Natural Language Processing (NLP) Based Innovations for Smart Healthcare Applications in Healthcare 4.0

Nemika Tyagi and Bharat Bhushan

April 2, 2024

Natural Language Processing (NLP) based Innovations for Smart Healthcare Applications in Healthcare 4.0

Nemika Tyagi[1], Bharat Bhushan[1]
[1]School of Engineering and Technology, Sharda University, Greater Noida, Uttar Pradesh, India

Abstract
Technology and computation have changed the backdrop of various aspects of our fast-paced lives. Healthcare is one such aspect that has been affected by this change and faces new challenges every day including the challenge of extracting relevant and valuable information from the enormous amount of data that is generated endlessly in this sector. Smart data analytics provides a solution to this problem through the use of Artificial Intelligence and Natural Language Processing (NLP). This paper elucidates the core concept of textual data analytics that is, NLP, its composition, and architecture. We also present the framework of Healthcare 4.0 and NLP's role in it. Subsequently, we give an elaborated and concise overview of state-of-art NLP technologies that have been employed in various aspects of healthcare and medicine. This paper aims to highlight the role of NLP in smart healthcare and its potential to solve the rising challenges of today's data-driven society.

1. Introduction

The Information and Communication Technology (ICT) advances are being reflected in terms of our changing lifestyles through the concept of smart cities and sustainable development. The primary technologies leading the charge of the ongoing tech revolution also called Fourth Industrial Revolution or Industry 4.0 are the Internet of Things (IoT), Wireless Sensor Network (WSN), big data, cloud computing, machine learning, and data analytics [1]. When applied in the field of medicine, this phase of development is named Healthcare 4.0 and it aims to provide efficient and high-quality healthcare services [2]. The recent decade has especially seen higher investments and focus on the healthcare department from all over the world due to the globalization of this domain and increase in foreign direct investments towards the medical sector in terms of finance, infrastructure as well as technology [3].

Aceto et al. [4] has categorized the ICT aspect of Healthcare 4.0 into four subsets: Sensing – use of IoT devices, wearables, embedded systems; Actuation – robotics, IoT actuators; Communication – WSN, cloud computing, networks; and Processing – Machine Learning, Artificial Intelligence (AI), and Data analytics. Since Healthcare 4.0 is driven by data and the involvement of AI in processing medical data not only provides more efficient and precise results but also gives physicians more direct time with patients [5]. Healthcare analytics requires strategies for dealing with versatile, unstructured data from heterogeneous sources that vary based on semantics, ontologies, and conceptual understanding [6]. In this scenario, Natural Language Processing (NLP) comes into the picture as it presents an AI solution to language data-related problems in Healthcare 4.0.

NLP is an extensive domain of AI and is used in collaboration with machine learning and deep learning algorithms for clinical informatics and data processing. NLP helps in understanding human languages and can assist in automatically analyzing and making decisions by interpreting a large amount of textual data in a very short period. Most of the Clinical-NLP (CNLP) approaches are either rule-based corpus reliant or statistical methods using supervised or semi-supervised algorithms [7]. In healthcare, the data for processing is available in the form of pathology and radiology reports, clinical notes, Electronic Health Records (EHR), nursing documentation, and discharge summaries [8]. CNLP systems are modeled for the requirement of healthcare systems to fit the existing medical pipeline and provide smart solutions. The utilities of CNLP range from providing quick emergency department admissions [9], predicting mortality risk [10], disease prediction from clinical notes [11], identification of patients [12], prognosis of disease [13-14], auto-prescriptions [15], probability of re-admission [16], predicting discharge disposition [17], providing healthcare assistants [18] and end-to-end decision support systems [19]. It is quite evident how CNLP has become an essential instrument in the intelligent medical domain and can take clinical analytics to new heights in the future as well.

In this paper, we shed light on the field of NLP and how it has integrated with smart healthcare over time to solve various medical problems that arise due to the increase in patient demands and simultaneous lack of resources. We found that the scope of NLP in medicine has been on a steady rise and detail-oriented studies have not been conducted that analyze the scope of NLP's applicability in Healthcare 4.0 and provide a review of research work conducted in those areas. The key features of this paper are given below:

- This work provides state-of-art NLP technologies that are setting a benchmark in the healthcare sector.
- This work also focuses on providing a deep understanding of CNLP and the various language tasks that contribute to the NLP pipeline
- The structure of Healthcare 4.0 has been discussed, including the essential aspect of identification of medical data origin sources.
- The applications of NLP in healthcare have been surveyed and analyzed by dividing into 3 separate categories: Healthcare Management and Administration, Assistive care and Clinical Decision Support Systems (DSS), and Disease Diagnosis, Prediction and Treatment.
- This work provides a review of Clinical-NLP systems used in the above domains and the types of methodologies adopted by them.

This paper has been organized as follows: Section 2 provides an overview of NLP, its language tasks, and the structure of the NLP pipeline. Section 3 explores the composition of Healthcare 4.0 and the role of smart data analytics in it. Section 4 throws light on the existing up-to-date NLP practices applied in the various healthcare domains such as management, assistance, and diagnosis. Finally, section 5 condenses our thoughts about this study.

2. Natural Language Processing: an Overview

Processing human language-based data using computational methods or automated techniques is becoming a big challenge in today's data-driven society. NLP is a branch of computational science that helps in understanding as well as generating natural human language queries and responses. Therefore NLP is the right fit to bridge the gap between the rapid generation of language data and the need to rapidly process this data to produce relevant information. NLP uses a combination of Artificial Intelligence and linguistics to analyze the subject matter written or spoken in human languages [20]. NLP when used with machine learning and deep learning technologies produces highly accurate solutions to problems that require text analysis, classification, segmentation, summary generation, and machine translation among others [21].

2.1. What is NLP?

NLP is a branch of Artificial Intelligence used for processing all kinds of language information such as shape, sound, tonality, semantics, and context [22]. It can be broadly classified into two major paradigms: Natural Language Understanding (NLU) and Natural Language Generation (NLG). NLU is the branch of NLP that is closely related to linguistics and it focuses on coarse to fine grain levels of language interpretation. NLU includes the study of Syntax – the structure of the sentence, Semantics – the meaningfulness of sentences or paragraphs, Phonology – the sound of the language, Morphology – the process of formation of words and Pragmatics – the understanding of contextual meaning of the language [20]. NLG is deployed in systems that require giving natural language responses. It consists of the following phases: Selection of Content – deciding the subject matter relevant to the query, Linguistic Resources – making use of the available resources for response generation, Textual organization – arranging the text to make syntactically and semantically correct sentences and Realization – the last phase where the final response is delivered in text or audio format. Figure 1. depicts these sub-sections of NLP.

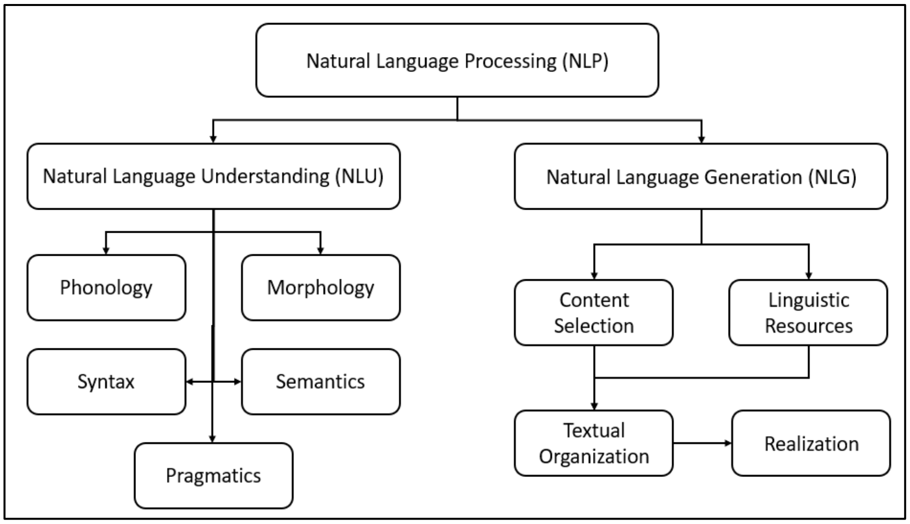

Fig 1. Branches of NLP and their sub-sections

NLP does not refer to a single algorithm or program but it is a consolidation of several processes and tasks that take place for understanding the actual meaning and context of the words including the temporal, spatial, and semantic insinuations. Consequently, NLP is regarded as a sub-domain of Artificial Intelligence since it requires intelligence on par with humans for the interpretation of language. As of now, we rely more on human-AI collaboration for executing NLP tasks than being solely dependent on computational acumen. The reason is, that several NLP processes still heavily depend on supervised learning approaches for high performance and need human intervention for understanding the NLP model's behavior, justify the predictions made by the model, and assess the uncertain or incorrect predictions [23].

2.2. Language Tasks in NLP

Depending on the purpose of the system or application, different NLP tasks can be used individually or in collaboration. Having an overview of the underlying NLP tasks can be useful while trying to solve a specific problem. This sub-section throws light on some of the most important natural language tasks and their role in the textual data processing.

