AI Solutions for Health Equity: NLP and ML Approaches to Analyzing Social Determinants of Health

A Technical Report

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Andrew Lucktong

Spring 2024

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Rosanne Vrugtman, Department of Computer Science

AI Solutions for Health Equity: NLP and ML Approaches to Analyzing Social Determinants of Health

Andrew Lucktong
Computer Science
The University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia USA
anl8rpm@virginia.edu

ABSTRACT

Healthcare workers often experience trouble when it comes to standardization and integration of Social Determinant of Health (SDoH) data. Utilizing Artificial intelligence (AI) would allow for efficient processing and analyzing of patient data to predict at-risk groups and individuals based on the social determinants of health. Applying natural language processing (NLP) to medical records would allow a computer model to extract key information from a patient's electronic chart, providing more structured and organized data. This data could be used as input for a machine learning (ML) based clinical decision support system (CDSS) to provide personalized care plans and insights to health care providers caring for patients in areas negatively impacted by SDoH. Collecting SDoH data with AI would allow healthcare workers to create patient-centered care plans, reducing health disparities with proper semi-automated intervention. The future of this project involves designing a system to integrate into electronic health record systems, allowing streamlined workflow for healthcare providers.

1. INTRODUCTION

Many non-clinical factors, known as the social determinants of health (SDoH), can be used to determine potential health risks and outcomes. Exploring and analyzing social determinants of health and determining their effects on patient outcomes is important in providing equal and equitable care, but current data collection methods lack structure and uniformity. A 2023 survey of medical professionals conducted by NORC at the University of Chicago revealed that nearly eight in ten respondents collected SDoH data at their organization but experienced trouble with integration and standardization of this information (Social Determinants of Health Data, 2023).

AI offers a promising solution for automatic analysis of social determinant patient data. NLP is a subset of AI that gives computers the ability to understand and interpret human language in a meaningful way (IBM, n.d). By applying NLP to electronic health records (EHRs), computer models can extract vital information from patient charts, ensuring structured and organized data. With structured SDoH data, a machine learning (ML) based clinical decision support system (CDSS) could be used to help healthcare providers create patient-centered care plans. This project focuses on applying AI techniques to provide decision support for physicians with the goal of incorporating SDoH recommendations into personalized care plans.

2. RELATED WORKS

Clinical AI systems have seen an uptick in popularity as accuracy of predictive models

have improved dramatically in recent years. JvionAI, a clinical AI company, has developed an AI system called Jvion CORE that is able to provide specific recommendations for at-risk patients based on several different factors analyzed by machine learning. The system takes advantage of a combination of EHR and publicly available data and integrates seamlessly with existing healthcare technologies (Jvion, n.d.). By utilizing both current and prior data along with machine learning techniques, the system can learn and adapt to different healthcare systems depending on the needs of both patient and physician.

Many studies have also proposed AI-based solutions to predicting health outcomes based on SDoH, some with promising results. In a 2020 study, machine learning was used to predict whether patients would have emergency department or inpatient utilization within 90 days based on health determinant data. This study found that the model was able to predict utilization correctly 83% of the time when given unseen testing data. This is promising, considering machine learning is anticipated to improve as time goes on (Chen, et al., 2020).

3. PROPOSED DESIGN
There are many human and technical factors that would go into designing a system to analyze SDoH using AI and ML. Creating a system that is efficient is important, but focusing on human-centered design is crucial for designing a system that is effective. For this project, I propose two main technological pieces that work together: an NLP system to analyze patient chart data and extract SDoH data; and an ML-based CDSS to give healthcare workers recommendations based on this data. By using both technologies together and focusing on human-centered design, healthcare workers will be able to

collect and analyze SDoH data without unnecessarily increasing workload.

3.1 Architecture Overview
The NLP model must collect data in a way that can be interpreted by both human and machine. For this reason, I propose using an NLP technique known as named-entity recognition (NER). NER uses a predefined set of named entity categories to classify or identify words or phrases from unstructured text data. This is done by training an AI model on annotated data and then using that model to classify new text based on the training set. This technique would be used to identify and extract the many different social determinants of health, each with a specific label or category.

With structured data, an ML-based CDSS could be used to provide recommendations to healthcare providers. Non-knowledge-based clinical decision support systems use AI to create suggestions instead of using a predefined ruleset. These systems do not require rules for inferencing and can work with incomplete information, making them advantageous for analyzing SDoH data where there is no real ruleset for care plans (Sutton, et al., 2020). ML algorithms also tend to improve performance with more data, allowing this type of system to make better estimates of SDoH-related issues over time.

