International Journal of Scientific & Engineering Research Volume 12, Issue 8, August-2021
ISSN 2229-5518

Management of Individualized Therapy Options for Chronic Diseases via NLP through Clinical Document Analysis (CDA)

Bright Kyeremeh

Abstract—Improved therapies which supplement well beyond scientific proof treatments are needed due to the rising popularity of chronic diseases even amongst the worldwide people. A possible treatment is the secondary utilization of EHRs (electronic health records), wherein patient information is reviewed in order to carry out medical as well as translational research. Approaches related to Machine learning for analyzing EHRs are leading in a deeper understanding of patient medical histories as well as chronic disease risk predictions, enabling what was once an ability to obtain completely undiscovered treatment options. Behind unrestricted written clinical descriptions, though, a variety of clinical histories stay undetected. Consequently, to fully understand the possibilities of EHR statistics, the NLP (natural language processing) techniques which can normally move medical language into organized medical studies which can assist treatment strategies and perhaps prevent or avoid sickness development are necessary. The current study purpose was to give a detailed summary on NLP (natural language processing) approaches for unrestricted medical records linked with chronic diseases, as well as their development and implementation. When a patient having chronic disease (like Hypertension, Diabetes etc.) visits a doctor clinic, and the patient data is entered into the system, the NLP is to be used to process the previous clinical documentation and identify the case history identify red flags like previous heart surgery etc. and provide treatment options to the doctors. This helps to save time and effort for the doctors in analysis of previous history and treatment plan analysis. By giving therapy options along with their effectiveness, it will benefit the patient to have different options to choose better. The data collected is the patient case history information. Keywords that are programmed in the system like surgery, hypertension, and chemotherapy and so on are identified and processed in the Electronics health records and treatment plans are suggested.

Keywords—NLP, Chronic diseases, Therapy options

———————————— ♦ ————————————

I. INTRODUCTION

The coronavirus outbreak has somehow put a huge strain mostly on the global public health system, but has also revealed a deficit. Hospital staff, doctors, medical equipment, hospital wards, as well as other healthcare systems are all in high demand. Elective operations as well as normal treatments are viewed as non-essential, so they do not receive the recognition they deserve. For chronic disease people, this has created a twofold problem. They are not only receiving no care, but they are also at the highest risk of contracting Corona. Since COVID-19 has restricted masses of humans to their residences, the first and only realistic potential treatment for patients with chronic conditions is distant and digitized treatment.

Chronic diseases (for example diabetes, failure of heart, and hypertension) account for a lot of medical costs; the issues are that there are many more patients with not enough doctors to treat them. Structural algorithms are being used in various businesses to generate effective production better accessible, economical, and reliable. Is it important to improve treatment experience and minimize variability by making patient - centered care better computational, accessible, as well as cost-effective? That's one of the topics covered in this study.

Chronic Disease Management (CDM) refers to the integrated method of care for managing chronic diseases or ailments with long-term consequences. Cardiovascular disease as well as diabetes are the most costly and prevalent chronic diseases that people face, costing the healthcare system billions of dollars each year to treat and manage. Screening tests, consult, tracking and arranging medication, as well as patient care are among the most challenging components of handling chronic disease, besides from significant healthcare expenses Nothing will ever substitute a clinician's research evidence in healthcare delivery, but updated data technologies such as artificial intelligence (AI) have been helping to offer the right care to right person at the right time frame. As patient care evolves from responsive illness therapy to proactive, preventative care, many businesses are shifting to digitalization in healthcare to assist people generate useful insights from existing large data sources. To assist them, technological advances including machine learning and artificial intelligence are also being applied.

Recently, AI techniques have made huge waves in the healthcare industry, sparking a heated debate on if AI doctors will ever be able to take the position of human doctors. AI can undoubtedly help doctors in creating good medical judgment or perhaps even improve conventional judgment in certain aspects of patient care (e.g. radiology). The growing number of information regarding health as well as the rapid growth of big data analytical approaches have made considerable productive usage with AI in healthcare viable. Whenever driven by appropriate patient inquiries, strong AI algorithms may discover medically essential information contained in huge amounts of information that can improve clinical decision making [13–15].