2.2.1. Information Extraction

Information Extraction (IE) is one of the most important NLP processes and it is used for scraping the significant information from a large collection of related data. IE is a complex procedure and can be enhanced to extract information from structured as well as unstructured data sources [24]. This information is labeled and stored for further analysis under classes such as entity, object, temporal sequence, relations, types, and more. Fei et al. [25] made use of biomedical knowledge graphs on a big scale for biomedical IE to capture the contextualized concepts from available data. To identify, extract and timely arrange a mental health-related disorder, Zirikly et al. [26] used NLP to extract all the relevant information for diagnostic and treatment purposes.

2.2.2. Named Entity Recognition

Named Entity Recognition (NER) is a sub-task within IE and its main role is to extract proper nouns from a given structured or unstructured textual data [27]. It can be used to identify people, places, organizations, occupations, products, and other designators from the text. NER is a basic step that leads to several NLP tasks such as IE, question answering, machine translation, and even text summarization [28]. With a pre-trained BERT model, Li et al. [29] used clinical NER for the classification and identification of medical entities in Chinese medical records. Weber et al. [30] proposed a novel NER tagger called *HunFlair* for identifying biomedical entities.

2.2.3. Sentence classification

Classification of text is another important NLP task that makes use of word embeddings and other representations of words to classify them separately. Bag of words models and word vector models are popular sentence classification approaches but nowadays deep learning and neural network especially Convolutional Neural networks (CNN) and Recurrent Neural networks (RNN) are producing more effective results [31]. Guo et al. [32] proposed a data augmentation method-based approach for the classification of sentences using word embeddings and sentence embeddings. To obtain a patient's past medical history from EHRs, Bagheri et al. [33] proposed enrichment by topic modeling for the task of sentence classification.

2.2.4. Document classification

To maintain a large set of files and documents for archival, procedural, and management purposes document classification is necessary. Manual classification of documents can be a tedious process especially when big corporations and organizations are considered. To automate this process computer vision, deep learning classifiers, and NLP are used in combination to produce fast and accurate results [34]. Intending to evaluate the performance of Deep Neural Networks (DNN), Behera et al. [35] analyzed the different models used for biomedical document classification. Nadif et al. [36] made use of supervised and unsupervised text mining along with vector-based word embeddings for classifying documents.

2.2.5. Text summarization

Going through an entire lengthy document to find significant information can be tiresome and wasteful. To accelerate this procedure automatic text summarization, both abstractive and extractive, can provide a gist of important subject matter from the document. Extractive summarization makes use of the frequency of word appearances to construct the summary, whereas abstractive summarization reconstructs the document information to focus on the central idea [37]. Belwal et al. [38] proposed a graph-based text summarization technique to rank sentences based on similarities with each other and the entire document. To improve the performance of the extractive summarization technique, Belwal et al. [39] further introduced another technique by integrating semantic measures and topic modeling with a vector space system.

2.2.6. Question Answering

Question answering systems have been gaining popularity due to the increase in the variety and usage of Artificial Intelligence empowered voice assistant services such as Apple's Siri, Amazon's Echo, and Google assistant. Besides the smartphone-based systems, many chatbot services and even IoT-based devices also make use of question answering either in text or speech mode. QA systems are used in combination with Information Retrieval (IR) to understand queries in human languages and deliver precise and appropriate responses [40]. Yin et al. [41] created a Spatio-temporal analysis and visualization tool using NLP-enabled-QA to give real-time instructions and queries. In another work, Meichanetzidis et al. [42] proposed a Quantum NLP model for constructing a grammar-aware QA model.

2.2.7. Machine Translation

Machine translation is a very practical NLP task that refers to the automatic translation of human languages using computers or machines. The most popular approaches for carrying out this task are rule-based, statistical, and neural machine translation [43]. The main focus during machine translation is retaining the actual meaning, grammar rules, and ideas behind the sentences where this translation is applied. Laskar et al. [44] demonstrated the use of neural machine translation with the help of two different models for translating English to the Hindi language. In another interesting scenario, Rahit et al. [45] proposed a machine learning system to convert natural language instructions to programming language codes. Table 1 gives a summary of the various language tasks in NLP and their recent advancements.

Table 1. Comprehensive overview of NLP tasks and their recent applications

S. No.	NLP Task	Purpose	Recent Works	Unit of Annotation
1.	Information Extraction	Extracting task-relevant information from a large collection of data	Fei et al. [25], Zirikly et al. [26]	Lexicon
2.	Named Entity Recognition	Identifying nouns from the data that can be tagged as an entity, event, person, and other classes	Li et al. [29], Weber et al. [30]	Lexicon
3.	Sentence Classification	Classifying sentences from data records to find a specific collection of information	Guo et al. [32], Bagheri et al. [33]	Sentence

4.	Document Classification	Classifying entire documents for better retrieval of information and reducing the amount of search	Behera et al. [35], Nadif et al. [36]	Document
5.	Text Summarization	Summarizing the content of large datasets into humanly comprehensible summaries	Belwal et al. [38], Belwal et al. [39]	Document
6.	Question Answering	Creating interactive dialogue to enable human-computer communication	Yin et al. [41], Meichanetzidis et al. [42]	Document
7.	Machine Translation	Translating one human language to another computationally	Laskar et al. [44], Rahit et al. [45]	Sentence / Document

2.3. NLP Pipeline

Getting an overview of the language tasks in NLP gives an insight into how these tasks can not only be applied individually but also in combination with other tasks to solve a problem. Most NLP modules consist of four essential phases: *Data collection* – here the data that is to be processed is gathered from various sources be it text or speech, structured or unstructured. A lot of times if machine learning approaches are being used for carrying out a language task, the training corpus is also gathered or created in this phase. *Data preprocessing* – in this phase the collected data is processed to remove any noise, discrepancies, or inconsistencies to make it more uniform and accurate. Avoiding this phase may lead to complete failure of NLP application and heavily impact the performance of the system. *Feature engineering* – this is the process of obtaining required attributes from the pre-processed data and modifying them to fit the requisite NLP approach (rule-based, statistical, or neural network-based) of the given system. *Modeling* – this is the final step in which the kind of natural language task or combination of tasks required to tackle the problem is decided and modeled to fit the NLP pipeline. The final output is produced after the modeling phase and evaluated based on various performance metrics. The flow of the NLP pipeline is depicted in figure 2.

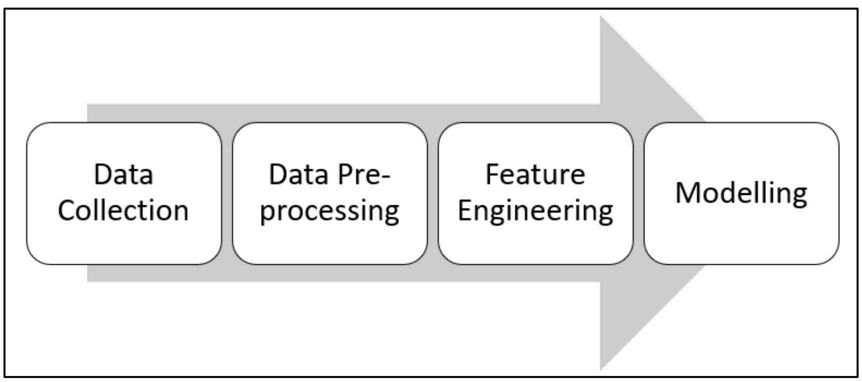

Fig 2. NLP pipeline sequence

3. Healthcare 4.0

With the developments in ICT, we have witnessed the revolutionization of cities, industries, energy production, transportation, education, and healthcare services. Today we live in the era of Healthcare 4.0 where the healthcare industry is driven by numerous smart technologies such as IoT, Big Data Analytics (BDA), Artificial Intelligence, Machine Learning, Cloud computing, Virtual reality, Computer vision, and Augmented reality [46]. One common denominator in all these techniques is the presence and requirement of a large amount of data. Smart devices and embedded systems generate and collect an enormous quantity of medical data every single day. This data is stored via cloud architecture and processed using Machine Learning and Artificial Intelligence to conduct BDA to produce valuable insights. Manual data processing has taken a setback at present and given way to the large-scale automated processing of data for utilities such as NLP, computer vision, Virtual/Augmented reality, and a lot more. The decision-making process is no longer solely human-reliant but is actively influenced by the knowledge provided via computational analytics.

