

Chapter 14

The Emergence of Natural Language Processing (NLP) Techniques in Healthcare AI



Rakesh Kumar Attar and Komal

Contents

1	Introduction	285
2	Background and Motivations	287
3	Natural Language Processing in Healthcare	289
4	Role of Deep Learning Techniques Based NLP Systems in Healthcare	298
5	NLP and Covid-19	302
6	Conclusions and Future Directions	304
	References	305

1 Introduction

Natural Language Processing, a specialized branch of AI, aims on the interpreting and manipulating of human generated written or spoken data. NLP mainly focuses on inference of meaningful data by analyzing speech and text data. Natural language processing (NLP) can be termed as the ability for computational machines to better understand the human spoken terms and written text. NLP is extensively used in modern technology to provide assistance for privacy of spam email, personal voice assistants and application of language translation. NLP is a very demanding and appealing research domain in linguistic informatics. NLP has the capability of analysis and extraction of information from distinct unstructured sources, rendering question answer process in automatic manner and summarization of textual data and conducting analysis on sentiments. NLP is often used to include a set of methods

R. Kumar Attar (✉)

Department of Computer Science, University College Miranpur, Patiala, India

Komal

Department of Computer Science, M.M. Modi College, Patiala, India

involving the processing of unstructured texts. However, some methods make use of less semantic knowledge, and are purely based on the occurrence of words in text. The semantic knowledge required in these is the knowledge of what comprises a word; these methods rely on the approach of bag-of-words or a keyword. One instance of a method is a search engine, which make use of only words, extracting all the records containing the occurrence of a mixture of words in a group, although these words may totally differ from each other in the extracted documents. For instance, a method of machine learning that make use of words independently to build a statistical model. Generally, more advanced semantic methods of NLP pursue to ascertain all or some of the linguistic structure in text and to construe meaning of related information in text. Natural language processing is being utilized to portray the methodology of exploiting computer algorithms for identification of main elements in everyday language and extracting meaning from unstructured written or spoken input. NLP needs special skills in AI, computational linguistics, and other disciplines of machine learning [1].

Natural Language Processing, being formulaic, makes easy reason for computational methods to process it. Being formulaic, it contains discrete words, and grammar rules defining how diverse semantic elements can be joined to create or build a sequence of words representing a well-formed phrase or sentence that is capable of conveying a specific meaning. NLP is emerged from AI that is purely devoted for development of algorithms and designing models capable of using natural language of the human beings [2]. It is extensively used in searches and translations made using Google or in virtual personal assistants like “Alexa” and “Siri”. NLP is capable of analyzing and extracting information from various unstructured sources, automating question answering process and conducting analysis on sentiments and summarization of text data [3]. As public healthcare and medical institutions acquire and exchange knowledge with natural language as main source of communication, NLP can play a vital role to unlock the potential of artificial intelligence in biomedical sciences. In modern scenario, techniques of machine learning are extensively used to build most of NLP platforms [4]. The main aim of Natural Language Processing (NLP) is for automation of tasks with help of computing techniques that are capable of representation of the specific information in the textual or written form reliably. Machine learning models consists of four main components: a model; input data; a data fitting function; and a training algorithm for model [5]. Recent developments in these areas have paved the way to improve NLP models largely with use of deep learning techniques, emerging from machine learning [6]. New innovations and developments in the diverse types of models, such as convolutional neural network-based (CNN), recurrent neural network-based (RNN), and attention-based models, has facilitated capturing and modelling of more complex semantic associations and concepts than simple keyword presence using recent NLP systems [7]. The approaches of vector embedding has this effort for preprocessing the data that is capable of encoding the words before passing data into a model. These approaches are capable of recognizing that different meaning of word in different contexts (e.g. the meanings of “shot”, “patient,” “virus”, and “disorder,” may change according to context) and treating them as points in a

conceptual space rather than separated entities. The advent of transfer learning has augmented the performance of these models as once a model is trained its output can be fed to another model as input for training the model in a related task. The performance of NLP models has received a great boost with advancements in hardware and steep rise in freely available databases [8]. Novel evaluating tools and yardsticks such as GLUE, BioASQ and superglue are providing help to a great extent in broadening our understanding information type as well as scope these novel models can capture [9, 10].

This chapter is structured around the concepts of the emergence of Natural Language Processing (NLP) techniques in healthcare AI. Firstly, an in-depth detailed introduction of Natural Language Processing in healthcare are presented in Sect. 1. In the subsequent Sect. 2, the background and motivations behind use of Natural Language Processing is given. Section 3 presents a detailed study on use of NLP in healthcare. NLP techniques and role of deep learning and NLP is presented in Sect. 4. Section 5 is focused on NLP and COVID-19 information. Finally, in Sect. 6 we conclude our findings on emergence of NLP techniques in healthcare and present its future directions.

2 Background and Motivations

Natural Language Processing (NLP) aims for rendering support for the advancement of the basic aims of biomedical and healthcare informatics, which comprises the new discovery and supporting scientifically validation of available knowledge, enhancement in the quality and expenses of healthcare facilities, and full backing to health professionals and patients. In healthcare and biomedical fields, a large scale of data and knowledge are circulated in the literature of scientific nature in text form as articles, as text fields in large distributed databases, and as administrative nature or technical reports or articles on the Internet. In healthcare centers, patient health records mostly generates in the form of handwritten narrative notes and large computerized reports. The pervasive acceptance of electronic health records (EHR) in clinics, hospitals, nursing homes, medical centers and other healthcare facilities is the main source of generation of huge real-world health information, which is very beneficial for carrying out clinical or medical research. From the last many years, electronic health records (EHR) systems are being greatly welcomed by hospitals clinics, medical centers and other healthcare facilities. Analysis of giant volumes of data is the underlying principle for providing patients improved and better healthcare facilities. Nevertheless, manual review process of this huge amount of data procuring from multiple sources is very expensive and sluggish creating many challenges while reviewing this healthcare data in a meaningful or intelligible form.

Because of the increasing acceptance of electronic health records (EHRs), it is a trend nowadays for patients to make available their health records at more than one healthcare facilities, and the involvement of several notes in the health chart of a single patient at one healthcare facility. These days' scientists and medical professionals are finding it very difficult to keep them up to date or vigilant about the

latest discoveries or research studies due to the easy availability of huge volumes of online textual health information. As a result, they are feeling dire need of help for the acquiring, managing, and analyzing the huge amounts of knowledge and data available online. On the Internet, people are keen on finding and exchanging exchange health-related available information, and patients and health professionals are often flooded by the huge volume of the health related information accessible to them procuring from wither health related websites or through online socially health forums or communities. A lot of information is also occurring verbally through scientific or medical interactions at workshops or conferences, in health related lectures at hospitals and medical centers, and in doctor-patient consolations meetings.