NLP (Natural language processing) is a method which uses complex algorithms to detect definitions in free text, and is likely to be advantageous to the healthcare sector overall, rather than always to doctors and other medical experts. Natural language processing's possible advantages go beyond doctors and scientists to patients, according to modern innovations, including improving CDI programs (clinical documentation improvement), aiding research into chronic diseases, and supporting patients in taking a more active role in their own treatment. Clinical documentation improvement specialists are accustomed to manually examining records in the absence of NLP, with just minimal search functionality at their disposal.

CDI experts are useful to seek at a range of electronic data whenever they initiate a patient encounter, containing the physical assessment and the historical background, consultation notes, as well as a practical analysis. Without using an NLP technology, experts must individually scan every record for each patient's files for proposed amendments. That work can be taken up by NLP and completed in a shorter time. Because of its ability to remove the enormous quantities of information contained in EHRs, NLP is discovering a role in clinical investigation. Even when investigation into how natural language processing could indeed assist accelerate clinical research is now in its initial phases, Jonathan L. Haines, Doctorate, a scholar, board member of epidemiology as well as biostatistics, along with the director of Case Western Reserve University Institute of Computational Biology, thinks the possibilities for this to discover indications regarding different diseases is unlimited. *"Because of the volume and breadth of data collected, EHR data has enormous study potential,"* he says. *"One of the most significant challenges is that the majority of the data is unstructured, with the most of it being in free text. NLP is our method of converting unstructured input into structured data that may be used for research."*

The solution is based on three primary components:

- A comprehensive patient ontology
- Goal-based monitoring with linguistic involvement to help patients stay motivated with their health objectives.
- Interactive drawings to help patients do chores at home.

According to NLP algorithms created expressly in wellbeing scenarios, the digital assistant can recognize a lot of medical concepts through language and vocabulary. The patient and their physician can use this information to track their treatment plan's progress and overall success.

In order to assist individuals who've been identified with a chronic illness, AI technology may also detect disease indicators in people before they develop into chronic diseases. Machine learning algorithms based on AI can identify patients at risk of hypertension, cardiovascular disease, as well as pre-diabetes, allowing for early intervention and preventative treatment techniques. AI may potentially play a role in reducing hospitalization of chronic patients who are currently undergoing therapy. By monitoring and controlling patient vitals as well as medication adherence, methods based on AI can monitor the likelihood of worsening situations, requiring hospitalization.

II. LITERATURE REVIEW

Chronic diseases, including diabetes, cancer, as well as hypertension, are commonly acknowledged as major healthcare challenges. Although great progress throughout the detection of different therapies as well as inhibition approaches, such problems not only persists, however is becoming more prevalent [1], affecting patient standard of living as well as healthcare costs. As a consequence, to lessen the effect of chronic diseases on contemporary society, new solutions which enhance and go beyond conventional evidence-based treatment are necessary.

The subsequent use of EHRs (electronic health records) is an important area which can identify patient information, stimulate clinical research, as well as effectively assist treatment decisions. Techniques usually on EHR research [2] have resulted in improved patient classification as well as risk prediction [4-6], and also a deeper understanding of specific treatment histories [3]. Machine learning, specifically deep learning, applied to EHRs offers a once-in-a-lifetime chance to uncover earlier unknown treatment options [7]. This is certainly pertinent for chronic illnesses that generate a large and constant amount of information where the clinically significant trends can indeed be removed as well as used to inform medical judgment containing postponing or avoiding disease progression. EHRs are hard to analyze and predict owing to their high intricacy, noise, variability, sparseness, inadequacy, inaccuracies, and systematic biases. Moreover, since writing text is still the most intuitive as well as expressive way of recording clinical events, a great deal of patient's health history information is frequently buried behind unrestricted clinical narratives [8]. To turn medical text into organized medical studies which can be treated easily through machine learning algorithms, NLP techniques must be created. Considering implementations including detecting biological ideas using reports from radiology [9], paperwork for nurses [10], as well as directions for discharge [11], NLP is becoming increasingly commonly used in the healthcare sector. Structures dependent upon natural language processing used in clinical narratives, on the other hand, have not been frequently used in medical practice to assist decision support systems or processes.