3.1. Smart Data Analytics using Machine Learning

Big data is defined by the 4 V's namely : Volume – the sheer amount of data being generated every day, Velocity – the high speed at which this data is being generated, Variety – the different formats in which this data is collected, and the Veracity – the inconsistencies and trustworthiness of the data that is collected. In the medical sector as well big data is generated around the clock through sensing and monitoring devices, patient logs, discharge summaries, clinical notes, prescriptions, reports, scans, and other paperwork. In this scenario, healthcare facilities have turned toward Smart Data Analytics techniques to increase the pace of decision-making systems. Most of the collected data consists of either redundant, inaccurate, irrelevant, or incomplete information which poses challenges for data analytics due to resource as well as time constraints [47]. Machine learning provides a respite from these impediments through various data mining, text mining, and data analytics methods. These machine learning-based smart data analytics techniques range from regression, classification, clustering, sequential pattern mining, NLP, temporal reasoning, anomaly detection, association rule mining, and a lot more [48]. These algorithms have contributed to the diagnostic and prognostic evaluation of patient reports [49] as well as the prediction of the mortality rate of patients based on their records [50]. The scope of machine learning algorithms in the healthcare industry has no bounds and new ways are discovered daily for optimizing the tedious processes and enhancing the decision-making processes in the medical field using this technology.

3.2. Medical Data Sources

The sources of data in the medical sector are diverse and it is imperative to pay attention to the quality of data as it leads to serious implications in healthcare analysis. Recognizing the source of data assists in devising more suitable strategies for dealing with the inconsistencies, inaccuracies, or redundancy of the data. The requirement of every data analytics model is different and even the pre-processing techniques of analysis also adapt accordingly. Being able to deal with data quality-related problems at the source itself reduces the amount of work that goes into processing low-quality data and prevents erroneous outcomes.

IoT devices and wearable gadgets produce a constant stream of timestamped data that is channeled via local networks and preferably stored using cloud systems. This data is generated at a very high speed and is also prone to latency issues [51] and inconsistent real-time streaming of sensor data. This kind of data requires high-speed processing with less computational overhead and minimum inaccuracies. Monitoring devices with embedded systems [52] in healthcare facilities also produce constant data streams that are collected over time and are used in systems to alert patients and authorities instantaneously. The most prominent form of data in the medical sector to date is the manually generated or entered data, be it in electronic format or paper-based format. The admission forms and discharge summaries consist of valuable information about the state of a patient and the kind of diagnosis they have gone through. The clinical notes written by the doctors, nurses, or medical staff helps in understanding the treatment process of the administered patients. Reports in the form of descriptive, numeric, graphic, or visual data are used to gain insight into the present condition of a healthcare patient. Figure 3. shows the variants of medical data sources and the analytical processes applied on them.

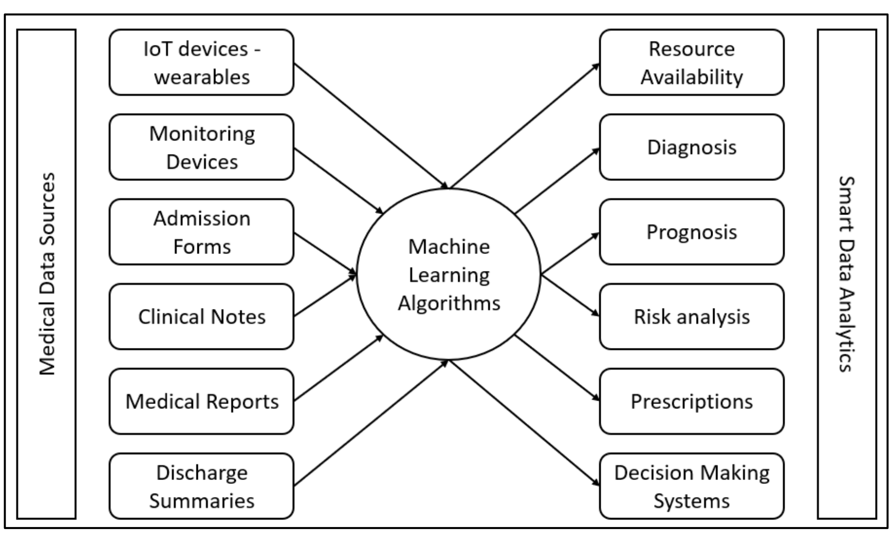

Fig 3. Process of Medical data analytics

A wide range of data sources are used for analytical procedures such as: checking the availability of resources, urgent patient care, diagnostic analysis, prognostic analysis, post-treatment risk analysis, and automatic prescriptions. Such kinds of analysis contribute to the formation of intelligent DSS. Zhan et al. [53] developed an assistive diagnosis and smart decision-making system that used sensor data for conducting an early diagnosis of NSCLC or Non-Small Cell Lung Cancer. To observe and effectively diagnose type-2 diabetes patients Abdel-Basset et al. [54] introduced an IoT and soft computing-based intelligent framework. In another work, Wadia et al. [66] implemented a NLP-based Clinical DSS for identifying post-colonoscopy patient cases that needed close follow-up and scheduling. Similarly, Shen et al. [68] proposed a CDSS to determine the order of sedation for outpatient endoscopy patients to rectify the sedation-type errors caused by human interventions. Table 2. gives a relative summary of the various CDSS used in medical healthcare system.

Table 2. Overview of NLP based CDSS

S. No.	**Work, Year**	**Disease**	**Data Source**	**Lexical Unit**
1.	Wadia et al. [66], 2017	Colon Cancer Care	Pathology Reports	Lexicon
2.	Abdel-Basset et al. [54], 2019	Type-2 Diabetes	IoT device observations	Lexicon
3.	Shen et al. [68], 2020	Outpatient Endoscopy	EHR	Sentence
4.	Zhan et al. [53], 2021	Non-Small Cell Lung Cancer	Medical Records	Document

4. Motivation

Clinical experts have always made use of medical data to conduct relevant studies but as healthcare becomes more accessible to the general public the amount of medical data is rapidly rising [8]. Manually conducting such analysis is becoming infeasible given the constraint of resources in healthcare facilities and time constrictions. This creates a need for automation of data processing and NLP provides the biggest solution to process language-based medical data. The popularity of Clinical-NLP has increased significantly in the last few years because of the introduction of several open-source NLP platforms such as Clinical Language Annotation, Modeling, and Processing Toolkit (CLAMP), clinical Text Analysis Knowledge Extraction System (cTAKES), SemEHR a tool for information extraction and general-purpose NLP tools like General Architecture For Text Engineering (GATE) [56]. An increase in the availability of electronic documents such as Electronic Health Records (EHRs) facilitates the easy structuring of data for obtaining diagnostic, semantic, and temporal information. Semantic analysis is an important aspect of CNLP for understanding the context behind words and making applications smarter and not just statistically driven.

Clinical text is different from regular language text since it consists of medical vocabulary and even telegraphic speech patterns [55]. The corpus used for training purposes is specialized to conduct domain-specific analysis and accurate tagging, annotation, and parsing of data. With further enrichment in the field of CNLP research and development, the prospects of improving smart healthcare systems are exceedingly bright. Healthcare 4.0 is a data-driven phase of the medical revolution and technology like NLP offers the chance to accelerate our vision of providing more efficient and accessible healthcare to people all around the world. Figure 4 depicts the current state of healthcare and its infrastructure in the top 5 most populated countries in the world as of 2022 [56-57].

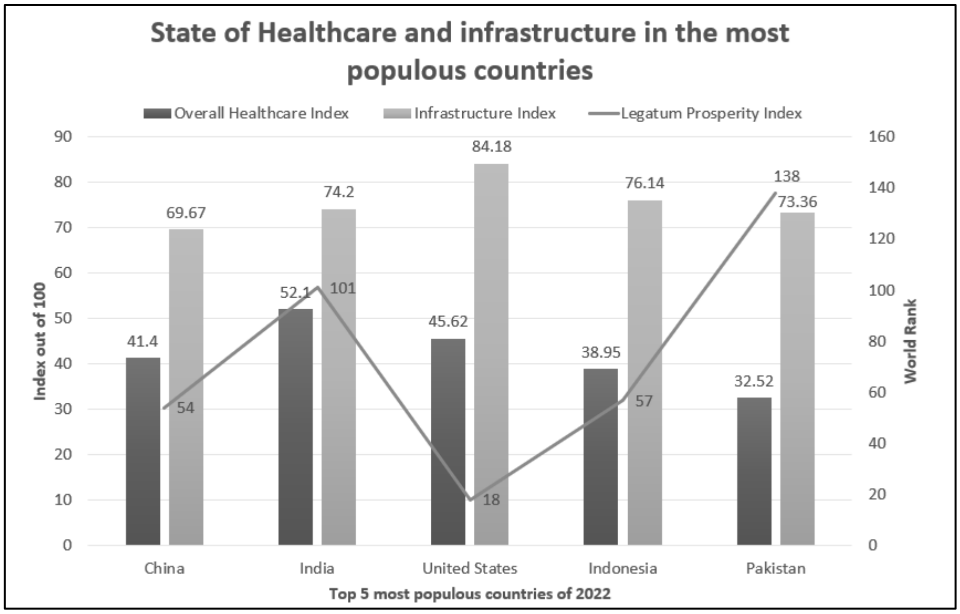

Fig 4. Healthcare statistics of most populous countries of 2022

5. Smart Healthcare Applications of NLP

The possibilities of deploying NLP in Healthcare 4.0 are plenty especially since the commencement of the Precision Medicine Initiative (PMI) back in 2015. PMI helps in giving access to and connecting the large-scale medicinal and biological databases that consider individual variabilities in the form of genomics, proteomics, metabolomics, and mobile health technologies with powerful ways of analyzing these datasets using precision computing [58]. An initiative like this has encouraged researchers to find novel and creative ways for incorporating precision technologies using machine learning and NLP on textual data in every medical department. This section provides a deeper insight into the development of NLP models over time to suit the requirements of various clinical tasks. The applications of NLP in smart healthcare have been divided into 3 sub-sections: Healthcare Management and Administration, Assistive Care and Clinical DSS, and Disease Diagnosis, Prediction, and Treatment.