In this chapter, we are concentrating on the written or textual form of health data. While there is very vital information communicated in textual form, it is not in a conducive format to further computerized processing. The intrinsic characteristics and diversification in languages renders the job of text processing very challenging. Mostly automated applications prefer the use of systematic regulated data so a notable quantity of manually work is particularly designed to map text or written health information into a structured, coded or systematic representation: in the healthcare domain, for example, professionally trained programmers allocate billing number/codes analogous to diagnoses and rules and regulations to healthcare facilities admissions; and database designers extracting genomic and phenotypical type information on organisms from the available literary works. Moreover, the large availability for tons of text or written information makes the manual work expensive, sluggish and very arduous task to keep vigilant and up to date. Hence, the role of Artificial Intelligence (AI) techniques are emerging as very vital in enhancement of clinical or medical research and healthcare. Natural Language Processing (NLP) mainly aims to automate the tasks using computerized techniques or methods that can represent the pertinent information in the textual form with higher reliability and validity.

As a huge numbers of electronic health records (EHR) are entwined in clinic written notes, the use of Natural Language Processing (NLP), Machine Learning (ML) and Deep Learning (DL) techniques have been exploited for mining or extracting information [11]. As for medical imaging the techniques of computer vision (CV) are preferred but to analyze the unstructured data or information contained in the electronic health records (EHR) NLP techniques are mostly used; and for robotics-assisted operations, techniques of reinforcement learning are used. During text analysis and determination of relationship among phrases grammatically, NLP algorithms are mainly used for identification of clinically phenotypes. To obtain highly sensitivity such as to identify true cases on a large scale and getting prediction value positively very high in healthcare records, NLP techniques using rule based approach can be more useful.

Healthcare being the most preferable domains where computer science is acting as supportive of diverse tasks. Artificial Intelligence is very willingly accepted across the healthcare domains covering from very basic level activities to moderate

ones, and moderates to specialization practices, and a great number of appealing AI applications using Natural language processing (NLP) to maximum extent. These AI techniques being potentially capable to identify the distinctive medical or clinical features among patients that further offers new ventures for clinic care and reducing methodological variations in medical or clinical research concentrating on diverse health diseases. The availability of giant volumes of textual information through published works related to science, clinical or medical care or on the internet can be used to maximum extent to obtain and documenting well knowledge procuring from the information expressed or communicated in textual form, and paving way for promotion through discovery of new occurrences [12, 13]. To cite an example, the information communicated through patient text notes, while not initially prepared for purposes to discover new things, but originally aiming rather for the better care of patients individually, can be analyzed, collected and extracted to find new or similar patterns across different patients.

For health professionals and doctors accessing electronic health records for treatment of a specific patient, Natural Language Processing can be supportive in healthcare domains: when studying and analyzing the patient information sheet, NLP can be influenced to collect and integrate information distributed across various health notes and summary reports, and to accentuate relevant details of the patient. Information contained in health notes extracted through NLP can impart to the decision support systems in the electronic health records at the time of actual health care and decision-making [14]. At the time of making report of patient information by health professionals, NLP oriented methods can contribute in generating qualitative health quality notes and reports. Eventually, Natural Language Processing can be supportive to patients and healthcare professionals seeking informative notes of a specific disease or treatment, with question understanding that can then provide easy availability of related information, meeting their needs, and enhancing their health education levels analyzing information contained in document and the lexis used in the health notes.

3 Natural Language Processing in Healthcare

The acceptance and application of natural language processing in healthcare sectors is increasing on a great extent due to its identified potential for searching, analyzing and interpreting huge amounts of patient health databases [15]. With the use of ever-advanced medical models, machine learning along with NLP technologies in healthcare have the ability to exploit associated concepts and insights from data that was previously taken as buried in textual form [11]. NLP in healthcare media can precisely bring life to the unstructured data of the healthcare world, delivering unbelievable insight into quality of understanding, gaining improvement in methods, and providing better outcomes for patients [16].

Healthcare professionals takes a lot of time entering the possible reasons behind, the happening and what's happening to their clients into handwritten chart notes.

These chart notes are not in a form that data available in these can be easily extracted and analyzed by a computational machines or computers. When the doctor listen your ailments at the time of appointment, and records your appointment in a chart note, these narrations becomes the part of the electronic health record systems (EHRs) and get stored as free form textual data [17]. Although, mammoth volumes of unstructured patient health data is entered daily into EHRs, but it is a very cumbersome task for a computer to assist healthcare professionals gather that crucial data [18]. Structured data like CCDAs/claims/FHIR APIs may give assistance to ascertain disease burden, but provides us a restrained view of the actual patient health record.

According to big data analytics in healthcare demonstrates that around 80% of healthcare records are unstructured, and therefore remain mostly unutilized, since text mining and extracting of information of this health data is arduous and resource consuming. Without the use of NLP, this healthcare data is not in a read to use format for latest computer-oriented models to retrieve. Natural Language Processing in Healthcare utilizes dedicated engines capable of cleansing huge amounts of unstructured healthcare data for discovery of previously ignored or inappropriately recorded patient health conditions [19]. NLP healthcare records with use of machine learning models [20] can help in uncovering diseases that could not have been identified previously, fundamental characteristics to make discovery of HCC disease. Electronic health records and healthcare professionals do not always follow same direction. The additional load of entering data responsibilities brings many challenges, and can lead to frustration and fatigue. Researchers comes with a conclusion, some healthcare professionals endure torture of EHR burnout and willing to get retired from job much earlier rather than enduring torture through the many clicks and computer screens needed for navigation of the electronic health records [18]. NLP in medical sector is steadily becoming as a solution to this problem since tools of NLP healthcare can easily use and correctly extract meaning from clinical records.

The availability of large amounts of learning data directly lead to an increase in the accuracy of NLP models in healthcare. The regular use of NLP models in healthcare helps us getting more accurate and efficient NLP systems as these systems are capable of learning from their experiences and training data. Even some suppliers of NLP systems for healthcare demonstrates the capability and functionalities system can offer in healthcare with a particular medical group fulfilling the needs of that specific group. NLP models in healthcare also provide the advantage of synthesizing and summarizing all the information contained into lengthy chart notes into some pertinent points. In the past, thus process of synthesizing and summarizing all information could vary from several weeks to even years following the manual process for reviewing and processing piles of chart notes from patient health records, just to extract the vital information. NLP systems in healthcare sector can easily look through medical textual data within seconds and find out the relevant that should be retrieved from these records. This help in getting some leisure time for healthcare professionals and focusing more intensively on the complex issues and minimizes the time eaten in frequently occurring administrative tasks [19, 21]. When computers

can become self-sufficient in understanding doctors' notes correctly and does processing tasks on extracted data accordingly, these systems can prove very beneficial in delivering valuable decisions. This can lead to carrying out drug and medicinal research in near future providing greater assistance to health professionals and patients.