3.2 Human-centered design
The system would be designed with both patient and provider needs in mind. The models must be designed and implemented in a way that is non-intrusive and does not require a steep learning curve. Considering EHRs are used in most major healthcare settings across the United States, embedding and integrating the system into EHRs would create the most effective workflow. This involves working alongside EHR companies and healthcare workers to create a user-

friendly front end that allows users to interact with the system. NLP would be able to silently collect SDoH data in the background as the provider takes notes within the EHR and data would be made available to providers for research and statistical analysis. A CDSS could then be used to provide the provider with multiple care options based on their notes, helping them create the most effective care plan based on current and prior data. The system allows healthcare workers to focus more on the patient by requiring less time for manual SDoH data while also enriching their knowledge about patients and their community.

3.3 Challenges and Limitations

Current technologies and infrastructures have many limitations that are important to consider while designing this system. The biggest challenge to any technology in the healthcare industry is information safety and HIPAA compliance. Although there are no current written HIPAA restrictions on AI systems, we can expect to see them in the very near future. The system must also protect all private patient information to create public trust and ensure confidentiality. Integration also poses a potential challenge as most EHR providers are private companies with their own proprietary software that is not open source. This means developing a generalizable system is much more difficult, as integrating with different EHR systems may require different software solutions.

3.3.1 Solutions

Working with EHR companies may provide the simplest solution to most of the problems that come with designing and AI system for healthcare. These companies are already very knowledgeable about healthcare restrictions and requirements and work closely with users to design human-centered systems. Integrating into these systems would be simpler if working directly with the company that designs and maintains the infrastructure.

There may also be potential to create a generalizable system as research has been done to create a portable NLP system for analyzing SDoH data. In their 2023 paper, Magoc, et al. describe how they were able to successfully port an NLP system from Indiana University to a Florida Health system with both models showing similar identification accuracy. This research shows promise for creating a system that is compatible with a multitude of healthcare EHR systems and would be the most effective solution for creating a working system.

4. ANTICIPATED RESULTS

This system is anticipated to improve healthcare especially in areas affected by the social determinants of health. By leveraging AI and creating personalized care plans, it may be possible to decrease health disparities and provide equal and equitable care to everyone. With wide adoption, this system could aid research around SDoH by collecting structure data. The decision support system also has the advantage of using machine learning, meaning that the system will likely improve in performance as time goes on and more data is collected.

An AI system for collecting SDoH data may also help to decrease administrative workload for users by automating monotonous tasks. Typical SDoH collection methods require healthcare workers to manually label or fill in a patient's information. This system would remove that time-consuming process and make collection much easier. Furthermore, by using the recommendation system, healthcare workers gain a holistic view of patient information and prior care plans, allowing them to make more informed decisions.

5. CONCLUSION

As AI becomes more intertwined with everyday life, it is important to consider what applications it may have in the healthcare industry. By leveraging different AI techniques like NLP and ML, healthcare providers can offer more personalized and effective care plans. The proposed system helps to streamline data collection and can improve over time, ultimately working towards reducing health inequalities and enhancing patient outcomes in areas disproportionately affected by the social determinants of health.

6. FUTURE WORK

Developing an AI system for healthcare purposes would require working closely with EHR companies. This would greatly simplify implementation as many organizations already use these systems daily. EHR collaboration will also make NLP-based chart parsing and data collection much easier as the system could be integrated into the same environment.

Determining how healthcare workers use technology is essential to creating a usable system that will not interrupt regular workflow. Therefore, additional research needs to be done to determine clinician technology interaction patterns. Observing healthcare workers and determining how AI could assist will help to create a system that works for the client and not against them.

REFERENCES

Chen, S., Bergman, D., Miller, K., Kavanagh, A., Frownfelter, J., & Showalter, J. (2020). Using Applied Machine Learning to Predict Healthcare Utilization Based on Socioeconomic Determinants of Care. 26.

IBM. (n.d.). What is natural language processing?. IBM. https://www.ibm.com/topics/natural-language-processing

Implementing Advanced Technology in Healthcare. (n.d.). Jvion. Retrieved October 27, 2023, from https://jvion.com/approach/technical-implementation/

Magoc, T., Allen, K. S., McDonnell, C., Russo, J.-P., Cummins, J., Vest, J. R., & Harle, C. A. (2023). Generalizability and portability of natural language processing system to extract individual social risk factors. International Journal of Medical Informatics, 177, 105115.

Social Determinants of Health Data: Survey Results on the Collection, Integration, and Use. (2023). American Health Information Management Association (AHIMA).

Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. *Npj Digital Medicine*, *3*(1), Article 1. https://doi.org/10.1016/j.ijmedinf.2023.105115