5.1. Healthcare Management and Administration

Overcrowding in emergency departments (ED) leads to crucial waste of time and to resolve this issue Zhang et al. [59] proposed a prediction model for ED admission using initial presentation data. In a similar work, Sterling et al. [60] used nursing triage documentation to predict final ED disposition. Risk analysis of a patient is a fundamental research field to estimate the severity, time of stay, chances of mortality, and cost of procedures. For predicting the severity of an Intensive Care Unit (ICU) admitted patient’s ailment and the in-hospital mortality risk Jin et al. [61] proposed a neural network-based NLP model.

Nawab et al. [62] developed a system to extract insight from patient feedback forms for reimbursement-related studies. Discharge planning involves predicting the length of stay of a patent and Bacchi et al. [63] proposed a model to predict the general medical admission length of stay with NLP and deep learning. In 2021 Arnaud et al. [64] proposed a prediction model using triage notes to project medical specialties at hospitals in advance. To compare the performance of the Confusion Assessment Method for intensive care unit (CAM-ICU) and Natural Language Processing (NLP) diagnosed behavioral disturbance (NLP-Dx-BD) for delirium, Young et al. [65] developed an analytical system. Table 3. provides an overview of the NLP systems described in this section.

Table 3. A comprehensive review of NLP applications in Healthcare Management and Administration

S.No.	Author, Year	Purpose	Data Source	Methodology	Performance	Limitation
1.	Jin et al. [61], 2018	Mortality risk analysis of ICU patients	MIMIC-III data set	Word embeddings, LSTM, and multi-modal Neural	AU-ROC (%) – LSTM: 0.8531; Multi-modal:	Extracted entities could not be normalized

				Network	0.8734	
2.	Zhang et al. [59], 2019	Prediction of admission to ED	NHMACS US ED dataset	LR and MLNN with NLP	AUC – LR: 0.824; MLNN: 0.823	Contextual information was ignored
3.	Sterling et al. [60], 2019	Predict ED disposition at admission	Data from 3 academically-affiliated EDs	Neural network regression-based NLP model	AOC – Bag-of-words: 0.737; para-vectors: 0.785	Prone to mis-classification and biases
4.	Nawab et al. [62], 2020	Analysis of patient experience	Press Ganey Database of patient feedback	POS tagging and sentiment analysis	Positive and negative feedback classification	Lacks entity linking and contextual understanding
5.	Bacchi et al. [63], 2020	Length of stay prediction	AMU of The Royal Adelaide Hospital	NLP and LR, CNN, ANN, Random Forrest	ANN had highest AUC: 0.75	Limited sample size
6.	Arnaud et al. [64], 2021	Medical specialties at hospital prediction	Amiens-Picardy University Hospital dataset	NLP, word embeddings, CNN and MLP	Accuracy: 0.68	Model sensitivity was low
7.	Young et al. [65], 2022	NLP-Dx-BD vs CAM-ICU to assess delirium in ICU patients	Nursing Progress Notes	NLP-based behavioral diagnosis and CAM analysis	NLP-Dx-BD identifies more patients likely to receive antipsychotic medications	No evaluation was done by independent, psychiatrically trained clinician

(Abbreviations: MIMIC - Medical Information Mart For Intensive Care, AU-ROC - Area Under The Receiver Operating Characteristic Curve, NHAMCS - National Hospital Ambulatory Medical Care Survey, LR - Logistic Regression, MLNN - Multilayer Neural Network Models, POS – Part of Speech, AMU - Acute Medical Unit, MLP - Multi-Layer Perceptron)

5.2. Assistive Care and Clinical DSS

Decision support systems are great assistance to healthcare professionals when making a dynamic prediction based on knowledge, cognition, and analytical skills. Wadia et al. [66] described a Clinical-DSS based on NLP designed for colon cancer care coordinators which can be used to monitor post-colonoscopy patients. In another work, Zikos et al. [67] introduced a CDSS reference model for designing contextually relevant DSS. Shen et al. [68] developed a CDSS for adopting a suitable endoscopy sedation strategy for patients dealing with outpatient endoscopy using NLP and consensus-derived logic. A Portuguese CDSS system was proposed by Leite-Moreira et al. [69] to promote the use of NLP in clinical decision-making.

Social or nursing robots in healthcare and assistive chatbots have a lot of use in the care, monitoring, and treatment procedures of patients. NLP provides a means of communication and understanding between these assistive care units and the end-users. Dino et al. [70] developed a Socially Assistive Robot, Ryan, for providing internet-delivered cognitive-behavioral therapy (CBT) to elderly depression patients. For assisting patients that don't know the workings of the outpatient department, Chen et al. [71] proposed an attention-based bidirectional long short-term memory (Att-BiLSTM) system for service robots. Christopherjames et al. [72] introduced a healthcare chatbot for assisting people with their medical concerns and corresponding treatments. In a similar work, Hassan et al. [73] presented a chatbot system for e-mental health care. Table 4. provides an overview of the NLP systems described in this section.

Table 4. A comprehensive review of NLP applications in Assistive Care and Clinical DSS

S.No.	Author, Year	Purpose	Data Source	Methodology	Result / Performance	Limitation
1.	Wadia et al. [66], 2017	CDSS for post-colonoscopy patients	VA Connecticut Health Care System pathology reports	Structured Query Language (SQL) function and text categorization using NLP	Accuracy: 0.985	Does not have a web-based user interface
2.	Zikos et al. [67], 2018	CDSS reference	External data providers	Conceptual development of	Generalized CDSS-RM	Infrastructural aspects of health

		model		CDSS using NLP	framework	facilities not considered
3.	Dino et al. [70], 2019	CBT for older adults with depression	Dialogue training on human language data	AIML-based dialogue manager system using NLP	Positive survey responses	Small-sample size and short research period
4.	Shen et al. [68], 2020	CDSS for outpatient endoscopy	EHR from Brigham and Women's Hospital	SQL, NLP using heuristic check and validation	Sedation-type order error rate decreased to 0.037%	Relied heavily on historical endoscopy data
5.	Chen et al. [71], 2020	Service robot for outpatient department	Dialogue data from Taiwan E Hospital	NLP, TF–IDF for text processing and Att-BiLSTM	Accuracy: 0.96; Precision: 0.96	The dialogue system is not operative for other departments
6.	Christopherjames et al. [72], 2021	Healthcare chatbot	Chatbot trained with Google assistant	NLP and ML algorithms by Dialogflow	Real-time application	Does not provide health parameters to the user
7.	Hassan et al. [73], 2021	e-Mental health chatbot	MBTI dataset	NLTK, LSTM and computer vision	A web-based application	Dataset requires further expansion
8.	Leite-Moreira et al. [69], 2022	Portuguese CDSS	EHR: clinical notes, discharge summaries, test results	Transition-based parser, MATPD, and SAACE	Favorable and rapid predictions	Semantic contextualization is lacking

(Abbreviations: TF–IDF - Term Frequency-Inverse Document Frequency, NLTK - Natural Language Toolkit)

5.3. Disease Diagnosis, Prediction, and Treatment

EHRs, clinical notes, reports, and discharge summaries contain a lot of information about a patient's medical history, present condition, statistics, clinical procedures or surgeries, prescriptions, and final comments. These data sources have been used in CNLP to make predictions about diseases and risks related to them, the prospective diagnosis after analyzing the present condition, suggesting treatments and medication, forecasting post-operative complications, and finding patterns in clinical data. More importantly, chronic diseases benefit the most from smart analysis of data since they have a longitudinal nature and consist of plenty of temporal information and trends that might be overlooked when conducting clinical analysis [74]. This section reviews various CNLP models that were used for the diagnosis, prediction, and treatment of the following medical conditions: Mental healthcare, Cancer research, Circulatory system diseases, and miscellaneous diseases.

5.3.1. Mental Healthcare

Early prediction of mental health-related illnesses at the time of their onset can help doctors and physicians devise better treatments at an early stage itself. Jackson et al. [75] used NLP to construct a language model that can detect the symptoms of severe mental illness (SMI) from clinical text. Detecting patients at risk of self-harm can be done using NLP of case notes in EHR, as demonstrated in a work by Van Le et al. [76]. In another work, Mulyana et al. [77] proposed case-based reasoning (CBR) computer system to diagnose medical illnesses from health records using NLP to compensate for the lack of professional psychologists. An Internet-Delivered Psychological Treatment (IDPT) system was developed with the help of patient-authored clinical data and NLP [78]. Ridgway et al. [79] introduced an NLP model to detect mental disorders or indicators of substance abuse among HV patients.