As per a research report in Harvard Business Review [12], Paschalidis with his associates in Australia worked on projects during 2017 that was using electronic health records of patients as well as machine learning to forecast hospitalization owing to heart disease and diabetes. The scientists found that they might predict hospitalization up to a year beforehand using that same technique, with such an efficacy speed of approximately 82 percent. EHRs including real-time medical information, such as data through wearables, implanted procedures, as well as home-based connected medical tools, will also be used by Paschalidis along his associates to develop much more comprehensive estimation accuracy. This shows how machine learning algorithms could be used to classify individuals who are at a greater risk for cardiovascular or diabetes. Healthcare practitioners can employ such algorithms to allow for initial identification as well as tailored handlings for high-risk individuals.

The majority of the work in the Cardiovascular Diseases field centered on using natural language processing to assess the risk of coronary artery disease. For instance, Chen et al [16] proposed a unique workflow based on machine learning as well as guidelines to discover clinically relevant information pertaining to cardiovascular disease risk and track disease progression across collections of longitudinal patient data, comprising medical documentation (same as Torri et al [17]). Karystianis et al [18] as well as Yang et al [19] looked into identifying heart disease risk factors from diabetes patients' medical documentation. Roberts et al [20] developed a perhaps different approach, concentrating on estimating heart disease risk who used an eight-factor categorization system (such as aspirin). Aspirin are using as a health risk [21,22], echocardiographic heart functioning assessment [23], deep vein thrombosis particularly pulmonary embolism [24], as well as low-density lipoprotein content and statin use [25] have all been studied in this field. Among patients with atrial fibrillation, structured data as well as medical information are used to determine the effects of strokes and substantial bleeding [26], while medical documents only are being used to detect patients with heart failure [27]. Medical papers provided with, in Italian language, have also been used to identify arrhythmia episodes [28].

Most hypertension studies that focus on extracting essential indications, comorbidities, as well as pharmaceutical therapies are done through natural language processing [21]. Medical observations as well as different forms of clinical documents were used to detect hypertension patients that use the open-source medication IE (information extraction) software MedEx [49].

III Methodology

The journey starts with a diagnosis of diseases that may be prompted by patient-reported problems or regular check-ups. The data entered are of two types, structured and unstructured. The structured data refers to the patient demographics and basic information. The unstructured data refers to the clinical history, test results, doctor prescriptions etc. which are a part of the EHR. The picture, EP, as well as genetic information are machine-understandable following proper pre-processing and quality assurance steps, permitting various ML algorithms to be deployed right away. Furthermore, a significant portion of medical data, like a physical exam, diagnostic medical results, operations records, as well as discharge instructions, is now in the type of narrative language that is unorganized and unreadable to a computer system. In this circumstance, NLP seeks to extract useful information within narrative text to enhance therapeutic judgment. Once the patient data is retrieved, the system is trained to identify keywords pertaining to patient data and from the knowledge base created can identify the red flags. Probable therapy plans for the patient are presented which are then discussed with the patients by the doctor. Patients stick to the strategy and are continually monitored for effectiveness. On assessment, the doctor can update the plan if necessary.

The two most important aspects of an NLP pipeline are text processing as well as categorization. By text processing, the NLP discovers a series of disease-related terms in medical documentation looking at historical records. After that, a selection of the keyword is picked once their impacts, mostly on categorization of normal and anomalous situations, are evaluated. The verified words will then be added to the organized information that helps with medical decision support. NLP pipelines are designed to assist clinical decision-making regarding domains like notifying therapy regimens, tracking adverse effects, and etc. When such suggestions are provided to the doctor, the therapy options are presented to the Patient which helps to find the best course of treatment.