All the doctors do not "speak the same way", and should be vigilant enough that their chart notes and narratives will likely be consumed by their colleagues, patients and even computational systems, adhering to data privacy policies of their organizations. The use of standard natural language in creation and maintenance of chart notes and narratives is very pertinent. Many of the NLP systems in healthcare are designed and developed with the aim to serve a wide variety of notation terminologies used in healthcare [22]. However, the use of nonstandard and uncommon notations can lead to confusions among readers of these chart notes and NLP systems. In last 3 years, the developing system for gaining improvement in NLP healthcare data has proven a very arduous task. If the outcomes of NLP systems in healthcare suggests too many results, or artificial results that are not correct, users will have been forced in ignoring the intelligence and move towards a system that is capable of minimizing overall business output. NLP systems in healthcare should focus around least noised and robust signal data outcomes about what healthcare professionals intended to do. Medical NLP provide a great opportunity for computers to center around the things that they meant to do.

3.1 Challenges in NLP

Despite the recent developments, hurdles to extensive use of NLP methods and technologies still exist. Similar to other AI techniques, Natural Language Processing is very much dependent on the easy availability, qualitative and natural orientation of the training datasets [23]. Smooth access and easy availability of appropriately annotated datasets (using supervised or semi-supervised learning in effective manner) are basic for training as well as implementation of powerful NLP models. For instance, the development and use of algorithms that are capable of carry a well-organized synthesis of reported research work on a specific topic or a comparative study and extraction of data from electronic health records (EHRs) needs access without any imposed restrictions to publisher or healthcare/primary care databases [18]. Although in recent years, the number of datasets in biomedicine and healthcare that are easily accessible without any hindrance and pre-trained learning models has been arising, but the number of dealing persons with public healthcare concepts stays very limited [24, 25]. The ability to remove biasness from data (i.e. by offering the ability for inspecting, explaining and adjusting data in ethical manner) indicates another major issue for the training purpose and using NLP models in public healthcare sectors.

The Failure for accounting for biasness present in the developing (e.g. annotating data), deploying (e.g. using pre-trained systems) and evaluating of NLP based

models could lead to compromising of the model outcomes and reinforcing existing health prejudice [26]. However, it is worth mentioning that even when annotated datasets and its evaluations are a modified for biases, this does not provide guarantee for an equivalent impact across morally applicable stratum. For instance, using healthcare data provided through social platforms must be considered the particular age group and socio-economic groups using this health data [27]. A monitoring system trained on data obtained from social media application Facebook is no doubt to have biasness towards healthcare data and semantic quirks specifically related to a set of population older than system trained on data obtained from another social application Snapchat [28]. In the recent times, many model unbiased tools have been developed to identify and bring unfairness in machine as well deep learning models according to the efforts imitated by the government agencies and academic organizations for defining unacceptable artificial intelligence development [29, 30]. At present, one of the main hindrance for further development of NLP platforms is limited access of data in public healthcare sectors [31, 32].

In a recent study at Canada, data related to public health are generally maintained and controlled on the regional basis and, due to confidentiality and security issues, there is hesitation for providing freely access to these NLP systems and their incorporation with other available datasets (e.g. data linkage) [23]. There are also some other challenges with public perception of privacy and freely access of data. A recent survey, conducted on users of social media platforms, demonstrated that the major chunks of the users have taken into account comparative analysis of data on their social platforms account for identification of health issues related to mental condition “intrusive and exposing” and they would not be willing to give their consent to this [33]. Prior to main NLP public health related activities could be ascertained at large scale, such as the comparative analysis in real-time of trends in disease nationwide, judiciary will require to jointly ascertain a reasonable scope and effortless access to public healthcare data sources (e.g. data related to health and administration). In order to prevent and eliminate privacy violations and misusing of data, NLP applications, in near future, to analyze personal healthcare data are contingent on the capability for embedment of distinguished privacy into learning models [34], both during at stages of training and after the deployment of model. Access to vital data is also restrained through the methods used currently to freely access full textual oriented publications. Unhindered and freely access to databases containing journal publications or novel models for storage of data is the basic need for realizing fully automatic PICO-oriented knowledge extraction and synthesis [35].

Eventually, meeting the requirements of new technologies, more intention must be given for identification and evaluation of Natural Language Processing based models for ensuring their working with intention and pace according to the society’s ever changing views about ethics. These NLP models and technologies are required to be analyzed for ensuring their functionality according to their expectation and accounting for biasness [25]. Although in the present time improved or equivalent to human scores on text analysis tasks are posted by many approaches, the equating high scores with true natural language understanding is not of too much importance.

However, it is also equally of very much significance not to visualize a lack of true natural language understanding as a lack of utilization. NLP Models with a “relatively poor” in-depth understanding can still be of very much effective at the time of extracting information, classifying and predicting tasks, specifically with the enhanced availability of data of labelled type [11].

3.2 Opportunities for NLP in Healthcare

Public healthcare organizations work with the motive to gain optimum health results within and across diverse sets of populations, mainly by focusing on the development and implementation of interventions that are meant for alterable causes of poor health condition [8, 15]. The success solely relies on the capability of quantifying the impact of disease in effective manner or its risk factors within the population and also identifying clusters that are influenced in disproportionate manner or at the verge of risk; detect best practices (i.e. optimum prevention or therapeutic procedures); and measure results [23]. There is a decision making model based on evidence-informed that uses “PICO concept (patient/problem, intervention/exposure, comparison, outcome)” [36]. This PICO concept based model offers a strategy to identify optimum knowledge for framing and answering particular clinical or public health queries [37]. The decision-making based on evidence-informed is typically well grounded on the thoroughly and well-organized comprehensive review and synthetization of data according to the PICO concept elements.

In present scenario, information is being generated and published (e.g. technical reports, science literary text, medical information, surveys, socially media generated data, and other well documented data) at incredible rates [38]. NLP is capable of rapidly analyzing huge amounts of unstructured data or semi structured data, thus paving the way to open up tremendous opportunities for proof-informed decision-making process and text-based research [17, 37]. NLP has emerged as a potentially robust tool to support the rapidly identifying populations, interventions and results of interest that required for surveillance and prevention of disease and promoting health. For instance, the use of NLP systems that can identify specific characteristics of individuals in unstructured medical documents or social media content can be utilized for enhancement of existing surveillance structures with real-world testament [39].

One latest study revealed the capability of NLP methods to assess the existence of depression much before it appeared in health related records. The ability of NLP to carry out real-time text mining of scientific research publications or reported articles for a specific “PICO (Patient/Problem, Intervention, Comparison, and Outcome)” concept [36] creates enough opportunities for decision makings to rapidly offer recommendations and provide suggestions on the prevention of disease or management that are made aware of latest body of evidence when timely recommendations and suggestions are sought, such as during an outbreak of disease or epidemic. NLP enabled Chatbot’s and question-answering systems [36] also possesses the much

required potential to gain improvement in health promoting activities by creating environments for individuals to get engaged in these health activities and offering personalized assistance or advice.