5.3.2. Cancer Research

Research in the field of cancer has been a very demanding and progressive field for decades including studies related to pathology, radiology reports, narrative reports, histopathology, information extraction from clinical notes, and more. Karunakaran et al. [80] applied NLP to find lung cancer patients from CT-scan notes created by radiologists using Geisinger's Close-the-Loop clinical program. In another approach, Si et al. [81] proposed an NLP model for extracting information about cancer diagnosis, cancer therapeutic procedure, and tumor description from EHR. To determine the onset of familial breast cancer and colorectal cancer from Family health history (FHH), Mowery et al. [82] made use of NLP classifiers. Alawad et al. [83] introduced a deep learning NLP-based privacy-preserving transfer learning model for cancer-related information from pathology documents. Deshmukh et al. [84] used NLP for extracting vital cancer information at the prognostic stage itself from breast cancer medical records.

To quantify the privacy of deep learning models for IE from cancer pathology reports Yoon et al. [85] developed a vocabulary selection NLP method.

5.3.3. Circulatory System Disease

Circulatory system disorders are related to heart failure, cardiology, arterial diseases, cardiovascular disorders, and risk prediction for the onset of such diseases. To investigate how cardiovascular disease operates in HIV patients, Patterson et al. [86] developed an NLP system to analyze echocardiogram reports. Afzal et al. [87] introduced an algorithm for identifying Critical limb ischemia (CLI) cases in narrative clinical notes of peripheral artery disease (PAD) patients. For classifying patients with heart disease and predicting various cardiovascular disorders, Thaiparnit et al. [88] proposed a technique for clinical data extraction. Bagheri et al. [89] designed a system for cardiovascular risk prediction using multimodal BiLSTM on structured EHR documents. In another work, Sammani et al. [90] aimed to create a model for automated classification of reliable ICD-10 codes using free medical text available in the cardiology department. Zaman et al. [91] created a framework using semi-supervised NLP for automated diagnosis categorization of Cardiovascular MRI from clinical text reports.

5.3.4. Miscellaneous

The prospects of applying NLP in smart healthcare prediction, diagnosis, and prognosis are immense and go beyond the categorizations provided in the above sections. Weng et al. [92] made use of NLP for the classification of medical notes into sub-domains using a machine-learning-based model. For bridging the gap between clinical experts and NLP, Trivedi et al. [93] proposed an interactive NLP tool called NLPReViz. Chen et al. [94] used EHR clinical notes to automatically procure patients' geriatric syndromes from the medical text. Septic shocks are a major health concern, yet often overlooked by smart healthcare experts. Using a gradient boosting algorithm on clinical notes, Liu et al. [95] computed the time-impending risk of developing septic shocks. Oliwa et al. [96] developed a predictive model to identify symptoms of lost to follow-up (LTFU) HIV patients from unstructured notes. For detecting adverse drug events (ADEs), Chen et al. [97] introduced a knowledge-based NLP system and relation identifier using an attention-based BiLSTM network. An NLP tool called COVID-19 SignSym for extracting Covid-19 symptoms from free text EHRs was created by Wang et al. [98] using CLAMP. In a different work Song et al. [99] used NLP for the automatic extraction of medical data from unstructured esophagogastroduodenoscopy (EGD) reports. Table 5. provides an overview of the various NLP models discussed in the above section in a detailed yet precise manner.

Table 5. A comprehensive review of NLP applications in healthcare diagnosis, prediction, treatment

Cate-gory	Author, Year	Purpose	Data Source	Methodology	Result / Performance	Limitation
Mental Health	Jackson et al. [75], 2017	Detecting SMI from clinical text	CRIS data resources and EHR from hospitals	Information extraction, annotation, and classification	Median F1 score: 0.88 for 46 SMI symptoms	Underestimation of symptoms occurred
	Van Le et al. [76], 2018	Risk analysis of self-harm-prone patients	EHR of de-identified forensic inpatient notes	Bagging, J48, Jrip, Logistic Model Trees (LMT), LR, and SVM	LMT and SVM algorithm were best	Synonyms and language variants were not incorporated
	Mulyana et al. [77], 2019	CBR system for mental disorders	Medical records	NLP and CBR processing	Pattern matching was successful	Limited dataset and scope
	Mukhiya et al. [78], 2020	IDPT system to diagnose depression	PHQ-9 questionnaire	Depression2Vec for word embedding and cosine similarity	Performance is comparable to WordNet	Evaluation is done on a human-annotated dataset
	Ridgway et al. [79], 2021	Mental illness or substance abuse in HIV patients	Structured and unstructured EMR data	CoreNLP and NegEx for NLP	PPV of 98% and NPV of 98%	Cases of false detection are present
Cancer Research	Karunakaran et al. [80], 2017	Identifying lung cancer patients from radiology reports	Scanned radiology reports	Hadoop for data storage, cTAKES PoS tagging, NER	F1-score: 0.908; Precision: 0.873	Study on a specific use case and has not been generalized
	Si et al. [81],	Cancer-related	Clinical notes	Bi-LSTM CRF	F1: cancer	Multi-step

	2018	information extraction	from UT Physicians data warehouse	Networks and word embeddings	diagnosis - 93.70; therapy procedure: 96.33	evaluation not conducted
	Mowery et al. [82], 2019	Extract breast and colorectal cancer from FHH	EHR from the University of Utah Health Enterprise	Frequency pattern and NLP classifiers	NLP classifier outperformed; Precision: 96%	Study based on a single academic healthcare network
	Alawad et al. [83], 2020	Cancer IE from pathology documents	Text corpora of cancer pathology reports	Multitask learning CNN NLP algorithm	F1-score: Micro - 0.823; Macro - 0 .580	Performance is not as good in low prevalence class labels
	Deshmukh et al. [84], 2021	Prognostic stage of breast cancer prediction	Pathological and clinical reports	NLP, Decision tree, K-fold cross-validation	Prediction accuracy: urban - 0.92; rural – 0.82	Small-sample size to quantify the privacy data
	Yoon et al. [85], 2022	Securing patients' information while IE	Cancer pathology reports	Multitask CNN and Membership inference attacks	Lower privacy vulnerability and good IE performance	The study requires multiple data providers
Circulatory System Disease	Patterson et al. [86], 2017	Cardiovascular disease identification in HIV patients	General clinic notes, echo-cardiogram, and radiology reports	Apache UIMA AS) framework for NLP	F1-scores: 0.872, 0.844, and 0.877 for each report type	Poor performance for rare measures
	Afzal et al. [87], 2018	Identifying CLI cases among PAD patients	Clinical notes from Mayo clinical data warehouse	CLI-NLP algorithm for patient classification	CLI-NLP PPV: 0.96; F1-score: 0.90	Data from only a single medical center was used
	Thaiparnit et al. [88], 2019	Extraction of data for cardio-disease	University of California Irvine dataset	Vertical Hoeffding Decision Tree algorithm	Accuracy: 0.8543;	The system can be made more accurate
	Bagheri et al. [89], 2020	Cardiovascular risk prediction	EHR from SMART study	Multimodal BiLSTM for NLP	AUC: 0.847; Misclassification rate: 0.143	Not integrated with Clinical DSS
	Sammani et al. [90], 2021	Automated classification of reliable ICD-10 codes	Discharge letters from cardiology patients	Bidirectional Gated Recurrent Unit Neural Network	F1 scores: 0.76–0.99 for 3-char codes; 0.87–0.98 for 4-char codes	Required manual assessment for discrepancies in performance
	Zaman et al. [91], 2022	Automatic Diagnosis Labeling of MRI reports	Cardiac MRI reports	BERT, rule-based model, and SVM model	BERT model performed best; F1-score: 0.86	The model was not fully accurate across all categories
Miscellaneous	Weng et al. [92], 2017	Sub-domain classification	MGH, and iDASH data repository	CRNN with neural word embedding and TF-IDF, cTAKES	CRNN AUC: iDASH - 0.975; MGH - 0.991	cTAKES may not be the most suitable NLP tool here
	Trivedi et al. [93], 2018	Interactive NLP tool for binary concept	Colonoscopy reports	NLPReViz using bag-of-words and SVM	F1 scores: between 0.78 and 0.91	Integration with other NLP tools not explored
	Chen et al. [94], 2019	Determining patients with geriatric syndromes	Anonymized EHR data multispecialty organization	cTAKES and CRF	Macro F1-score: 0.834; Micro F1-score: 0.851	Annotations were slightly inconsistent
	Liu et al. [95], 2019	Predicting septic-shock onset patients	MIMIC-III clinical database	text2vec, GloVe, and XGBoost	AUC: 0.92; PPV: 0.49; early warning: 7hrs	Better labels for annotations are available
	Oliwa et al. [96], 2020	Classify LTFU and retained HIV patients	Outpatient HIV care clinical notes	Bag-of-words, TF-IDF, 10-fold cross-validation	Weighted F1-score mean of 0.912	An automatic ontology match was not used
	Chen et al. [97], 2020	ADEs detection from clinical text	MIMIC-III clinical care database	Knowledge-based NLP system and BiLSTM	F-measure of 0.9442	Sense ambiguity, relation classifier matching error