A. Working of Data Processing background of the Clinical Data

The procedure began with the creation of a keyword library related to the signs and symptoms of interest. The NLP modules were built using ANTLR (Another Tool for Language Recognition), which is a Java-based open source parsing developer. As per our specifications, this has been changed to contain numerous components. The program was educated just on the move with 2 small pilot local medical databases that were made accessible for development and testing in order to expand the lexicon and add regularly deployed shorthand abbreviations as well as terminology. Manual insertion of possible terms based on clinical notes was part of the training. It uses a rule-based approach to concept extraction. The tool employs part - of - speech tags, prefix, using dynamic typing, as well as ontologies and grammatical interpretation, to identify indicators and ailments from free-text documents. Another Tool for Language Recognition (ANTLR) refers to a powerful parsing engine able to read, process, execute, and translate structured text and binary files. It's commonly utilized in the creation of programming languages, tools, and frameworks. From grammar, ANTLR generates a parser which can construct and explore parse trees. ANTLR is a parser generator for computer-based language recognition which implements LL(*) for interpreting.

The parser LL is a highest level parser for one subset of context-free dialects (Left-to-right that is derivation towards Leftmost). When processing the inputs from left to right, this helps in accomplishing the Leftmost derivation of the statement. If such an LL parser uses k tokens from look ahead while processing a sentence, it is termed as LL(k) parser. The LL(k) grammar is something that can be constructed with the LL(k) parser. A formal language is called an LL(k) language while it possesses an LL(k) sentence structure. The collection of LL(k) dialects is adequately included in the group of LL(k+1) dialects for any k larger than 0. A corollary of this is that an LL(k) parser will not recognize all context-free languages. LL(*) and LL(finite) are two types of nomenclature outlier parsers (finite). If a parser utilizes the LL(*)/LL(finite) parsing strategy, it is called LL(*)/LL(finite). The LL(*) as well as LL(finite) parsers are much more equivalent to PEG parsers in terms of functionality. This research only involves using the ANTLR tool, not the details of its background programming.

B. Data Source Description

All the patient data collected were for patients having chronic diseases and those who visit the doctors on a regular basis.

Following are the stages in this research:

1. Training and verification of the Electronic Health Records also determine the accuracy of the data. Natural Language processing using the NLTK which is the Natural Language Processing Toolkit refers to a Python package which can be used for NLP.
2. Unstructured data with human-readable text makes up a large portion of the data you could be examining. Data must be pre-processed first before one can analyze it programmatically.
3. Processing patient information during a visit, fetching the patient data and analysis.
4. Providing treatment options to the doctor which in-turn be conveyed to the patient.

The data source considered was the patient data taken from the clinic under study. The clinic is a multispecialty/General consultation clinic. Doctors of all specializations like Diabetologist, Cardiologist, Neurologist, Urologist, and General Physician are a few to mention who visit the clinic. Patient data was collected over a period of 2 months. This time period was chosen to ensure that the patients come for monthly follow ups and visit is recorded twice at least to see the effectiveness of the system. In addition, the study was also included to see

C. Process Model

The model gives an insight into the system that is implemented. NLP (natural language processing) refers to a branch of computer science which aims to help computers comprehend natural human language.

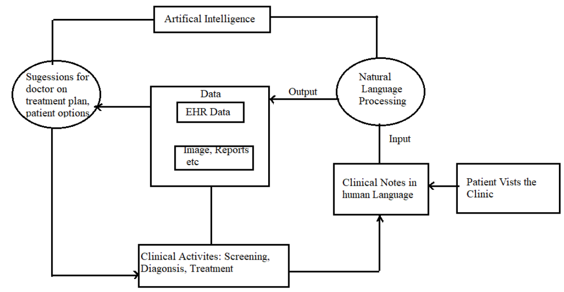


Figure 1: Research Model

Electronic Health Records present in the clinic were analyzed with the physical prescriptions and reports to check the validity of the EHR system. Patient data for the past five years consisting of 1250 records were identified through random sampling. The extracted EHR data was verified against the manual data using physical verification of the prescription and the electronic translation. Discrepancies were found and identified. With 625 manually annotated records, the algorithm was taught to create different words in order to increase its effectiveness, spelling mistake, as well as abbreviated techniques in the sequence and grammatical libraries which required to be upgraded, or removing keywords which resulted in a large number of false positives. The data was gone through and after numerous rounds of training, good results with the training dataset were obtained. Following training, the algorithm accuracy in detecting signs or symptoms were assessed mostly on leftover 625 comments which were not part of a training dataset. The accuracy of the findings were. In comparison to the adjudicated reference standard, performance was measured in terms of recall, precision, and F-measure for identifying symptoms and signs. This was done on a phrase per phrase and episode by episode basis. True positives are signs and symptoms which the algorithms correctly recognized, false - positive are signs and symptoms which the algorithms incorrectly identified, as well as false negatives are signs and symptoms which were overlooked, as described mostly by following formulae:

Precision = (True Positive + False Positive) / (True Positive + False Positive)

Recall = (True Positive + False Negative) / (True Positive + False Negative)

F–measure = (2 [Precision x Recall]) / (Precision + Recall)

Precision refers to the frequency in which problems indicated by the instrument are significant (positive predictive value). Recall refers to the regularity in which related symptoms are noted (sensitivity). To provide an overall impression of the tool performance, the F-measure, the weighted harmonic mean for recall and precision, was utilized. The F-measure can be used because real negatives are not available but do have a reference standard to compare with [31].

Finally, based on the symptoms identified and the patient history, possible treatment options are obtained and presented to the doctor. A qualitative examination of examples in which the NLP system has failed to accurately describe diseases, assertion levels, as well as symptom durations, and also a discussion of future development prospects.

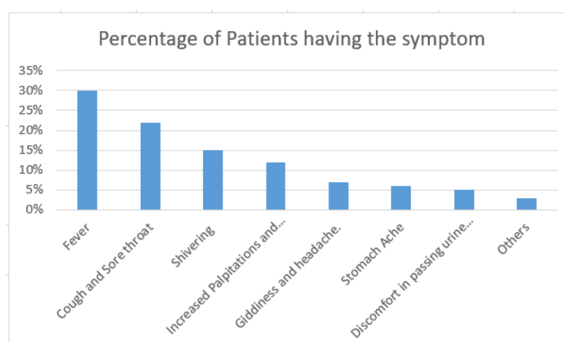
IV. RESULT

The results section gives a brief insight on the results obtained in this research. The table below, shows the commonly detected symptoms from the Electronic Health Records form the 1250 patient data that were analyzed. Symptoms sorted on the episode times occurring frequency:

Table 1: Patient Symptoms and Occurrence Percentage

Symptoms in Patient	**Percentage of occurrence of the symptom**
Fever	30%
Cough and Sore throat	22%
Shivering	15%
Increased Palpitations and breathlessness	12%
Giddiness and headache	7%
Stomach Ache	6%
Discomfort in passing urine and motion	5%
Others	3%

The graphical representation is also shown for the same. Whenever body has any infection, the first way the body shows is by developing a fever. The research here also indicates the same that in chronic patients also, the exhibited symptom of fever is high. Respiratory disorders/ infections, people with low immunity tend to develop a fever often, indicating that some additional steps need to be taken to find the root cause or preventive care.



The chronic diseases can be respiratory in the form of cough and sore throat, cardiology/ hypertension in the form of increased palpitations, giddiness etc., gastro related issues like frequent stomach ache, discomfort etc. Diabetic issues are related to low levels of sugar associated with shivering, sudden onset of weakness etc. Urology issues are also commonly occurring chronic issues. For example, in the case of kidney stones etc., discomfort in passing urine, pain in the related area and other related symptoms are observed.

Using the data present, the algorithm is trained to identify the symptoms and relate them to the knowledge base. The related information is then presented as therapy options to the doctor using internal healthcare tools. From the therapy options presented, the patient chooses the feasible one after consulting with the doctor and further course of action is taken.

A. The following is a result of determining the performance of the measurements developed here.

In order to determine the accuracy of the system developed here, three factors of precision %, Recall %, F – measure % are found in comparing the two cases –by use of algorithm after training of the algorithm is done and the standard set of values. Recalling the function of each of these parameters again. Precision is indeed the sum of positive category forecasts which specifically relate to the positive class. Recall seems to be the number of positive category assumptions made from several positive cases in the database. F-Measure calculates an overall value which takes into consideration all recall and precision problems. The formula mentioned within the methodology part is used to calculate the prediction performance.