3.3 *NLP Applications in Healthcare*

Natural language processing (NLP) has a very vast range of possible applications. The following are some very important applications of Natural language processing technology for healthcare:

1. **Information Extraction:** It is the process of locating and structuring particular information in text, most widely used application of NLP in field of biomedicine, and usually carried out without having semantic analysis of the text, but focusing and finding on the patterns present in the text. Once task of extraction and structuring of text information is completed, then textual information can be utilized for various diverse tasks [11]. In the field of bio surveillance, for example, symptoms from a main complaint field in a handwritten note can be extracted when a patient was directed to get admitted in the emergency wards of a healthcare facility [40] or from handwritten notes of electronic health records meant for ambulance services [41]. The gathered data from a number of patients after extraction can provide help in understanding the prevailing and progressive stages of particular epidemic or pandemic. In the field of biology, bio molecular conservations as well as interactions between patients and health professionals on written notes can be extracted (if taken from one single note) or merged (if taken from multiple articles) for construction of pathways of bio molecular type. In the clinical sector, NLP can be employed on large number of health records of patients for obtaining structure data to be used in pharmacovigilance systems for discovery of adverse drug happenings [11]. In named-entity recognition process, the information extraction techniques may be used to the identify names of persons or places, numerical expressions, and dates, or to several types of terms present in text (e.g. mentioning of proteins or medications) and then can be transformed into standardized or canonical forms. This process of transformation is termed as named-entity normalization. More well organized and reliable techniques are used for identification and representation of the modifiers related to a named entity. Such advanced techniques are needed for reliable extraction of information because the accurate meaning of a term typically associated and may vary in association with other terms in a provided sentence. For example, the term pain has different meanings in no pain, high pain, pain lasted 3 days, and low pain. Another important method of information extraction is to identify relations among named entities. For instance, when locating and finding adverse events in association with a medication, the phrases “the patient developed a rash from ciprofloxacin” and “the patient came in with a rash and was given benadryl” must be differentiated. In both sentences, a relation between a

rash and a drug is showing, but the first sentence indicates a possible adverse drug happening whereas the second sentence indicates a treatment for an adverse happening. As extraction of entities takes place from multiple sources, need for one very crucial step of reference resolution arises. Reference resolution is the process to recognize two mentioning in two different source locations referring the same entity [19]. In some cases, Reference resolution is a very arduous task. For instance, mentions of fracture in two different articles related to same patient could indicate the same fracture or two very distinguished fractures; some more contextual information and additional domain knowledge is often required to resolve this complex problem.

2. **Information Retrieval:** Information Retrieval (IR) and Natural Language Processing coincide in many of the methods in use. The methods of Information Retrieval (IR) are aimed for providing support to user for gaining access of documents in huge databases, such as the scientific research literature, electronic health records (EHRs), or the WWW. This is a very important application in health and biomedicine, due to the outburst of vast information present in electronic form. The fundamental aim of information retrieval is to check a user's question against a collection of documents and retrieve an organized list of related documents [11]. A search operation is conducted on an index of the large collection of documents. The most fundamental form of indexing separates simple terms and words, and therefore, utilizes very less semantic knowledge. Many of modern approaches makes use of Natural Language Processing (NLP) oriented methods of same kind which are used in information extraction, finding entities of complex nature and ascertaining their associations in order to increase the accuracy level of information retrieval. For example, one could look out for Cough and have the search being operated at the conceptual level, retrieving documents mentioning the phrase Asthma in addition to the phrases that mentions cough only. Furthermore, one could look out for cough in a particular context, such as study of the disease or treatment.
3. **Question Answering:** Question Answering (QA) is termed as a process in which a user submits a question expressed in natural language and answer is then automatically presented by a QA system. The easy availability of required information in published articles or works and on the Internet empowers these type of systems increasingly very significant as healthcare professionals as well as consumers, and researchers in biomedical field frequently explore the Web to procure information about a disease/medical condition, its cure, or a clinical procedure. A question answering system can be very beneficial for procuring the answers to factual questions, like "In children with an acute febrile illness, what is the efficacy of single-medication therapy with acetaminophen or ibuprofen in reducing fever?" The answer to above query produced by QA system is like "Ibuprofen provided greater temperature decrement and longer duration of antipyresis than acetaminophen when the two drugs were administered in approximately equal doses" [36]. QA systems enriches an Information Retrieval

(IR) system with additional **capabilities**. In an IR system, a query is generated after user translates a question into a list of keywords, but a Question Answering (QA) system performs this step automatically. In the addition, user obtains an actual answer (often more than one passages mined from the original documents) presented by a QA system, in the place of a list of related original documents. A QA has aimed to focus on the scientific literature thus far [36].

4. **Text Summarization:** It is the process of producing a single output file containing rational text synthesizing the basic points from one or multiple documents taking as input. Using text summarization users are able to derive sense of a mammoth amount of data, by finding and reporting the essential points in retrieved texts in automatic manner. Text summarization can be either question-focused or general purpose (i.e. focusing on a specific information need into consideration when choosing significant content of inputted documents). Question-focused text summarization can be regarded as a post-processing stage of information revival and question answering: the relevant paragraphs or documents related to an input query are further refined into a standalone, rational text. Text summarization process consists of several steps such as selection of relevant content (recognizing essential points of information in the inputted documents), organization of content (recognizing duplicity and contradictory issues present among the selected passages of information, and placing them into order so the final summarization is coherent), and regeneration of content (presenting natural language from the organized passages of information). Following the similar way of question answering, text summarization has aimed to focus on the scientific literature [11, 19].
5. **Text Generation:** It is the process of formulating sentences of natural language from a provided source of information that human are not able to read directly. Generation can be employed for creation a text from a well-organized database, such as summarizing patterns and in data procured from laboratories [11].
6. **Machine Translation:** It is the process of transforming text in one source language (e.g. Hindi) into another target language (e.g. English). These applications could prove very beneficial in multilingual environments in which translation by humans is either very time consuming or too expensive.
7. **Text Readability Identification and Text Simplification:** Text Readability Identification and Text Simplification is proving very beneficial to the healthcare sectors, as health professional as well as consumers and patients search and explore more and more health related information on the WWW [19]. However, they lacks in their literacy levels according to the health documents available on the Web and seeks support for easy readability.
8. **Emotion Detection and Sentiment Analysis:** Emotion Detection and Sentiment Analysis are latest applications of Natural Language Processing and responsible for making the process of content analysis automatic. There are encouraging research outcomes revealing that patients' discourse can be studied analytically in automatic manner for identification of their mental states [30].

Table 14.1 enlists some examples of possible applications of NLP in public healthcare that have showed at least some success.