	Wang et al. [98], 2021	Extracting covid-19 symptoms	Notes from MIMIC-III, UTP, KUMC, Johns Hopkins	COVID-19 SignSym, deep learning, pattern-based rules	F1-measure: 0.972; recall: 0.992	Not generalized, more extraction of information is possible
	Song et al. [99], 2022	Gastric disease detection	EGD reports	IE and concept summarization	F1 score: 0.967; PPV: 0.972	Not adaptable with document formats

(Abbreviation: CRIS - Clinical Record Interactive Search, PPV - Positive Predictive Value, NPV - Negative Predictive Value, UIMA-AS - Unstructured Information Management Architecture Asynchronous Scaleout, SMART - Second Manifestations Of Arterial Disease Study, BERT - Bidirectional Encoder Representations From Transformers, MRI - Magnetic Resonance Imaging, iDASH - Integrating Data For Analysis, Anonymization, And Sharing, CRF - Conditional Random Fields, GloVe - Global Vectors For Word Representation, CLAMP - Clinical Language Annotation, Modeling, And Processing)

6. Conclusion

The healthcare domain is a challenging technical field, especially since the flooding of the enormous amount of data started due to ICT advancements. The 21st century is regarded as the century of data which is quite evident from how much information is being generated every single day. In the field of medicine and healthcare, language-based data is present from the admission into a hospital to the discharge of a patient while also including the monitoring data, reports, diagnosis, and clinical notes. There is a lot of valuable information in this data that can be extracted for progressive healthcare research. Manually doing data analytics is not a feasible option due to time and resource restraints and hence AI technologies like NLP have taken the forefront. This paper begins by throwing light on the concept of NLP and the tasks used to make an NLP pipeline. Next, the features of Healthcare 4.0 were explored along with the concept of smart data analytics. Then the paper devolves into the details of specific NLP models and systems that have been created to assist smart healthcare. This is a study that provides an impression of present NLP technologies in Healthcare 4.0 and it intends to serve as a basis for understanding the landscape of smart healthcare systems and their applications in the future as well.

References:

[1] Quasim, M. T., Khan, M. A., Algarni, F., & Alshahrani, M. M. (2021). Fundamentals of smart cities. In *Smart cities: A data analytics perspective* (pp. 3-16). Springer, Cham.
[2] Malik, P., Pathania, M., & Rathaur, V. K. (2019). Overview of artificial intelligence in medicine. *Journal of family medicine and primary care*, *8*(7), 2328.
[3] Shenkar, O., Liang, G., & Shenkar, R. (2022). The last frontier of globalization: Trade and foreign direct investment in healthcare. *Journal of International Business Studies*, *53*(2), 362-374.
[4] Aceto, G., Persico, V., & Pescapé, A. (2018). The role of Information and Communication Technologies in healthcare: taxonomies, perspectives, and challenges. *Journal of Network and Computer Applications*, *107*, 125-154.
[5] Malik, P., Pathania, M., & Rathaur, V. K. (2019). Overview of artificial intelligence in medicine. *Journal of family medicine and primary care*, *8*(7), 2328.
[6] Kraus, J. M., Lausser, L., Kuhn, P., Jobst, F., Bock, M., Halanke, C., ... & Kestler, H. A. (2018). Big data and precision medicine: challenges and strategies with healthcare data. *International Journal of Data Science and Analytics*, *6*(3), 241-249.
[7] Wen, A., Fu, S., Moon, S., El Wazir, M., Rosenbaum, A., Kaggal, V. C., ... & Fan, J. (2019). Desiderata for delivering NLP to accelerate healthcare AI advancement and a Mayo Clinic NLP-as-a-service implementation. *NPJ digital medicine*, *2*(1), 1-7.
[8] Sheikhalishahi, S., Miotto, R., Dudley, J. T., Lavelli, A., Rinaldi, F., & Osmani, V. (2019). Natural language processing of clinical notes on chronic diseases: a systematic review. *JMIR medical informatics*, *7*(2), e12239.
[9] Klang, E., Kummer, B. R., Dangayach, N. S., Zhong, A., Kia, M. A., Timsina, P., ... & Oermann, E. K. (2021). Predicting adult neuroscience intensive care unit admission from emergency department triage using a retrospective, tabular-free text machine learning approach. *Scientific reports*, *11*(1), 1-9.
[10] Mahbub, M., Srinivasan, S., Danciu, I., Peluso, A., Begoli, E., Tamang, S., & Peterson, G. D. (2022). Unstructured clinical notes within the 24 hours since admission predict short, mid & long-term mortality in adult ICU patients. *Plos one*, *17*(1), e0262182.

[11] Abokhzam, A. A., Gupta, N. K., & Bose, D. K. (2021). Efficient diabetes mellitus prediction with grid based random forest classifier in association with natural language processing. *International Journal of Speech Technology*, *24*(3), 601-614.
[12] Weissler, E. H., Zhang, J., Lippmann, S., Rusincovitch, S., Henao, R., & Jones, W. S. (2020). Use of natural language processing to improve identification of patients with peripheral artery disease. *Circulation: Cardiovascular Interventions*, *13*(10), e009447.
[13] Sohn, S., Wi, C. I., Wu, S. T., Liu, H., Ryu, E., Krusemark, E., ... & Juhn, Y. J. (2018). Ascertainment of asthma prognosis using natural language processing from electronic medical records. *Journal of Allergy and Clinical Immunology*, *141*(6), 2292-2294.
[14] Hossain, M. T., Talukder, M. A. R., & Jahan, N. (2022). Depression prognosis using natural language processing and machine learning from social media status. *International Journal of Electrical and Computer Engineering*, *12*(3), 2847.
[15] Mahatpure, J., Motwani, M., & Shukla, P. K. (2019). An electronic prescription system powered by speech recognition, natural language processing and blockchain technology. *International Journal of Science & Technology Research (IJSTR)*, *8*(08), 1454-1462.
[16] Dhanalakshmi, T. S., & Meleet, M. (2020, June). Predicting Clinical Re-admission using Discharge Summaries (PCRUDS). In *2020 5th International Conference on Communication and Electronics Systems (ICCES)* (pp. 772-777). IEEE.
[17] Muhlestein, W. E., Monsour, M. A., Friedman, G. N., Zinzuwadia, A., Zachariah, M. A., Coumans, J. V., ... & Chambless, L. B. (2021). Predicting discharge disposition following meningioma resection using a multi-institutional natural language processing model. *Neurosurgery*, *88*(4), 838-845.
[18] Bharti, U., Bajaj, D., Batra, H., Lalit, S., Lalit, S., & Gangwani, A. (2020, June). Medbot: Conversational artificial intelligence powered chatbot for delivering tele-health after covid-19. In *2020 5th international conference on communication and electronics systems (ICCES)* (pp. 870-875). IEEE.
[19] Laxmi, P., Gupta, D., Gopalapillai, R., Amudha, J., & Sharma, K. (2021). A Scalable Multi-disease Modeled CDSS Based on Bayesian Network Approach for Commonly Occurring Diseases with a NLP-Based GUI. In *Intelligent Systems, Technologies and Applications* (pp. 161-171). Springer, Singapore.
[20] Khurana, D., Koli, A., Khatter, K., & Singh, S. (2017). Natural language processing: State of the art, current trends and challenges. *arXiv preprint arXiv:1708.05148*.
[21] Lavanya, P. M., & Sasikala, E. (2021, May). Deep learning techniques on text classification using Natural language processing (NLP) in social healthcare network: A comprehensive survey. In *2021 3rd International Conference on Signal Processing and Communication (ICPSC)* (pp. 603-609). IEEE.
[22] Liu, D., Li, Y., & Thomas, M. A. (2017, January). A roadmap for natural language processing research in information systems. In *proceedings of the 50th Hawaii international conference on system sciences*.
[23] Lertvittayakumjorn, P. (2021). Explainable NLP for Human-AI Collaboration.
[24] Adnan, K., & Akbar, R. (2019). An analytical study of information extraction from unstructured and multidimensional big data. *Journal of Big Data*, *6*(1), 1-38.
[25] Fei, H., Ren, Y., Zhang, Y., Ji, D., & Liang, X. (2021). Enriching contextualized language model from knowledge graph for biomedical information extraction. *Briefings in bioinformatics*, *22*(3), bbaa110.
[26] Zirikly, A., Desmet, B., Newman-Griffis, D., Marfeo, E. E., McDonough, C., Goldman, H., & Chan, L. (2022). Information extraction framework for disability determination using a mental functioning use-case. *JMIR Medical Informatics*, *10*(3), e32245.
[27] Liu, X., Chen, H., & Xia, W. (2022). Overview of Named Entity Recognition. *Journal of Contemporary Educational Research*, *6*(5), 65-68.
[28] Li, J., Sun, A., Han, J., & Li, C. (2020). A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*, *34*(1), 50-70.
[29] Li, X., Zhang, H., & Zhou, X. H. (2020). Chinese clinical named entity recognition with variant neural structures based on BERT methods. *Journal of biomedical informatics*, *107*, 103422.
[30] Weber, L., Sänger, M., Münchmeyer, J., Habibi, M., Leser, U., & Akbik, A. (2021). HunFlair: an easy-to-use tool for state-of-the-art biomedical named entity recognition. *Bioinformatics*, *37*(17), 2792-2794.
[31] Hassan, A., & Mahmood, A. (2018). Convolutional recurrent deep learning model for sentence classification. *Ieee Access*, *6*, 13949-13957.

