NLP in Healthcare

Use Cases and Trends in 2022 and Beyond

Table of Contents

Abstract

In the Healthcare industry, the amount of data that is now generated, stored, and analyzed has gone up exponentially. Electronic Health Records (EHR) in particular dominate the sector. EHRs provide data-driven opportunities for Healthcare providers to deliver better care and improve diagnostic and treatment outcomes for patients. However, since EHRs usually contain unstructured textual data, providers often struggle to put it to work. Natural Language Processing (NLP) provides the solution to this challenge.

NLP enables Healthcare entities to convert unstructured EHR data into a more digestible format and make better sense of it. They can leverage NLP to improve clinical data integrity, summarize lengthy blocks of medical text, find symptom attributes, identify at-risk patients, and segment named entities into predefined categories. Many NLP systems even “learn” over time to improve results’ accuracy and usefulness.

This eBook explores NLP use cases in Healthcare. The goal is to demonstrate the tangible ways NLP can impact Healthcare, and improve the quality and longevity of human life.

Recent NLP Breakthroughs and Implications for Healthcare

> *I am convinced that machines can and will think in our lifetime*
>
> *Oliver Selfridge, The Thinking Machine (1961)*

Recently, several breakthroughs have blown open the possibilities of NLP in Healthcare.

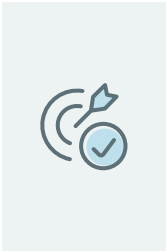
Pre-trained Medical Language Models

In the 2020 NLP Industry Survey[1] , over 40% of respondents said that they consider accuracy as the most important criterion when evaluating NLP technologies. Here, accuracy refers to the availability of pre-trained models and algorithms that allow users to input text to get common outputs, such as tokens, lemmas, part-of-speech, similarity, and entity recognition. In Healthcare, transformer models that are already pre-trained on clinical data and familiar with medical and clinical jargon, language, codes, etc., can deliver state-of-the-art performance on many clinical NLP tasks. Moreover, they also help eliminate one of the most painful problems hitherto seen in Healthcare NLP – the shortage of labeled data.

Accuracy in Healthcare NLP is important because it allows providers to use EHRs in more meaningful ways to deliver better clinical intervention and care. Further, clinicians can integrate NLP into their daily workflows with minimal risk of inaccuracies that are not acceptable in a life-critical sector like Healthcare.

Explainable & Interpretable NLP

In the past, NLP models provided no human-readable explanations for the predictions they made. This "black box" approach made it difficult for users to trust the model's outcome, and thus prevented them from making the best possible decision or action to benefit patients.



Explainable/interpretable NLP provides insights for transformer models, which clinicians can use to understand the models' choices. They can also understand the model's impact and potential biases, and investigate, track and optimize the model's performance to ensure that it maintains accuracy, fairness, and transparency. All these qualities can increase clinician trust in the system, and enable them to make better NLP-powered decisions.

Bidirectional Encoder Representations from Transformers (BERT)

Bidirectional Encoder Representations from Transformers or BERT is an open-source, pre-trained neural language model for NLP. BERT helps machines understand the meaning of language in the text by using surrounding text to establish context.

Recent developments in BERT are driving advances in NLP for Healthcare. BERT-based models can now deliver state-of-the-art performance for biomedical and clinical entity normalization. They can also be fine-tuned to normalize different kinds of named entities.

The Med-BERT[2] model adapts BERT to the structured EHR domain. Even with small local training datasets, Med-BERT can accurately predict diseases from clinical databases. In future, Med-BERT will help accelerate developments in AI and NLP-aided Healthcare.

John Snow Labs **has trained its own healthcare-specific BERT** word embeddings, and is unique in providing Healthcare-specific sentence embeddings. These embeddings support transfer learning, represent implicit relationships between words, and make it possible to represent Healthcare-specific text numerically and contextually in a way that NLP algorithms and machine learning models can understand and work with.

GPT-3

Generative Pre-trained Transformer or **GPT-3** is a highly capable language model for NLP. This Open AI model contains 175 billion parameters, which constitute nearly a trillion words. Previous language models were applied to NLP tasks by first providing an initial task model via transfer learning, and then tuning the model using traditional machine learning. GPT-3 not only eliminates the need for fine-tuning models for each NLP task, but it can also be used directly for many different tasks with very limited guidance.

In Healthcare, **GPT-3** may be leveraged to simplify medical information searches, summarize medical articles, and pull out relevant patient information from various documents. **GPT-3** could also play a major role in the development of health apps and software.

NLP Use Cases in Healthcare

> **Modern NLP technologies** are helping to transform raw Healthcare data into valuable information for a wide range of use cases.

Improved Patient Flow Forecasting

In clinical settings unexpected surges inpatient demand lead to bed and staffing shortages can often be life-threatening. To ensure that enough beds and nurses are available to take care of incoming patients, accurate forecasting is critical.

As part of a broader AI solution, NLP can analyze key of several factors that influence patients flows, such as volume and timing of arrivals, patients' length of stay, admission specialties, and more. By predicting future bed demand, NLP can help staff make real-time networked decisions and perform strategic planning to eliminate bottlenecks and optimize patient flow management.