Table 14.1 Some possible Applications of NLP

Activity	Objective of Public Healthcare	Examples based on using NLP
Identifying populations at risk	To measure the occurrence and pervasiveness of diseases and risk factors related to diseases continuously (i.e. monitoring)	A comparative analysis of semi structured or unstructured textual data from social media or EHRs [33]
	To detect and find highly vulnerable and at-risk individuals in population	A comparative analysis of risk behaviors using social platforms [42]
Identifying of health related interventions	Developing the optimum suggestions/interventions	Automatic systematic literature review and comparative study of the information conveyed through scientific research published as well as unpublished articles [33]
	Identifying best practices	Recognition of potential public healthcare interventions through comparative analysis of online available peer reviewed and grey literature
Recognition of healthcare outputs with use of real-world evidence	To assess the gains of healthcare interventions	A comparative analysis of semi structured or unstructured textual data from online social media, EHRs and publications to ascertain the impact of public healthcare suggestions and interventions [23]
	To recognize unforeseen adverse outputs related to health interventions	A comparative analysis of semi structured or unstructured textual data from online social media, EHRs and publications to assess possible adverse happenings of health interventions
Generation and Translation of Knowledge	To provide assistance to public healthcare research	A comparative analysis of semi structured or unstructured textual data from EHRs and scientific research published articles for generation of knowledge
	To provide assistance to decision making process using evidence information	Use of questioning answering systems, chatbots and text summarization systems by providing personalized information to people needing advice for improvement of their health and prevention of disease [23]
Scanning of Environment and awareness of situations	To carry risk assessments related to public healthcare and providing situational awareness	A comparative analysis of online available content for detecting and mitigating critical events in real time [26]
	To observe activities that can affect decision making in public healthcare	A comparative analysis of decisions of national and international stakeholders [32]

4 Role of Deep Learning Techniques Based NLP Systems in Healthcare

Most of the NLP based systems are developed with separate or dedicated components that take care of different functionalities offered by these systems. The dedicated components typically roughly occur simultaneously with the semantic levels. Generally, the outcomes of each lower level is fed to the next higher level as input [43]. To exemplify, the outcomes of tokenization process converts a text string into separated tokens that will have to bear lexical analysis for determination of their parts of speech (POS) and other linguistic properties as inscribed in a lexicon; the POS tags along with the equivalent semantic definitions will then serves as the input to syntactic analysis phase that will ascertain the structure of the sentence; the structure of sentence will be then serves as input to semantic analysis phase for the interpretation of the meaning [7]. Each NLP systems bundles these stages of processing differently. At each processing stage, the component meant for that stage works with the aim for regularizing the data in some aspect to minimize variety at the time of perseverance the informative data as much as possible.

There are two approaches to Natural Language Processing:

- A rule-based approach in which systems have to follow predefined rules in the algorithms.
- Machine Learning approach in which learning method based on supervised and unsupervised are used for training of the systems. In supervised Learning, the machines are able to learn according to predefined rules under the guidance of human while in the case of unsupervised learning machine learns without any human guidance or interaction [44, 45].

Firstly, the information contained in electronic health records is extracted by NLP algorithms, which is then processed for classification of patients into a sub categories according to the predefined rules and learners. The Nature of NLP procedures NLP is very complicated due to combination of various techniques together. NLP systems in healthcare will associates words or phrases to concepts of interest, and it requires very cautious approach to pre-process the extracted text and it needs to be transformed into document from the human form [46].

Following are few instances of low-level Natural Language Processing tasks (pre-processing of text):

- Detection of sentence boundary (usually a period indicates sentence boundary)
- Tokenization (partitioning a sentence into separated tokens)
- Stemming (deriving the root form of a word)
- Lemmatization (applying Lemma rules on tokens)

Following are few instances of high-level Natural Language Processing tasks:

- Recognition of Named entities
- Formulating rules for negation

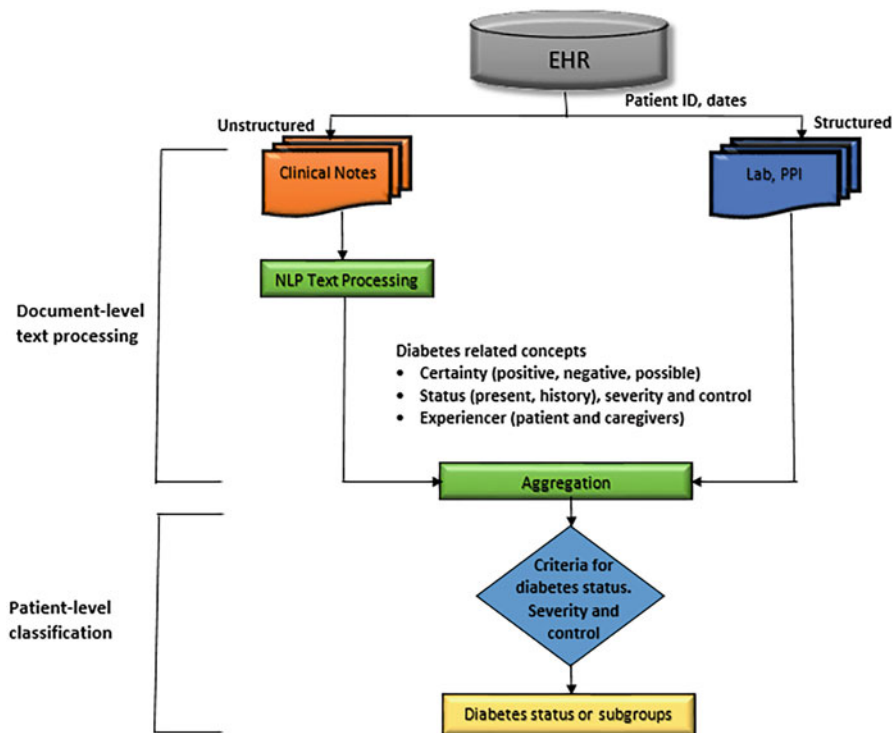


Fig. 14.1 Document-level processing of text and patient-level classification using NLP algorithms

Figure 14.1 demonstrates the preprocessing and classification of text of a diabetes patient.

The healthcare information can originate from four sources in terms of users such as Physicians, Patients, Staff members of paramedical team and Pharmaceuticals. Proper diagnosis of a disease directly relates to post-process stage of a disease. Patients can receive proper and timely treatment if the disease is detected timely and properly. Sometimes, delays in decision-making can lead to worse condition of patients. Deep learning is an approach consisting of a big network that is able to take a wide variation of inputs data types such as image, text, time-series data, audio etc. learns related features in its low-level networks corresponding to each data type. The data from each tower is combined to make it pass through higher levels, providing the provision to the Deep Neural Network in reaching to result based on evidence and reasoning across data types [11].

NLP can provide assistance to healthcare in extraction of information, transformation of unstructured to structured data, categorization of documents, and summarization of text. Eventually it will help in reducing administrative costs by means of efficient billing and correct prior authorized permission. It will also help in adding medical value by providing help for unproductive medical decisions and structured and up to date medical policy evaluation etc. Additionally, it will also provide

Table 14.2 Stages of Healthcare Information

Situation	Process	Tools/Devices/Actions/Reports
Pre-Disease	Health Awareness	Smart sensor devices to provide health updates
	Regular health check-ups	Devices for health check-ups
	How to stay away from disease?	Training programs and devices for various exercises
In Disease	Diagnosis	Blood sample, X-Ray Image, Urine Sample, Biopsy, Devices for checking vital signs in the body (Oxygen, sugar, heartbeat, pulses, etc.)
	Treatment	Pre-surgery treatments, Minor surgery, Major surgery, Combined surgery, post-surgery treatments
	Medication	Drug dosage based on the nature of the body for a single disease, multiple diseases
Post Disease	Follow-ups	Prediction of the number of follow-ups or time period of follow-ups required
	Improvement comparison	Compare Pre, In and Post disease reports
	Future predictions	

support for gaining improvement in patient interactions with health professionals and the electronic health records, increasing health awareness in patients, improving qualitative care, and identifying patients who need critical health care. In healthcare, technologies of NLP and sequential DL has provided a boost to the health applications lies within domains like EHRs [43]. The various stages of Healthcare information are given in Table 14.2.