[32] Guo, H., Mao, Y., & Zhang, R. (2019). Augmenting data with mixup for sentence classification: An empirical study. *arXiv preprint arXiv:1905.08941*.
[33] Bagheri, A., Sammani, A., van der Heijden, P. G., Asselbergs, F. W., & Oberski, D. L. (2020). ETM: Enrichment by topic modeling for automated clinical sentence classification to detect patients' disease history. *Journal of Intelligent Information Systems*, *55*(2), 329-349.
[34] Audebert, N., Herold, C., Slimani, K., & Vidal, C. (2019, September). Multimodal deep networks for text and image-based document classification. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 427-443). Springer, Cham.
[35] Behera, B., Kumaravelan, G., & Kumar, P. (2019, December). Performance evaluation of deep learning algorithms in biomedical document classification. In *2019 11th International Conference on Advanced Computing (ICoAC)* (pp. 220-224). IEEE.
[36] Nadif, M., & Role, F. (2021). Unsupervised and self-supervised deep learning approaches for biomedical text mining. *Briefings in Bioinformatics*, *22*(2), 1592-1603.
[37] Yadav, D., Lalit, N., Kaushik, R., Singh, Y., Yadav, A. K., Bhadane, K. V., ... & Khan, B. (2022). Qualitative Analysis of Text Summarization Techniques and Its Applications in Health Domain. *Computational Intelligence and Neuroscience*, *2022*.
[38] Belwal, R. C., Rai, S., & Gupta, A. (2021). A new graph-based extractive text summarization using keywords or topic modeling. *Journal of Ambient Intelligence and Humanized Computing*, *12*(10), 8975-8990.
[39] Belwal, R. C., Rai, S., & Gupta, A. (2021). Text summarization using topic-based vector space model and semantic measure. *Information Processing & Management*, *58*(3), 102536.
[40] Soares, M. A. C., & Parreiras, F. S. (2020). A literature review on question answering techniques, paradigms and systems. *Journal of King Saud University-Computer and Information Sciences*, *32*(6), 635-646.
[41] Yin, Z., Zhang, C., Goldberg, D. W., & Prasad, S. (2019, March). An NLP-based question answering framework for spatio-temporal analysis and visualization. In *Proceedings of the 2019 2nd International Conference on Geoinformatics and Data Analysis* (pp. 61-65).
[42] Meichanetzidis, K., Toumi, A., de Felice, G., & Coecke, B. (2020). Grammar-aware question-answering on quantum computers. *arXiv preprint arXiv:2012.03756*.
[43] Yang, S., Wang, Y., & Chu, X. (2020). A survey of deep learning techniques for neural machine translation. *arXiv preprint arXiv:2002.07526*.
[44] Laskar, S. R., Dutta, A., Pakray, P., & Bandyopadhyay, S. (2019, December). Neural machine translation: English to hindi. In *2019 IEEE conference on information and communication technology* (pp. 1-6). IEEE.
[45] Rahit, K. M., Nabil, R. H., & Huq, M. H. (2019, October). Machine translation from natural language to code using long-short term memory. In *Proceedings of the Future Technologies Conference* (pp. 56-63). Springer, Cham.
[46] Kumar, A., Krishnamurthi, R., Nayyar, A., Sharma, K., Grover, V., & Hossain, E. (2020). A novel smart healthcare design, simulation, and implementation using healthcare 4.0 processes. *IEEE Access*, *8*, 118433-118471.
[47] Li, W., Chai, Y., Khan, F., Jan, S. R. U., Verma, S., Menon, V. G., & Li, X. (2021). A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system. *Mobile Networks and Applications*, *26*(1), 234-252.
[48] Islam, M. S., Hasan, M. M., Wang, X., Germack, H. D., & Noor-E-Alam, M. (2018, May). A systematic review on healthcare analytics: application and theoretical perspective of data mining. In *Healthcare* (Vol. 6, No. 2, p. 54). MDPI.
[49] Fleuren, L. M., Klausch, T. L., Zwager, C. L., Schoonmade, L. J., Guo, T., Roggeveen, L. F., ... & Elbers, P. W. (2020). Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive care medicine*, *46*(3), 383-400.
[50] Thorsen-Meyer, H. C., Nielsen, A. B., Nielsen, A. P., Kaas-Hansen, B. S., Toft, P., Schierbeck, J., ... & Perner, A. (2020). Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. *The Lancet Digital Health*, *2*(4), e179-e191.
[51] Azari, A., Stefanović, Č., Popovski, P., & Cavdar, C. (2019). On the latency-energy performance of NB-IoT systems in providing wide-area IoT connectivity. *IEEE Transactions on Green Communications and Networking*, *4*(1), 57-68.
[52] Khan, S. F. (2017, March). Health care monitoring system in Internet of Things (IoT) by using RFID. In *2017 6th International conference on industrial technology and management (ICITM)* (pp. 198-204). IEEE.

[53] Zhan, X., Long, H., Gou, F., Duan, X., Kong, G., & Wu, J. (2021). A convolutional neural network-based intelligent medical system with sensors for assistive diagnosis and decision-making in non-small cell lung cancer. *Sensors*, *21*(23), 7996.
[54] Abdel-Basset, M., Manogaran, G., Gamal, A., & Chang, V. (2019). A novel intelligent medical decision support model based on soft computing and IoT. *IEEE Internet of Things Journal*, *7*(5), 4160-4170.
[55] Velupillai, S., Mowery, D., South, B. R., Kvist, M., & Dalianis, H. (2015). Recent advances in clinical natural language processing in support of semantic analysis. *Yearbook of medical informatics*, *24*(01), 183-193.
[56] Mittal, Y. K., Paul, V. K., Rostami, A., Riley, M., & Sawhney, A. (2020). Delay factors in construction of healthcare infrastructure projects: a comparison amongst developing countries. *Asian Journal of Civil Engineering*, *21*(4), 649-661.
[57] CEOWORLD magazine. (2021, April 27). Revealed: Countries With The Best Health Care Systems, 2021. https://ceoworld.biz/2021/04/27/revealed-countries-with-the-best-health-care-systems-2021/
[58] Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *New England journal of medicine*, *372*(9), 793-795.
[59] Zhang, X., Kim, J., Patzer, R. E., Pitts, S. R., Patzer, A., & Schrager, J. D. (2017). Prediction of emergency department hospital admission based on natural language processing and neural networks. *Methods of information in medicine*, *56*(05), 377-389.
[60] Sterling, N. W., Patzer, R. E., Di, M., & Schrager, J. D. (2019). Prediction of emergency department patient disposition based on natural language processing of triage notes. *International journal of medical informatics*, *129*, 184-188.
[61] Jin, M., Bahadori, M. T., Colak, A., Bhatia, P., Celikkaya, B., Bhakta, R., ... & Kass-hout, T. (2018). Improving hospital mortality prediction with medical named entities and multimodal learning. *arXiv preprint arXiv:1811.12276*.
[62] Nawab, K., Ramsey, G., & Schreiber, R. (2020). Natural language processing to extract meaningful information from patient experience feedback. *Applied Clinical Informatics*, *11*(02), 242-252.
[63] Bacchi, S., Gluck, S., Tan, Y., Chim, I., Cheng, J., Gilbert, T., ... & Koblar, S. (2020). Prediction of general medical admission length of stay with natural language processing and deep learning: a pilot study. *Internal and emergency medicine*, *15*(6), 989-995.
[64] Arnaud, É., Elbattah, M., Gignon, M., & Dequen, G. (2021, August). NLP-Based Prediction of Medical Specialties at Hospital Admission Using Triage Notes. In *2021 IEEE 9th International Conference on Healthcare Informatics (ICHI)* (pp. 548-553). IEEE.
[65] Young, M., Holmes, N., Kishore, K., Marhoon, N., Amjad, S., Serpa-Neto, A., & Bellomo, R. (2022). Natural language processing diagnosed behavioral disturbance vs confusion assessment method for the intensive care unit: prevalence, patient characteristics, overlap, and association with treatment and outcome. *Intensive care medicine*, *48*(5), 559-569.
[66] Wadia, R., Shifman, M., Levin, F. L., Marenco, L., Brandt, C. A., Cheung, K. H., ... & Krauthammer, M. (2017). A clinical decision support system for monitoring post-colonoscopy patient follow-up and scheduling. *AMIA Summits on Translational Science Proceedings*, *2017*, 295.
[67] Zikos, D., & DeLellis, N. (2018). CDSS-RM: a clinical decision support system reference model. *BMC medical research methodology*, *18*(1), 1-14.
[68] Shen, L., Wright, A., Lee, L. S., Jajoo, K., Nayor, J., & Landman, A. (2021). Clinical decision support system, using expert consensus-derived logic and natural language processing, decreased sedation-type order errors for patients undergoing endoscopy. *Journal of the American Medical Informatics Association*, *28*(1), 95-103.
[69] Leite-Moreira, A., Mendes, A., Pedrosa, A., Rocha-Sousa, A., Azevedo, A., Amaral-Gomes, A., ... & Pimenta, T. (2022). An NLP Solution to Foster the Use of Information in Electronic Health Records for Efficiency in Decision-Making in Hospital Care. *arXiv preprint arXiv:2202.12159*.
[70] Dino, F., Zandie, R., Abdollahi, H., Schoeder, S., & Mahoor, M. H. (2019, November). Delivering cognitive behavioral therapy using a conversational social robot. In 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 2089-2095). IEEE.
[71] Chen, C. W., Tseng, S. P., Kuan, T. W., & Wang, J. F. (2020). Outpatient text classification using attention-based bidirectional LSTM for robot-assisted servicing in hospital. *Information*, *11*(2), 106.

