	0.92	0.91
PPV	0.72	0.90	0.90	0.74	0.91	0.90	0.90
Sensitivity	0.99	0.87	0.83	0.99	0.91	0.89	0.88
Specificity	0.79	0.95	0.95	0.79	0.95	0.93	0.94