Figure 14.2 demonstrates various stages involved in processing of data.

- A. **Unstructured EHR data:** Heterogeneity is often found in health records according to the storage policies, data structures and working mechanism of a particular healthcare facility. Therefore, it will vary from one healthcare facility to another.
- B. **Data Standardization:** Data will be transformed into standard form into the similar format to associate data from various sources based on FHIR.
- C. **Sequencing:** To sequence data into a patient timeline temporally, time-oriented DL techniques can be utilized to the EHR datasets entirely for delivering predictions about individual patients. NLP systems can prove very beneficial in the neurology domain too. The NLP system, “Edinburgh Information Extraction for Radiology reports (EdIE-R)”, is a multiple staged pipelined process as demonstrated in Fig. 14.3, with XML rule-based text mining program at its internal level [47].

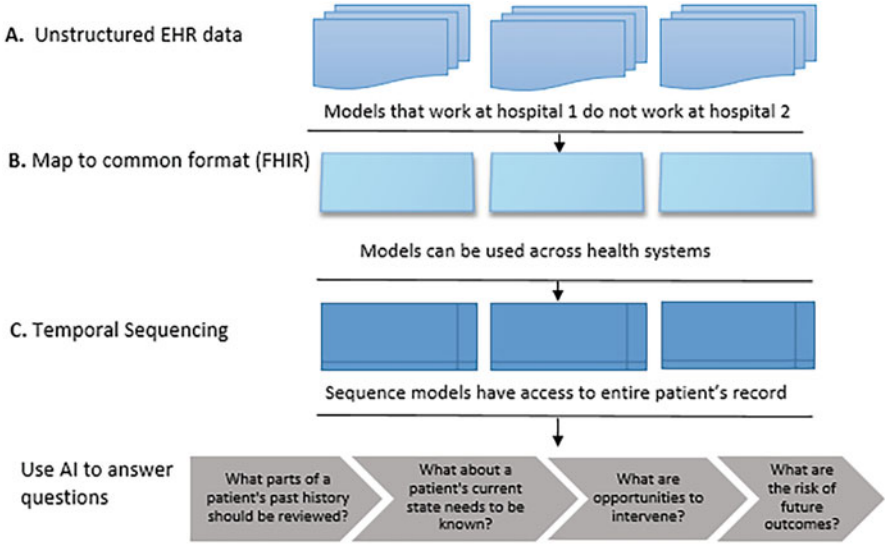


Fig. 14.2 Predicting using EHR

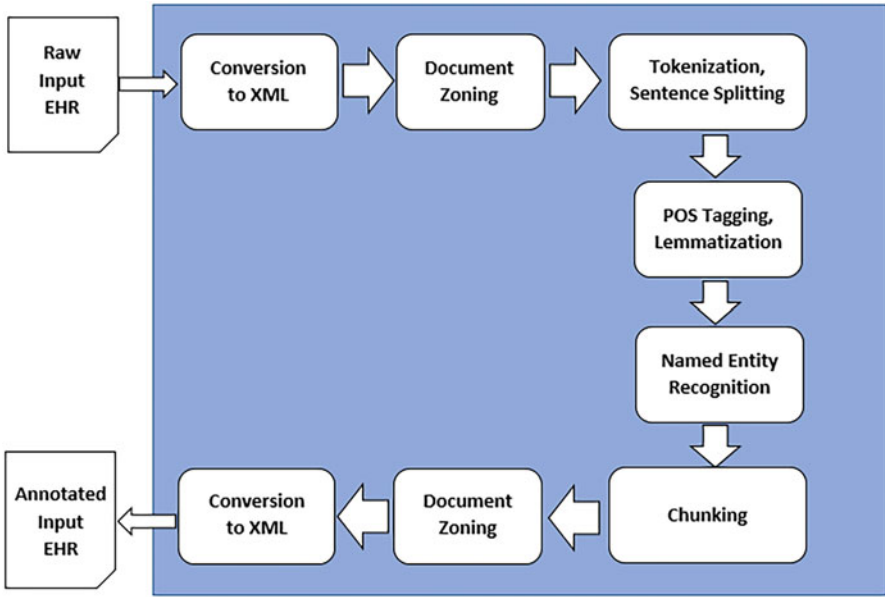


Fig. 14.3 Architecture of EdIE-R System

5 NLP and Covid-19

In the coming decades, people will remember year 2020 taking about novel coronavirus and its impact on so many global factors such as individual as well as global health, climate, economy, travelling, and population.

5.1 Role of NLP in Understanding the Landscape of COVID-19 Information

In this adverse time, as dark clouds looms, there is also a hope of coming from the collective efforts of healthcare organizations, institutions, and government agencies to search the best responses to this deadly challenge, specifically in the areas of repurposing of drugs and development of vaccines.

In these kinds of situation, NLP-based text mining come into scene playing a major role. Whenever researchers, scientists, healthcare professionals are met with such critical challenges, one crucial asset is to identify or stockpile as much informational data about the given problem as possible. Whatever kind of information available is, at regional and world level, that can be traced, collected and comprehended, will pave the way to take the right decisions. Relevant or related information on the biological nature of SARS-CoV2, from COVID-19 as a deadly disease, demographics of patients and co-morbidities, spreading locally or globally, and possible and viable drugs that might come handy in treating the symptoms and effects of this deadly disease is all just small part of huge data iceberg [48]. Most of this data has its existence or presence in the form of unstructured text such as scientific research papers, clinical health trial records, preprints, adverse event documents, EHRs, even news reports and social media content can all deliver information on related factors to epidemiologic, for instance. This is often where artificial technologies such as Natural Language Processing can play a vital role. NLP-based text mining employs a collection of methodologies involving machine learning, linguistic processing, ontologies, regular expression and more to covert the unstructured or free form text in reports and databases into organized, normalized data befitting to visualize or analyze. NLP empowers researchers to easily extract and use information from various sources such as clinical health trial records, scientific literature, preprints, insider sources, news and social media. The collection of vital information from various sources and incorporating into one place provides information consumers an in-depth understanding of almost everything that is happening all around. This approach can help in providing real time answers to main key questions to face the COVID-19 pandemic, such as:

- What are the best possible drugs for repurposing efforts?
- Where do I find the latest research, recent health trials, and the key health professionals and researchers in this field?

- Who are at more risk for deadly disease in the population?
- What are the main co-morbidities involved?
- What kind of additional health care activities are needed for patients post this deadly disease?