Automated Understanding of Scanned Medical Reports and Documents with OCR

In Healthcare, OCR is critical to read unstructured datasets, such as clinical notes, patient intake forms, medical tests, medical history records, etc, and present them in a more digestible, human-readable form. This formatted data can be fed into an NLP pipeline for further analysis.

Further, Healthcare professionals can leverage a combination of NLP, OCR, and computer vision models to:

- Classify documents
- Extract normalized entities and figures
- Find signatures on forms
- De-identify images.

Summarizing Medical Records and Other Documents

Clinicians and other Healthcare professionals are exposed to a lot of patient data through forms, medical history, discharge records, and other documents that can be very useful to improve the quality and timeliness of care. However, the sheer volumes of information generated means that a lot of it is not read, much less used. NLP systems can present key information from various text sources in a summarized format. By identifying key concepts or phrases in the source material, NLP can summarize lengthy blocks of text so Healthcare professionals can quickly see the most relevant information to better help patients, and make their own lives easier.

Detecting Adverse Drug Events

According to the latest data from the U.S. health.gov website[3]

Each year, **Adverse Drug Events (ADE)** such as medication errors, adverse drug reactions and overdoses affect about **2 million** hospital stays in inpatient settings. They also prolong hospital stays by **1.7 to 4.6 days**. In outpatient settings, ADEs result in about 1 million emergency department visits and **125,000 hospital** admissions. Unfortunately, many of these instances are never reported, making them harder to discover.

NLP solutions make it possible to monitor reactions to drug events, and thus effectively and accurately detect ADEs earlier. This helps decrease their likelihood of occurrence, and reduce their impact, ultimately, making the Healthcare system safer for both patients and Healthcare practitioners.

Building Patient Cohorts

NLP is useful for building patient cohorts – a key aspect of clinical research – by recognizing and normalizing clinical entities such as diagnoses, procedures, drugs, labs, and social determinants from patient records and assign an ICD-10-CM code to them. Such cross-referencing can help monitor mortality and quality statistics for different conditions to understand possible complications, and design better patient treatment plans. The use of Named Entity Recognition (NER) also enables providers to segment named entities, e.g. persons, locations, products, or organizations into predefined categories to further improve their ability to build the right cohorts.

Clinical Decision Support

Due to the extensive use of EHR platforms, Healthcare professionals now have access to a wide range of useful information from clinical notes, pathology reports, radiology impressions, etc. Real-time NLP pipelines can ingest these different types of clinical documents, extract information from them, and generate different risk stratification scores to improve clinical decision-making in many acute care settings.

One such NLP-enabled **Clinical Decision Support (CDS)** was tested and successfully demonstrated for cervical cancer screening and surveillance at Mayo Clinic Rochester between **2014** and **2016**. In the study, the NLP-enabled CDS system captured both discrete and free-text information in patient records to inform providers about the correct screening and surveillance intervals for preventing cervical cancer among young women. The new system not only provided real-time screening and surveillance to providers, but also improved surveillance rates from **5.7% to 23.7%**, opening new doors to improve cancer screening, surveillance and prevention.

Healthcare providers can leverage an NLP clinical assertion to analyze their clinical notes, check symptoms, and improve their ability to diagnose and treat patients. Further, NLP with predictive analytics can help providers and clinicians analyze free text reports and other documents to speed up diagnoses, discover critical complications early, and identify high-risk patients to improve care delivery.

Computational Phenotyping

Phenotyping enables clinicians to group patients to get deeper insights into data and compare cohorts. NLP allows clinicians to extract and analyze unstructured data, which not only makes more information available to create phenotypes, but also helps create richer phenotypes that can help with the identification of neuro-cognitive damages, cardiovascular diseases, psychological disorders, and other kinds of complex or latent diseases.

Patient Risk Adjustment

A risk adjustment model assigns risk scores to each patient to predict their future care costs. The overall financial performance of health plan providers as well as their reimbursements are tied to such models. To improve both care quality and financial performance, these providers need better ways to identify risk, close any gaps, and improve risk adjustment performance. NLP solutions can read unstructured patient data in medical records to make it easier to assess risks and assign risk scores to patients. These scores can then be used to predict Healthcare costs.

Future NLP Trends in Healthcare

> In the coming years, technological progress in NLP will accelerate even further. These key trends will drive this momentum, and deliver even more value for the Healthcare industry.

Smarter Pre-Trained Language Models

Future NLP models will be trained on vast amounts of medical information. These smart models will improve the accuracy of NLP results, and also perform transfer-learning between different medical sub-domains.

Multi-Modal Algorithms

The use of multi-modal algorithms that process both text and image will also increase. Such new algorithms already capture, read and combine signals from various sources, such as medical imaging devices and sensor data such as wearables, as well as audio and video inputs, free text (i.e. NLP), and “traditional” structured data. As these models become more commonplace in future, they will compensate for the lack of labeled data, and give Healthcare providers a fuller picture of a patient’s situation, so they can deliver better care.