An example of the calculation for precision is explained below:

Suppose that the model here makes prediction of 1250 examples using the manual and the symptoms are identified from the EHR and the treatment plan is presented to the doctor of which 1100 cases are correct and 150 are incorrect, then:

Precision = True Positive / (True positive + False Positive)

Precision = 1100 /(1100 + 150) = 0.88

% precision = 88%

Table 2: Comparative Performance of Manual and NLP Method for Different Factors

Performance measure	**Manual**	**NLP algorithm after training**
Precision %	88%	98%
Recall %	86%	97.6%
F – measure %	89.6%	98.7%

B. Principal Findings

The findings of this research have been presented in the previous sections. The key highlights are discussed here. It was found that the most commonly occurring symptom was fever, followed by cough and sore throat. These symptoms were found to be associated with chronic patients who had some allergies or prevailing respiratory issues causing low immunity overall. The system here was able to identify the same and from the knowledge base, patients having prolonged allergic reactions were given two options, to do an allergy test clinically or they could go to their environment and slowly try eliminating things and modifying the environment to avoid the onset of the allergy. Many patients were found to opt for this instead of the testing route. Similarly, in each of the chronic disease patients, the majority of the patients were found to choose a low intensity route that did not involve any aggressive treatment modes like heavy medication or surgery etc. depending on the illness.

However, it was found that, when the system alerts the doctor regarding red flags, the patients were found to go the heavy therapy route like emergency surgical intervention in case of cardiac issues, kidney issues and so on. It was also found on analysis of the patient file post the doctor visit that in most cases (89%) the doctor was found to make the same deductions the NLP algorithm system had identified in the treatment plans. The overall effectiveness and accuracy of the system was also found which indicated the success of this method.

V. CONCLUSION, LIMITATIONS, AND FUTURE SCOPE

A. Conclusion

The onset of healthcare system management through digital methods, tremendous amounts of data on patient activity, patient-reported outcomes, patient history, and treatment plan thanks are available, which collects data from various touch-points throughout the patient lifecycle—including wearable, mobile apps, and the hospital EHR system. With the use of AI tools, like NLP this data can offer new insights and opportunities for economical and scalable treatment. This data can be used to prioritize patients based on real-time needs, produce intervention warnings, and recommend follow-up activities using machine learning.

In conclusion, Natural Language Processing (NLP) tools have the potential to improve patients' and clinicians' point-of-care decision making, and a viable business model is required to ensure that safe, effective clinical AI systems are developed, validated, and sustainably deployed, integrated into EHR systems, and curated over time to maintain adequate accuracy and reliability. The system developed here effectively reduces a lot of time and effort for the doctor and is capable of presenting the treatment options to the patient, making both the sides to be at an advantage.

B. Research Limitations

In many places, doctors are very apprehensive about use of Natural language processing techniques and the AI methods for treatment planning as there are always chances of misdiagnosis when the system comes to the wrong deduction about patient illness. In addition, the training phase of the algorithm is tedious and there are also many instances where the system does not recognize some abbreviations and notations of the doctor when they are hand – written. In those instances, manual intervention is also necessary.

C. Future Scope

The future scope for this area of research is very vast and there is a lot of room for new findings. More technological advancements can be implemented in the research. Here, manual interference was there in order to translate some of the handwritings, medical abbreviations etc. of the doctor, however, a more automated NLP system can be tried in the future.

VI. REFERENCES

1. World Health Organization. WHO Global status report on non-communicable diseases 2014 URL: https://www.who.int/nmh/publications/ncd-status-report-2014/en/ [accessed 2021-07-22] [WebCite Cache ID 77Fa8uXax]

2. Kruse CS, Kothman K, Anerobi K, Abanaka L. Adoption factors of the electronic health record: a systematic review. JMIR Med Inform 2016 Jun 01;4(2):e19. [doi: 10.2196/medinform.5525] [Medline: 27251559]

3. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016 Dec 17;6:26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]

4. Jensen P, Jensen L, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012 May 02;13(6):395-405. [doi: 10.1038/nrg3208] [Medline: 22549152]

5. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc 2017 Jan;24(1):198-208. [doi:10.1093/jamia/ocw042] [Medline: 27189013]

6. Ye C, Fu T, Hao S, Zhang Y, Wang O, Jin B, et al. Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning. J Med Internet Res 2021 Jul 22;20(1):e22 [FREE Full text] [doi: 10.2196/jmir.9268] [Medline: 29382633]

7. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2017 May 06. [doi: 10.1093/bib/bbx044] [Medline: 28481991]

8. Jensen K, Soguero-Ruiz C, Oyvind MK, Lindsetmo R, Kouskoumvekaki I, Girolami M, et al. Analysis of free text in electronic health records for identification of cancer patient trajectories. Sci Rep 2017 Dec 07;7:46226 [FREE Full text] [doi: 10.1038/srep46226] [Medline: 28387314]

9. Flynn R, Macdonald TM, Schembri N, Murray GD, Doney ASF. Automated data capture from free-text radiology reports to enhance accuracy of hospital inpatient stroke codes. Pharmacoepidemiol Drug Saf 2010 Aug;19(8):843-847. [doi: 10.1002/pds.1981] [Medline: 20602346]

10. Popejoy LL, Khalilia MA, Popescu M, Galambos C, Lyons V, Rantz M, et al. Quantifying care coordination using natural language processing and domain-specific ontology. J Am Med Inform Assoc 2015 Apr;22(e1):e93-e103 [FREE Full text] [doi: 10.1136/amiajnl-2014-002702] [Medline: 25324557]

11. Yang H, Spasic I, Keane JA, Nenadic G. A text mining approach to the prediction of disease status from clinical discharge summaries. J Am Med Inform Assoc 2009;16(4):596-600 [FREE Full text] [doi: 10.1197/jamia.M3096] [Medline: 19390098]

12. *How Machine Learning Is Helping Us Predict Heart Disease and Diabetes*. (2020, November 4). Harvard Business Review.

https://hbr.org/2017/05/how-machine-learning-is-helping-us-predict-heart-disease-and-diabetes

13. Murdoch TB , Detsky AS . The inevitable application of big data to health care. JAMA 2013;309:1351–2.doi:10.1001/jama.2013.393

14. Kolker E , Özdemir V , Kolker E . How Healthcare can refocus on its Super-Customers (Patients, n =1) and Customers (Doctors and Nurses) by Leveraging Lessons from Amazon, Uber, and Watson. OMICS 2016;20:329–33.doi:10.1089/omi.2016.0077

15. Dilsizian SE, Siegel EL . Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. Curr Cardiol Rep 2014;16:441.doi:10.1007/s11886-013-0441-8

16. Chen Q, Li H, Tang B, Wang X, Liu X, Liu Z, Liu S, Wang W, Deng Q, Zhu S, Chen Y, Wang J. An automatic system to identify heart disease risk factors in clinical texts over time. *J Biomed Inform.* 2015 Dec;58 Suppl:S158–S163. doi: 10.1016/j.jbi.2015.09.002. https://linkinghub.elsevier.com/retrieve/pii/S1532-0464(15)00194-X.

17. Torii M, Fan J, Yang W, Lee T, Wiley MT, Zisook DS, Huang Y. Risk factor detection for heart disease by applying text analytics in electronic medical records. *J Biomed Inform.* 2015 Dec;58 Suppl:S164–S170. doi: 10.1016/j.jbi.2015.08.011. https://linkinghub.elsevier.com/retrieve/pii/S1532-0464(15)00178-1.

18. . Karystianis G, Dehghan A, Kovacevic A, Keane JA, Nenadic G. Using local lexicalized rules to identify heart disease risk factors in clinical notes. *J Biomed Inform.* 2015 Dec;58 Suppl:S183–S188. doi: 10.1016/j.jbi.2015.06.013. https://linkinghub.elsevier.com/retrieve/pii/S1532-0464(15)00123-9.

19. Yang H, Garibaldi JM. A hybrid model for automatic identification of risk factors for heart disease. *J Biomed Inform.* 2015 Dec;58 Suppl:S171–S182. doi: 10.1016/j.jbi.2015.09.006. https://linkinghub.elsevier.com/retrieve/pii/S1532-0464(15)00199-9.