5.2 Use of NLP for COVID-19 Understanding in Pharma and Healthcare Organizations

An innovative approach, for example, for taking better care of patients arises from a large United States Healthcare System. Healthcare professionals in the system were perturbed about less reporting of COVID-19 cases, so they employed NLP techniques to extract incoming emails and chatting messages from their helpline for patient, for tracing symptoms related to COVID-19. They then performed analytical study for classification of these patients according to the possibility of possessing COVID-19. This automatic process helped in enabling healthcare professionals to manage the patient population in more effective manner. One more instance can be taken of a pharma company who ought to extract social media content to comprehend spreading and risk factors involved COVID-19. Again, a set of questionnaire were prepared for categorization of patient behaviors according to a variety of factors, e.g., mentioning of medical facilities being attended. They also performed a comparative analysis on the social media messages to comprehend COVID-19 status and risk across a spread of occupations and professions.

5.3 Resources for NLP in Healthcare and COVID-19

One of the ways in which NLP field can achieve more progress is sharing of different datasets, developed tools and resources used by various teams and groups of researchers. Shared databases provide assistance to various different teams of researchers in testing and comparing their systems analytically on the same datasets. Shared databases with appropriate annotations are very crucial, as they provide the provision for training of their NLP systems as well. As such, these NLP systems are very beneficial to the community. In last few years, there has been a relatively increase in the clinical and biological NLP communities to develop publicly available shared resources and tools, and in smooth conduction of challenges posed by community. Here we are presenting a few of publicly available datasets on COVID-19 data, but ever-evolving resources in the field, this list could prove to be obsolete. Therefore, we suggest the readers to explore the literature on COVID-19 and the WWW for the latest updates. Healthcare organizations are also utilizing NLP techniques to access the landscape of scientific research papers related to COVID-19 pandemic. Researchers working on developing COVID-19 cures employs NLP

techniques to trace new and latest papers, specifically around safety of vaccine and drug. Many publishers have started offering free access to crucial scientific literature; for instance, “the COVID-19 Dataset compiled by Allen Institute for AI [48]; Coronavirus Dataset of Elsevier [49]; and Copyright Clearance Center COVID-19 resources [50]”. These excellent resources can be harvested to identify efficacy of drug or safety related profiles, understanding co-morbidity profiles, the natural history of the severe virus and disease, and who are at more risk in the population for deadly disease. NLP can provide assistance in some way with enabling easy access to data, and hopefully therefore to dispel some of that uncertainty.

6 Conclusions and Future Directions

NLP in healthcare is giving birth to new and promising opportunities for delivery of healthcare facilities and patient experience. It will not be far away when these specialized NLP systems in healthcare will grant the opportunity to healthcare professionals to provide more quality care time with their patients, while providing assistance in deriving insightful outcomes based on precise data. In the decades to come, we will feel fortunate hearing the news, and witnessing the potential and functionalities of NLP technology, as it enables healthcare providers to contribute towards health conclusions positively. Technologies of NLP and Deep Learning will come into picture playing a significant role to accelerate the decision-making process in the healthcare sector. However, the actual rewards of designing efficient algorithms will rely largely on the data quality that they obtain and preserve. The rapid process of decision-making will empower doctors and healthcare workers in focusing on the additional care of patients. Natural language processing with computer vision and deep learning can provide the greater help in processing a wide variety of data altogether to derive precise and accurate decisions. Collaborative research can provide the opportunity in achieving a higher level of medication and health treatment in healthcare sector. Keeping in mind, the impact of artificial technologies, systems using NLP require to be devised and developed using a cautionary approach in a larger socio-ecological view of healthcare facilities for provision of availing better care services in healthcare. Eventually, NLP tools may gain success in bridging the gap between the immeasurable amount of data accumulated on a daily basis and the limited coherent capability of the human mind. From the most arduous tasks to simpler tasks in healthcare, natural language processing has nearly never-ending possibilities to transform EHRs from burden to blessing. The success of NLP in healthcare will depend heavily on developing algorithms that are precise, intelligent, efficient, and healthcare-oriented and to design the user interfaces displaying healthcare decision support data effortlessly. If the healthcare organizations and industries succeed in meeting these dual goals of extracting and presenting information effectively, there will be no doubt about unbounded opportunities NLP in healthcare can bring in the future.

References

1. Sarosh P, Parah SA, Bhat GM, Muhammad K (2021) A security management framework for big data in smart healthcare. *Big Data Res* 25:100225
2. Ruder S. NLP-progress. Sebastian Ruder, London, UK. <https://nlpprogress.com/>. Accessed 18 Feb 2021
3. Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 25(1):44–56. <https://doi.org/10.1038/s41591-018-0300-7>
4. Jurafsky D, Martin JH (2019) *Speech and language processing*. Stanford University, Stanford, CA. <https://web.stanford.edu/~jurafsky/slp3/>. Accessed 18 Feb 2021
5. Nadkarni PM, Ohno-Machado L, Chapman WW (2011) Natural language processing: an introduction. *J Am Med Inform Assoc* 18(5):544–551. <https://doi.org/10.1136/amiajnl-2011-000464>
6. Gull S, Parah SA, Muhammad K (2020) Reversible data hiding exploiting Huffman encoding with dual images for IoMT based healthcare. *Comput Commun* 163:134–149
7. Tang B, Pan Z, Yin K, Khateeb A (2019) Recent advances of deep learning in bioinformatics and computational biology. *Front Genet* 10:214
8. Arcaya MC, Arcaya AL, Subramanian SV (2015) Inequalities in health: definitions, concepts, and theories. *Glob Health Action* 8:27106
9. Hirschberg J, Manning CD (2015) Advances in natural language processing. *Science* 349(6245):261–266
10. Kamili A, Fatima I, Hassan M, Parah SA, Vijaya Kumar V, Ambati LS (2020) Embedding information reversibly in medical images for e-health. *J Intell Fuzzy Syst* 39:8389–8398
11. Ghayvat H, Pandya SN, Bhattacharya P, Zuhair M, Rashid M, Hakak S, Dev K (2021) CP-BDHCA: blockchain-based confidentiality-privacy preserving Big Data scheme for healthcare clouds and applications. *IEEE J Biomed Health Inform.* <https://doi.org/10.1109/JBHI.2021.3097237>
12. Kumar I, Alshamrani SS, Kumar A, Rawat J, Singh KU, Rashid M, AlGhamdi AS (2021) Deep learning approach for analysis and characterization of COVID-19. *Comput Mater Continua* 70: 451–468
13. Rashid M, Singh H, Goyal V, Parah S, Wani A (2020) Big data based hybrid machine learning model for improving performance of medical Internet of Things data in Healthcare systems. *Healthc Paradig Internet Things Ecosyst.* <https://doi.org/10.1016/B978-0-12-819664-9.00003-X>
14. Treasury Board Secretariat. Directive on automated decision-making. Government of Canada, Ottawa, ON. <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592>. Modified 5 Feb 2019; Accessed 27 Feb 2021
15. Jackson B, Huston P (2016) Advancing health equity to improve health: the time is now. *Health Promot Chronic Dis Prev Can* 36(2):17–20
16. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H, Wang Y (2017) Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2(4):230–243. <https://doi.org/10.1136/svn-2017-000101>
17. Koleck TA, Dreisbach C, Bourne PE, Bakken S (2019) Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *J Am Med Inform Assoc* 26(4):364–379
18. Carchiolo V, Longheu A, Reitano G, Zagarella L (2019) Medical prescription classification: a NLP-based approach. In: *Proceedings of the 2019 federated conference on computer science and information systems, FedCSIS 2019*, vol 18, pp 605–609. <https://doi.org/10.15439/2019F197>
19. Friedman C, Rindflesch TC, Corn M (2013) Natural language processing: state of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine. *J Biomed Inform* 46(5):765–773