NLP Data in New Workflows

Currently, the incorporation of NLP data into end-to-end AI solutions, including business and clinical workflows, is still in its infancy. But as language models like BERT and its clinical variants progress further, the outputs from NLP and AI models will be significantly more accurate than they are today.

Increasing Adoption of AI Techniques

AI techniques, like neural nets, transfer learning, and active learning will be increasingly leveraged to improve the accuracy and relevancy of insights in the Healthcare sector. NLP and text analytics will also be used to drive greater value in Healthcare organizations, and to improve life expectancy and quality for patients.

Progress in Cognitive Computing and Knowledge Graphs

Using NLP and pattern recognition techniques, as well as complex data mining and sophisticated curation, cognitive computing applications will be able to draw important connections between seemingly unrelated pieces of Healthcare data. These applications will consider user intentions, provide actionable insights, and express results in a format that supports better analysis and decision-making. NLP will also help turn big data into smart data by intelligently analyzing vast data volumes, and delivering intuitive reports and insights that empower providers to focus on delivering quality patient care.

How Open-source Software Helps Advance NLP in Healthcare

Data Avalanche in Healthcare

The U.S. Healthcare industry produces about **1.2 billion**[4] clinical documents every year. That's a massive amount of unstructured text data. These comprehensive datasets can provide a more holistic view of a patient's health, inform drug discovery, support real-world safety assessments, and help with the design of new diagnostic and treatment pathways.

However, such data is generated in huge volumes making it challenging to measure, analyze or consume. NLP effectively addresses these gaps, allowing users to extract, codify and store unstructured clinical text in a structured format.

Roadblocks to NLP in Healthcare

Despite incredible innovations, not many Healthcare organizations fully leverage NLP. One reason is that traditional NLP technology was largely rule-based and grammar-based, and cannot reach the level of accuracy achievable with current state-of-the-art deep learning and transfer learning techniques. Another issue is that legacy data architectures cannot support unstructured data or its analyses – effectively processing the much larger volume of data available today, handling streaming data, and combining multiple data modalities. A lack of advanced analytics capabilities slows down time-to-insight, and has become a business critical priority given the fast progress in this field and its uneven application by different Healthcare organizations.

How NLP Libraries Unlock the Power of Healthcare NLP

Software Libraries like Spark NLP for Healthcare are purpose-built for Healthcare. They include algorithms and models designed to understand Healthcare-specific language, so Healthcare organizations can effectively transform their large volumes of textual data into novel patient insights.

Spark NLP is an open-source text processing library that seamlessly extracts, classifies, and structures text data with state-of-the-art accuracy. It uses scalable and trainable implementations of healthcare-specific deep learning and transfer learning techniques.

It also leverages **4,000+** pre-trained pipelines, models and transformers, and provides out-of- the-box named entity recognition of **100+** clinical and biomedical entities. Spark NLP for Healthcare includes are commercial extensions of the open- source library which include clinical and biomedical algorithms and **250+** additional models – which Healthcare data scientists can leverage.

4,000+
pre-trained pipelines and models

100+
clinical and biomedical entities

250+
additional models

Like Spark NLP for Healthcare, Spark OCR is another software library designed to help Healthcare organizations tap into the power of NLP to improve care and patient outcomes. Spark OCR applies deep learning techniques to accurately extract facts from visual documents, images, tables, and forms to help data scientists better understand their data – and their patients.

Conclusion

Increasing digitization has increased the volume of unstructured textual data generated in the Healthcare industry. NLP makes it possible to easily analyze this data, extract valuable insights from it, and leverage these insights to improve diagnoses, treatments, and decision-making.

With open-source NLP libraries, Healthcare-specific NLP software & language models, and OCR solutions, Healthcare providers have the power and potential of NLP at their fingertips. Further developments in Healthcare NLP will empower them to make more of their data to improve health outcomes for all stakeholders who rely on them.

Resources

1 Lucente, Ida. “The 2020 NLP Industry Survey Finds Increasing Enterprise Investment in Natural Language Processing, Despite Pandemic-Impacted IT Budgets.” John Snow Labs, 22 Sept. 2020, www.johnsnowlabs.com/the-2020-nlp-industry-survey-finds-increasing-enterprise-investment-in-natural-language-processing-despite-pandemic-impacted-it-budgets.

2 Rasmy, Laila, et al. “Med-BERT: Pretrained Contextualized Embeddings on Large-Scale Structured Electronic Health Records for Disease Prediction.” Nature, Nature, 20 May 2021, www.nature.com/articles/s41746-021-00455-y.

3 Adverse Drug Events.” Health.Gov, health.gov/our-work/national-health-initiatives/health-care-quality/adverse-drug-events. Accessed 11 Oct. 2021.

4 Health Story Project.” HIMSS, www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiS7Y6qgcjzAhVpyzgGHb0SBR8QFnoECAQQAQ&url=https%3A%2F%2Fs3.amazonaws.com%2Frdcms-himss%2Ffiles%2Fproduction%2Fpublic%2FFileDownloads%2FHIMSS%2520Health%2520Story%2520Project_FactSheet.pdf&usg=AOvVaw1oKGsV3mcnYNRMa4eC6qdg. Accessed 13 Oct. 2021.