[72] Christopherjames, J. E., Saravanan, M., Thiyam, D. B., Sahib, M. Y. B., Ganapathi, M. V., & Milton, A. (2021, August). Natural Language Processing based Human Assistive Health Conversational Agent for Multi-Users. In *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 1414-1420). IEEE.
[73] Hassan, A., Ali, M. D., Ahammed, R., Bourouis, S., & Khan, M. M. (2021). Development of NLP-Integrated Intelligent Web System for E-Mental Health. *Computational and Mathematical Methods in Medicine*, *2021*.
[74] Sheikhalishahi, S., Miotto, R., Dudley, J. T., Lavelli, A., Rinaldi, F., & Osmani, V. (2019). Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR medical informatics*, *7*(2), e12239.
[75] Jackson, R. G., Patel, R., Jayatilleke, N., Kolliakou, A., Ball, M., Gorrell, G., ... & Stewart, R. (2017). Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. *BMJ open*, *7*(1), e012012.
[76] Van Le, D., Montgomery, J., Kirkby, K. C., & Scanlan, J. (2018). Risk prediction using natural language processing of electronic mental health records in an inpatient forensic psychiatry setting. *Journal of biomedical informatics*, *86*, 49-58.
[77] Mulyana, S., Hartati, S., & Wardoyo, R. (2019, October). A processing model using natural language processing (nlp) for narrative text of medical record for producing symptoms of mental disorders. In *2019 Fourth International Conference on Informatics and Computing (ICIC)* (pp. 1-6). IEEE.
[78] Mukhiya, S. K., Ahmed, U., Rabbi, F., Pun, K. I., & Lamo, Y. (2020, July). Adaptation of IDPT system based on patient-authored text data using NLP. In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 226-232). IEEE.
[79] Ridgway, J. P., Uvin, A., Schmitt, J., Oliwa, T., Almirol, E., Devlin, S., & Schneider, J. (2021). Natural language processing of clinical notes to identify mental illness and substance use among people living with HIV: retrospective cohort study. *JMIR Medical Informatics*, *9*(3), e23456.
[80] Karunakaran, B., Misra, D., Marshall, K., Mathrawala, D., & Kethireddy, S. (2017, December). Closing the loop—Finding lung cancer patients using NLP. In *2017 IEEE international conference on big data (big data)* (pp. 2452-2461). IEEE.
[81] Si, Y., & Roberts, K. (2018). A frame-based NLP system for cancer-related information extraction. In *AMIA annual symposium proceedings* (Vol. 2018, p. 1524). American Medical Informatics Association.
[82] Mowery, D. L., Kawamoto, K., Bradshaw, R., Kohlmann, W., Schiffman, J. D., Weir, C., ... & Del Fiol, G. (2019). Determining onset for familial breast and colorectal cancer from family history comments in the electronic health record. *AMIA Summits on Translational Science Proceedings*, *2019*, 173.
[83] Alawad, M., Yoon, H. J., Gao, S., Mumphrey, B., Wu, X. C., Durbin, E. B., ... & Tourassi, G. (2020). Privacy-preserving deep learning NLP models for cancer registries. *IEEE Transactions on Emerging Topics in Computing*, *9*(3), 1219-1230.
[84] Deshmukh, P. R., & Phalnikar, R. (2021). Information extraction for prognostic stage prediction from breast cancer medical records using NLP and ML. *Medical & Biological Engineering & Computing*, *59*(9), 1751-1772.
[85] Yoon, H. J., Stanley, C., Christian, J. B., Klasky, H. B., Blanchard, A. E., Durbin, E. B., ... & Tourassi, G. D. (2022). Optimal vocabulary selection approaches for privacy-preserving deep NLP model training for information extraction and cancer epidemiology. *Cancer Biomarkers*, *33*(2), 185-198.
[86] Patterson, O. V., Freiberg, M. S., Skanderson, M., J Fodeh, S., Brandt, C. A., & DuVall, S. L. (2017). Unlocking echocardiogram measurements for heart disease research through natural language processing. *BMC cardiovascular disorders*, *17*(1), 1-11.
[87] Afzal, N., Mallipeddi, V. P., Sohn, S., Liu, H., Chaudhry, R., Scott, C. G., ... & Arruda-Olson, A. M. (2018). Natural language processing of clinical notes for identification of critical limb ischemia. *International journal of medical informatics*, *111*, 83-89.
[88] Thaiparnit, S., Kritsanasung, S., & Chumuang, N. (2019, July). A classification for patients with heart disease based on hoeffding tree. In *2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (pp. 352-357). IEEE.
[89] Bagheri, A., Groenhof, T. K. J., Veldhuis, W. B., de Jong, P. A., Asselbergs, F. W., & Oberski, D. L. (2020). Multimodal learning for cardiovascular risk prediction using EHR data. *arXiv preprint arXiv:2008.11979*.

[90] Sammani, A., Bagheri, A., van der Heijden, P] G., Te Riele, A. S., Baas, A. F., Oosters, C. A. J., ... & Asselbergs, F. W. (2021). Automatic multilabel detection of ICD10 codes in Dutch cardiology discharge letters using neural networks. *NPJ digital medicine*, *4*(1), 1-10.
[91] Zaman, S., Petri, C., Vimalesvaran, K., Howard, J., Bharath, A., Francis, D., ... & Linton, N. (2022). Automatic diagnosis labeling of cardiovascular MRI by using semisupervised natural language processing of text reports. *Radiology: Artificial Intelligence*, *4*(1).
[92] Weng, W. H., Wagholikar, K. B., McCray, A. T., Szolovits, P., & Chueh, H. C. (2017). Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach. *BMC medical informatics and decision making*, *17*(1), 1-13.
[93] Weng, W. H., Wagholikar, K. B., McCray, A. T., Szolovits, P., & Chueh, H. C. (2017). Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach. *BMC medical informatics and decision making*, *17*(1), 1-13.
[94] Chen, T., Dredze, M., Weiner, J. P., Hernandez, L., Kimura, J., & Kharrazi, H. (2019). Extraction of geriatric syndromes from electronic health record clinical notes: assessment of statistical natural language processing methods. *JMIR medical informatics*, *7*(1), e13039.
[95] Liu, R., Greenstein, J. L., Sarma, S. V., & Winslow, R. L. (2019, July). Natural language processing of clinical notes for improved early prediction of septic shock in the ICU. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 6103-6108). IEEE.
[96] Oliwa, T., Furner, B., Schmitt, J., Schneider, J., & Ridgway, J. P. (2021). Development of a predictive model for retention in HIV care using natural language processing of clinical notes. *Journal of the American Medical Informatics Association*, *28*(1), 104-112.
[97] Chen, L., Gu, Y., Ji, X., Sun, Z., Li, H., Gao, Y., & Huang, Y. (2020). Extracting medications and associated adverse drug events using a natural language processing system combining knowledge base and deep learning. *Journal of the American Medical Informatics Association*, *27*(1), 56-64.
[98] Wang, J., Abu-el-Rub, N., Gray, J., Pham, H. A., Zhou, Y., Manion, F. J., ... & Zhang, Y. (2021). COVID-19 SignSym: a fast adaptation of a general clinical NLP tool to identify and normalize COVID-19 signs and symptoms to OMOP common data model. *Journal of the American Medical Informatics Association*, *28*(6), 1275-1283.
[99] Song, G., Chung, S. J., Seo, J. Y., Yang, S. Y., Jin, E. H., Chung, G. E., ... & Han, H. W. (2022). Natural Language Processing for Information Extraction of Gastric Diseases and Its Application in Large-Scale Clinical Research. *Journal of Clinical Medicine*, *11*(11), 2967.