20. Rashid M, Singh H, Goyal V (2020) The use of machine learning and deep learning algorithms in functional magnetic resonance imaging—a systematic review. <https://doi.org/10.1111/exsy.12644>
21. Harris JK, Mansour R, Choucair B, Olson J, Nissen C, Bhatt J (2014) Centers for Disease Control and Prevention. Health department use of social media to identify foodborne illness - Chicago, Illinois, 2013–2014. *MMWR Morb Mortal Wkly Rep* 63(32):681–685
22. Rashid M, Yousuf MM, Ram B, Goyal V (2019) Novel Big Data approach for drug prediction in health care systems. In: 2019 International conference on automation, computational and technology management (ICACTM), London, UK, pp 325–329. <https://doi.org/10.1109/ICACTM.2019.8776823>
23. Bačić O, Tunis M, Young K, Doan C, Swerdfeger H (2020) Challenges and opportunities for public health made possible by advances in natural language processing. *Can Commun Dis Rep* 46(6):161–168. <https://doi.org/10.14745/ccdr.v46i06a02>
24. Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366(6464):447–453
25. Penning de Vries BB, van Smeden M, Rosendaal FR, Groenwold RH (2020) Title, abstract, and keyword searching resulted in poor recovery of articles in systematic reviews of epidemiologic practice. *J Clin Epidemiol* 121:55–61
26. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH (2018) Ensuring fairness in machine learning to advance health equity. *Ann Intern Med* 169(12):866–872
27. Jha N, Prashar D, Rashid M, Shafiq M, Khan R, Pruncu CI, Saravana Kumar M (2021) Deep learning approach for discovery of in silico drugs for combating COVID-19. *J Healthc Eng* 2021:6668985
28. Spasic I, Nenadic G (2020) Clinical text data in machine learning: systematic review. *JMIR Med Inform* 8(3):e17984
29. Holstein K, Vaughan JW, Daumé H, Dudík M, Wallach H (2019) Improving fairness in machine learning systems: what do industry practitioners need? In: CHI '19: Proceedings of the 2019 CHI conference on human factors in computing systems, vol 600, pp 1–16
30. Wiens J, Price WN, Sjoding MW (2020) Diagnosing bias in data-driven algorithms for healthcare. *Nat Med* 26(1):25–26
31. Montreal Declaration for a Responsible Development of Artificial Intelligence. Forum on the Socially Responsible Development of AI. 2–3 Nov 2017. Montréal, QC. Accessed 18 Feb 2021
32. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V (2019) Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med Inform* 7(2):e12239
33. Ford E, Curl Lewis K, Wongkoblap A, Curcin V (2019) Public opinions on using social media content to identify users with depression and target mental health care advertising: mixed methods survey. *JMIR Ment Health* 6(11):e12942
34. Zhou M, Duan N, Liu S, Shum HY (2020) Progress in neural NLP: modeling, learning, and reasoning. *Engineering* 6(3):275–290. <https://doi.org/10.1016/j.eng.2019.12.014>
35. Radebaugh C, Erlingsson U (2021) Introducing tensorflow privacy: learning with differential privacy for training data. <https://medium.com/tensorflow/introducing-tensorflowprivacy-learning-with-differential-privacy-for-trainingdata-b143c5e801b6>. Accessed 27 Feb 2021
36. Demner-Fushman D, Lin J (2007) Answering clinical questions with knowledge-based and statistical techniques. *Comput Linguist* 33(1):63–103. <https://doi.org/10.1162/coli.2007.33.1.63>
37. Kim ES, James P, Zevon ES, Trudel-Fitzgerald C, Kubzansky LD, Grodstein F (2020) Social media as an emerging data resource for epidemiologic research: characteristics of social media users and non-users in the Nurses' Health Study II. *Am J Epidemiol* 189(2):156–161
38. Chen IY, Joshi S, Ghassemi M (2020) Treating health disparities with artificial intelligence. *Nat Med* 26(1):16–17

39. Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, Forshee R, Walderhaug M, Botsis T (2017) Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. *J Biomed Inform* 73:14–29
40. Chapman D, Whitfield C, Felitti V, Dube S, Edwards V, Anda R (2004) Adverse childhood experiences and the risk of depressive disorders in adulthood. *J Affect Disord* 82:217–225. <https://doi.org/10.1016/j.jad.2003.12.013>
41. Hripcsak G, Elhadad N, Chen Y, Zhou L, Morrison F (2009) Using empiric semantic correlation to interpret temporal assertions in clinical texts. *J Am Med Informatics Assoc* 16:220–227. <https://doi.org/10.1197/jamia.M3007>
42. Conway M, Hu M, Chapman WW (2019) Recent advances in using natural language processing to address public health research questions using social media and consumer generated data. *Yearb Med Inform* 28(1):208–217
43. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J (2019) A guide to deep learning in healthcare. *Nat Med* 25(1):24–29. <https://doi.org/10.1038/s41591-018-0316-z>
44. Xu C, Doshi T (2021) Fairness indicators: scalable infrastructure for fair ML system. Google, Mountain View, CA. Accessed 27 Feb 2021
45. Yin Z, Sulieman LM, Malin BA (2019) A systematic literature review of machine learning in online personal health data. *J Am Med Inform Assoc* 26(6):561–576
46. Juhn Y, Liu H (2020) Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. *J Allergy Clin Immunol* 145(2):463–469. <https://doi.org/10.1016/j.jaci.2019.12.897>
47. Wheater E, Mair G, Sudlow C, Alex B, Grover C, Whiteley W (2019) A validated natural language processing algorithm for brain imaging phenotypes from radiology reports in UK electronic health records. *BMC Med Inform Decis Mak* 19(1):1–11. <https://doi.org/10.1186/s12911-019-0908-7>
48. The CORON-19 Dataset compiled by Allen Institute for AI. <https://www.semanticscholar.org/cord19>. Accessed 18 Feb 2021
49. Coronavirus Dataset of Elsevier. <https://www.elsevier.com/clinical-solutions/coronavirus-research-hub/data-scientist>. Accessed 18 Feb 2021
50. Copyright Clearance Center COVID-19 resources. <https://www.copyright.com/covid19-resources/>. Accessed 18 Feb 2021