20. Roberts K, Shooshan SE, Rodriguez L, Abhyankar S, Kilicoglu H, Demner-Fushman D. The role of fine-grained annotations in supervised recognition of risk factors for heart disease from EHRs. *J Biomed Inform.* 2015 Dec;58 Suppl:S111–S119. doi: 10.1016/j.jbi.2015.06.010. https://linkinghub.elsevier.com/retrieve/pii/S1532-0464(15)00120-3.

21. Pakhomov S, Shah N, Hanson P, Balasubramaniam S, Smith S. Automated processing of electronic medical records is a reliable method of determining aspirin use in populations at risk for cardiovascular events. *Inform Prim Care.* 2010;18(2):125–133. http://hijournal.bcs.org/index.php/jhi/article/view/762.

22. Zheng C, Rashid N, Koblick R, An J. Medication extraction from electronic clinical notes in an integrated health system: a study on aspirin use in patients with nonvalvular atrial fibrillation. *Clin Ther.* 2015 Sep;37(9):2048–2052. doi: 10.1016/j.clinthera.2015.07.002. [PubMed] [CrossRef] [Google Scholar]

23. Patterson OV, Freiberg MS, Skanderson M, Brandt CA, DuVall SL. Unlocking echocardiogram measurements for heart disease research through natural language processing. *BMC Cardiovasc Disord.* 2017 Dec 12;17(1):151. doi: 10.1186/s12872-017-0580-8. https://bmccardiovascdisord.biomedcentral.com/articles/10.1186/s12872-017-0580-8.

24. Tian Z, Sun S, Eguale T, Rochefort CM. Automated extraction of VTE events from narrative radiology reports in electronic health records: a validation study. *Med Care.* 2017 Dec;55(10):e73–e80. doi: 10.1097/MLR.0000000000000346. http://europepmc.org/abstract/MED/25924079.

25. Ross EG, Shah N, Leeper N. Statin intensity or achieved LDL? Practice-based evidence for the evaluation of new cholesterol treatment guidelines. *PLoS One.* 2016;11(5):e0154952. doi: 10.1371/journal.pone.0154952. http://dx.plos.org/10.1371/journal.pone.0154952.

26. Wang SV, Rogers JR, Jin Y, Bates DW, Fischer MA. Use of electronic healthcare records to identify complex patients with atrial fibrillation for targeted intervention. *J Am Med Inform Assoc.* 2017 Mar 01;24(2):339–344. doi: 10.1093/jamia/ocw082.

27. Pakhomov S, Weston S, Jacobsen S, Chute C, Meverden R, Roger V. Electronic medical records for clinical research: application to the identification of heart failure. *Am J Manag Care.* 2007 Jun;13(6 Part 1):281–288. https://www.ajmc.com/pubMed.php?pii=3319.

28. Viani N, Larizza C, Tibollo V, Napolitano C, Priori SG, Bellazzi R, Sacchi L. Information extraction from Italian medical reports: an ontology-driven approach. *Int J Med Inform.* 2018 Mar;111:140–148. doi: 10.1016/j.ijmedinf.2017.12.013.

29. Teixeira PL, Wei W, Cronin RM, Mo H, VanHouten JP, Carroll RJ, LaRose E, Bastarache LA, Rosenbloom ST, Edwards TL, Roden DM, Lasko TA, Dart RA, Nikolai AM, Peissig PL, Denny JC. Evaluating electronic health record data sources and algorithmic approaches to identify hypertensive individuals. *J Am Med Inform Assoc.* 2017 Jan;24(1):162–171. doi: 10.1093/jamia/ocw071. http://europepmc.org/abstract/MED/27497800.

30. Wikipedia contributors. (2021, July 19). *LL parser*. Wikipedia. https://en.wikipedia.org/wiki/LL_parser

31. Powers D. 2015. Mar 22, [2018-02-10]. What the F-measure doesn't measure: Features, Flaws, Fallacies and Fixes http://arxiv.org/abs/1503.06410 *webcite*.